MOLECULAR CHARACTERISATION OF THE GENE ENCODING Δ^1-PYRROLINE-5-CARBOXYLATE REDUCTASE ISOLATED FROM *Arabidopsis thaliana* (L.) Heynh.

PETER DEREK HARE

Submitted in fulfilment of the academic requirements for the degree of Master of Science in the Department of Botany, University of Natal

Pietermaritzburg

1995

PREFACE

The experimental work described in this dissertation was carried out in the Department of Botany, University of Natal, Pietermaritzburg, from January 1993 to December 1994, under the supervision of Doctor William A. Cress.

These studies represent work done by the author and have not otherwise been submitted in any form for any degree or diploma to any other University. Where use has been made of the work of others it is duly acknowledged in the text.

PD Hare

Peter Derek Hare

January 1995

ACKNOWLEDGEMENTS

Sincere thanks are due to many individuals who have contributed towards this project:

To my supervisor Dr W.A. Cress for his guidance and encouragement throughout the course of this project.

To my family for their invaluable support.

To my laboratory colleagues, especially Sandra du Plessis, for their interest, help and encouragement.

To Prof. J. van Staden for his interest and efficiency in running a well-organised department.

To my "E-mail pal" Diana Horvath for her sustained help, encouragement and unfailing sense of humour over the past eighteen months.

To the technical staff of the Botany Department for their efficient assistance when needed.

To Dr M. Raynal and Dr J. Fleck for providing the YAP057 and FAFJ25 cDNA clones respectively.

To Prof. D. Brothers and Paul Bayliss for advice and assistance in the use of NTSYSpc.

To Ron Berry for his assistance in retrieving many of the computer programs used.

To the Foundation for Research Development and University of Natal for financial assistance in the form of bursaries.

ABSTRACT

In *Arabidopsis thaliana* (L.) Heynh. the size of the pool of free proline increases up to 27-fold in response to osmotic stress. The magnitude of this accumulation is dependent upon the rate of imposition of the stress. Numerous reports have suggested a role for proline accumulation as a general adaptation to environmental stress. However, controversy surrounds the beneficial effect of proline accumulation in plants under adverse environmental conditions.

Stress-induced proline accumulation in plants occurs mainly by *de novo* synthesis from glutamate. The final and only committed step of proline biosynthesis in plants is catalysed by Δ^1-pyrroline-5-carboxylate reductase (P5CR). The sequence of an incomplete 999 bp cDNA encoding P5CR from *A. thaliana* was determined. This enabled a preliminary molecular study of the structure and function of both the gene and the corresponding enzyme.

The 999 bp cDNA insert in the clone YAP057 was sequenced on the sense and antisense strands following subcloning of four sub-fragments in appropriate orientations. Comparison with known plant P5CR sequences revealed that YAP057 does not encode the first 23 N-terminal amino acids of P5CR from *Arabidopsis*. However, it does encode the remaining 253 amino acid residues of *Arabidopsis* P5CR. The cDNA YAP057 is complete on the 3' end as indicated by the presence of a poly(A) tail. The nucleotide sequence determined shows complete homology to the corresponding exons of the genomic copy of a *bona fide* gene encoding P5CR in *A. thaliana* (Verbruggen *et al.*, 1993). The only difference observed between the sequence of YAP057 and that of a cDNA sequenced by these workers is that polyadenylation was initiated seven nucleotides earlier in YAP057 than in the sequence of the published cDNA.

Genomic Southern analysis suggests the presence of only a single copy of the gene encoding P5CR in *Arabidopsis*. Restriction mapping and sequencing the ends of another incomplete *Arabidopsis* P5CR cDNA clone FAFJ25 (664 bp) indicated that the regions sequenced were completely homologous to the corresponding portions of YAP057.

Analysis of codon usage in the *Arabidopsis* gene encoding P5CR revealed it to closely resemble the consensus pattern of codon usage in *A. thaliana*. This suggests that the gene is moderately expressed. Expression of the gene encoding P5CR in *Arabidopsis* is not likely to be subject to translational control.

Although P5CR from *A. thaliana* has a fairly high composition of hydrophobic amino acid residues, it does not possess any stretches of hydrophobic amino acids of sufficient length to act as membrane-spanning domains or to anchor the enzyme in a membrane. Neither does it contain an N-terminal leader sequence capable of directing it to either the plastid or mitochondrion. The enzyme therefore appears to be cytosolic.

The nucleic acid and deduced amino acid sequences of *Arabidopsis* P5CR were compared with those from eleven other organisms for which P5CR sequences are currently available. Except among the three different plants examined, P5CR sequences displayed less identity at the amino acid level than at the nucleotide level.

The deduced amino acid sequence of *Arabidopsis* P5CR exhibits high similarity to the corresponding genes and amino acid sequences of P5CR from soybean and pea. Lower but significant similarity was observed to the amino acid sequences of P5CRs from human, *Saccharomyces cerevisiae* and the bacteria *Escherichia coli*, *Pseudomonas aeruginosa*, *Thermus thermophilus*, *Mycobacterium leprae*, *Treponema pallidum* and *Methanobrevibacter smithii*. Similarity was also observed to the translational product of a gene from *Bacillus subtilis* with high homology to the *E. coli proC* gene. However, construction of a phenogram indicating the relatedness of the various P5CR enzymes suggests that sequence analysis of this enzyme is not a good indicator of evolutionary relatedness of organisms from different biological kingdoms.

Multiple alignment of the twelve known P5CR sequences indicated homology between the sequences across their entire lengths. Homology was particularly high in the C-terminal portions of the P5CRs studied. It is speculated that this region may be of importance in binding of the substrate Δ^1-pyrroline-5-carboxylate (P5C). Another region displaying high sequence conservation was found in the central portion of all P5CRs. All P5CRs studied, with the exception of P5CR from *T. pallidum* contained an N-terminal domain capable of binding a nicotinamide dinucleotide cofactor. Comparison of this region with consensus

sequences for NADH and NADPH binding sites in proteins suggests that NADPH is the preferred reductant used by P5CRs from plants and human. In contrast, the N-terminal domains of P5CRs from *S. cerevisiae*, *M. smithii*, *T. thermophilus* and *M. leprae* display greater similarity to a consensus NADH-binding site. The definite preference of plant P5CRs for NADPH in comparison with NADH suggests that P5CR may be involved in regulating the redox potential within plant cells and that this step in proline biosynthesis from glutamate may be of importance in overall metabolic regulation.

Three amino acid residues are universally conserved in all P5CRs studied. All are found within blocks of high sequence similarity. These residues are likely to be of importance in the structure or catalytic mechanism of P5CR. A number of other residues are common to several of the enzymes examined. These may also be of importance in subsequent manipulation of *Arabidopsis* P5CR at the molecular level.

Prediction of the putative secondary structures of *A. thaliana*, soybean, pea, human and *E. coli* indicated a high degree of similarity between the enzymes. This was particularly evident in the region of the putative P5C-binding domain. Considerable similarity exists in hydrophobicity profiles of P5CRs from these five organisms.

Proline levels in reproductive organs of unstressed *Arabidopsis* plants were considerably higher than those in vegetative tissues. This suggests differential expression of enzymes involved in proline metabolism in these organs. *In situ* hybridisation studies indicated an increase in levels of mRNA transcripts encoding P5CR in stem tissues in response to water deprivation stress. Regulation of levels of mRNA transcript encoding P5CR in *Arabidopsis* therefore appears to be an osmotically sensitive process. Furthermore, this accumulation of transcript occurred in a tissue-specific manner. In particular, an increase in levels of transcript encoding P5CR was observed in the cortical parenchyma, phloem, vascular cambium and pith parenchyma in the vicinity of the protoxylem.

The significance of these findings in contributing to a better understanding of the role of proline in adaptation to environmental stress is discussed.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

A	-	adenine
ABA	-	abscisic acid
ADP	-	adenosine diphosphate
ATP	-	adenosine triphosphate
bp	-	base pairs
C	-	cytosine
CaMV	-	cauliflower mosaic virus
cDNA	-	complementary DNA
dATP	-	deoxyadenosine triphosphate
dCTP	-	deoxycytosine triphosphate
(d)dNTP	-	(di)deoxynucleoside triphosphate
dGTP	-	deoxyguanosine triphosphate
dH_2O	-	distilled water
DIG	-	digoxigenin
DNase	-	deoxyribonuclease
dTTP	-	deoxythymidine triphosphate
dUTP	-	deoxyuridine triphosphate
EDTA	-	ethylenediamine tetra acetic acid
EST	-	expressed sequence tag
EtBr	-	ethidium bromide
FAD	-	flavin-adenine dinucleotide (oxidised form)
FUE	-	far-upstream element
G	-	guanine
GSA	-	glutamic-γ-semialdehyde
γ-GK	-	γ-glutamyl kinase
GUS	-	β-glucuronidase
IPTG	-	isopropylthiogalactoside
kb	-	kilobase
K_m	-	Michaelis constant
M_r	-	molecular mass
mRNA	-	messenger RNA
NAD(H)	-	nicotinamide adenine dinucleotide (reduced form)
NADP(H)	-	nicotinamde adenine dinucleotide phosphate (reduced form)
NAD(P)	-	oxidised form of either NAD or NADP

NAD(P)H	-	reduced form of either NAD or NADP
nt	-	nucleotide
NUE	-	near-upstream element
OAT	-	ornithine δ-aminotransferase
ORF	-	open reading frame
PAGE	-	polyacrylamide gel electrophoresis
PBS	-	phosphate-buffered saline
P2C	-	Δ^1-pyrroline-2-carboxylate
P5C	-	Δ^1-pyrroline-5-carboxylic acid
P2CR	-	Δ^1-pyrroline-2-carboxylate reductase
P5CR	-	Δ^1-pyrroline-5-carboxylate reductase
P5CS	-	Δ^1-pyrroline-5-carboxylate synthetase
PEG	-	polyethylene glycol
pfu	-	plaque-forming unit
pI	-	isoelectric point
RNase	-	ribonuclease
rRNA	-	ribosomal RNA
RWC	-	relative water content
SDS	-	sodium dodecyl sulphate
SE	-	standard error of the mean
snRNP	-	small nuclear ribonucleoprotein
ss	-	single stranded
SSC	-	standard saline citrate buffer
STS	-	sequence tagged site
®	-	registered trademark
T	-	thymine
TCA	-	tricarboxylic acid
TE	-	Tris-EDTA buffer
TEMED	-	*N,N,N',N'*-tetramethylethylenediamine
T_m	-	melting temperature
™	-	registered trademark
tRNA	-	transfer RNA
U	-	uracil
X-gal	-	5-bromo-4-chloro-3-indoyl-β-D-galactoside
Y	-	any pyrimidine (cytosine or thymine or uracil)

Three-letter and one-letter abbreviations of the twenty amino acids found in proteins

A	Ala	alanine
C	Cys	cysteine
D	Asp	aspartic acid
E	Glu	glutamic acid
F	Phe	phenylalanine
G	Gly	glycine
H	His	histidine
I	Ile	isoleucine
K	Lys	lysine
L	Leu	leucine
M	Met	methionine
N	Asn	asparagine
P	Pro	proline
Q	Gln	glutamine
R	Arg	arginine
S	Ser	serine
T	Thr	threonine
V	Val	valine
W	Trp	tryptophan
Y	Tyr	tyrosine

1. INTRODUCTION

The ability of a plant to grow depends not only on its photosynthetic capacity, but also on its ability to maintain dry matter accumulation in the presence of a variety of stresses. The sessile habit of plants requires that they must be able to adapt to environmental stress. Despite their developmental plasticity, the ways in which plants respond to their environment are constrained by their genetic makeup.

Furthermore, most plants do not achieve their full genetic potential for productivity because of environmental stress. For most crops growing under field conditions, average agricultural yields are only 12-30% of maximum yields. This suggests that current agricultural productivity, even in first world countries, is probably only 25% of its maximum potential (Boyer, 1982). Of the wide spectrum of abiotic stresses encountered by most plants in the field, hyperosmotic stresses caused by drought and salinity are the major constraints to crop yield (Boyer, 1982).

The reduction in plant growth accompanying environmental stress is usually beyond the control of agriculturalists. Accordingly, stress physiology is currently one of the most active areas of research in contemporary plant physiology. Better understanding of how plants respond to unfavourable physiochemical environments and how such stresses limit growth is likely to provide a rational basis for the development of stress tolerant varieties. Appreciation of the physiological mechanisms that enable plant survival in suboptimal environments is of particular interest in developing countries, where these are often the only habitats into which agriculture can expand. In the light of increasing population pressures, accelerated rates of desertification and salinisation and the widely-anticipated effects of global climate change, such research is likely to assume even greater importance in the future.

The success of traditional breeding practices in enhancing stress-tolerance is limited (Horsch, 1993). Since genetic recombination does not usually occur at sites that are close together, traditional approaches normally result in the transfer of large blocks of genes from the donor plant to the recipient. Of these, only one or a few may be desired. The remainder may be tolerated usually without observable effect, although examples of unwanted linkages are well

known (Horsch, 1993). Alteration of precise physiological attributes of plants is difficult, if not impossible, using such approaches. Furthermore, the pool of genes available for transfer to crops is usually limited to those found in other varieties of the same crop or closely related wild species (Horsch, 1993).

The advent of genetic transformation of plants in the past decade has introduced the possibility of directly influencing particular metabolic responses to stress by a precise transfer of single genes, theoretically from any organism, into plants. One such response is the accumulation of the imino acid proline. Free proline has been reported to accumulate in a range of plants in response to stresses including water deprivation, salinity stress, temperature stress, heavy metal stress, anaerobiosis and pathogen infection. In some species, proline accumulated in response to osmotic stress may account for up to 10% of the dry weight (Stewart and Lee, 1974). In maize roots grown at low water potentials, proline accumulation accounts for approximately 45% of the total osmotic adjustment (Voetberg and Sharp, 1991). While several amino acids are known to accumulate in response to osmotic stress, proline has a specific protective role in the adaptation of plant cells to water deprivation (Handa *et al.*, 1986).

Since the discovery of this phenomenon over forty years ago (Kemble and MacPherson, 1954), a large body of literature has accumulated concerning the physiology of proline accumulation. However, controversy continues as to whether the response has any real adaptive value or is in part simply an incidental consequence of stress-induced changes in metabolism.

The objective of this study was the characterisation of the gene from *Arabidopsis thaliana* (L.) Heynh. that encodes the last enzyme in the proline biosynthetic pathways from glutamate and ornithine, where Δ^1-pyrroline-5-carboxylate serves as an intermediate. Not only should better understanding of proline accumulation at the molecular level complement the large body of physiological information that has accumulated concerning the process, but it may also pave the way for the development of transgenic crops exhibiting enhanced environmental tolerance. The choice of *Arabidopsis thaliana* as the plant of study was based on the use of this species as a model system in plant molecular biology.

2. Literature Review

2.1 Proline accumulation as a response to environmental stress

2.1.1 A definition of stress

The physical environment of all living organisms is in a constant state of flux. In order to survive environmental fluctuations, organisms have evolved adaptations to the environments from which they originated. Since mature plants, unlike animals, cannot escape from adverse environmental conditions, selection for the development of tolerance to the prevalent environmental conditions has been strong throughout their evolution. The absence of mobility among higher plants has resulted in their acquiring unique sets of responses to environmental insult.

Environmental variables that may impact on plant growth and distribution include temperature, nutrient deficiency, flood-induced anaerobiosis, high irradiance, metal toxicity and availability of water. In addition to these abiotic factors, biotic variables such as herbivory by insects, grazing by animals and infection by microbial or fungal pathogens may also limit plant growth and reproductive capacity.

Abiotic factors in the environment such as light, water, temperature and mineral nutrients may occur at either minimum and maximum threshold levels beyond which a plant cannot survive. However, these sub- and supra-optimal limits seldom coincide with the thresholds for damage (Levitt, 1980). Any deviation from the optimal level of the factor concerned may constitute a stress on the plant (Levitt, 1980). A precise definition of plant stress is difficult. In engineering terms, stress is "a force or a system of forces producing deformation or strain". By analogy, in a biological system stress may thus be defined as the result of any external force causing an internal strain (Levitt, 1980). Osmond *et al.* (1987) defined stress as "any factor that decreases plant growth and reproduction below the genotype's potential". Within the acceptable limits of the factor concerned, there are degrees of stress. The more removed the factor is from the level optimal for growth, the more stressful the environment.

Since the environmental conditions to which both native and agricultural plant communities are exposed are seldom optimal, plants can be said to be experiencing some stress at all times. Most natural environments are continuously sub- or supra-optimal with respect to at least one of the environmental parameters (Levitt, 1980). Furthermore, even the most moderate environments experience seasonal fluctuations in light, moisture, temperature, nutrients or exposure to pathogens, often to levels not optimal for plant growth (Chapin, 1991a). Often, these stresses do not occur in isolation, but are coupled (Nover, 1989). For example, high temperature and reduced water supply often occur at the same time. Stomatal closure resulting from deficit prevents transpirational cooling and exacerbates heat stress. Restricted water availability decreases ion mobility and may thus reduce the rate at which nutrients become available in soils (Chapin, 1991b). Water deficit also reduces the flux of nutrients to the root surface, the rate of nutrient uptake by plant roots, and the requirement of plants for nutrients (Chapin, 1991b). Soil salinity may also interfere with nutrient uptake. Often these indirect effects on plant nutrition are as important as the direct effects of water stress. Also, most environmental stresses increase susceptibility to infection by pathogens. Therefore, plants continuously need to adapt to new combinations of stresses they may encounter.

The acclimation of plants to environmental stresses must ultimately occur at the level of the genome. It may involve both short-term responses that confer immediate benefit as well as long-term physiological and/or morphological modifications that ensure survival in the event of the stress persisting or even increasing in severity. These changes help to minimise stress in the plant and to maximise the use of internal and external resources.

The advances in molecular technologies during the past decade have enabled preliminary dissection of genomic responses to stress and the mechanisms whereby these are regulated. To date, a number of plant genes have been identified, the expression of which increases rapidly in response to water deficit, temperature shifts, pathogenicity, anaerobiosis, hyperoxia and exposure to heavy metals (Sachs and Ho, 1986). Although much remains to be understood concerning the response of the plant genome to stress, these findings are already enhancing our understanding of how plants perceive a stressful situation and respond appropriately to the challenge.

2.1.2 The functional significance of proline accumulation during hyperosmotic stress

2.1.2.1 Adaptations to hyperosmotic stress

Water is an essential resource for plant life. At the cellular level, it is used in chemical reactions as well as to maintain subcellular compartmentation by conferring stability to membranes. At the whole plant-level, water is the main carrier for substances travelling among plant organs and tissues. Moreover, owing to the large difference in water potential between the hydrated plant cell and the much drier atmosphere, no other substance in plants is replaced in the same quantities as water (Boyer, 1982). Therefore, any limitation in water availability affects almost all plant functions, including photosynthetic carbon fixation, mineral nutrient uptake by roots and solute transport (Chapin, 1991a). Accordingly, stress physiologists concur that of all the major physiochemical resources exploited by plants, namely water, carbon dioxide, soil nutrients, oxygen and radiation, water is the most limiting resource (Boyer, 1982). In the field, the two conditions contributing most frequently to water deficit are drought and soil salinity (Boyer, 1982).

A direct and inseparable relationship exists between salinity stress and water deprivation. Salinity causes water stress by lowering the water potential of the rooting medium. The resultant lowering of osmotic potential interferes with the ability of plants to use available water. In the rest of this text, the term hyperosmotic stress will be used to describe both drought and salinity stress. However, it is important to note that salinity stress also has an ionic component, which may introduce stress effects not found in plants which have simply been deprived of water.

Considering the vital need for maintaining adequate levels of water, plants have evolved a number of responses to cope with osmotic stress. At the cellular level, these are manifested by chemical, molecular and physiological responses. In addition, structural and morphological modifications may occur at the level of the entire organism (McCue and Hanson, 1990). Of these adaptations, biochemical responses at the cellular level appear to be the most conserved throughout the plant kingdom (Hanson and Hitz, 1982). In contrast, morphological and long-term physiological responses are often more specialised and restricted to individual species.

2.1.2.2 Accumulation of organic osmolytes as a response to hyperosmotic stress

Water loss resulting from hyperosmotic stress initiates a number of regulatory processes that enable adaptation to the new cellular conditions (Hanson and Hitz, 1982). In particular, water deficits induce lowering of cellular osmotic potential as a mechanism of ensuring maintenance of cell turgor. This decline in the osmotic potential as a response to water deficit may be achieved by decreased cell volume, which results in increased concentrations of cellular solutes as water is lost from the vacuole. Such osmoregulation occurs until intracellular osmotic potential is approximately equal to the potential of the medium surrounding the cell (Turner and Jones, 1981).

However, the best characterised biochemical response of plant cells to water deficit is the accumulation of organic osmolytes, which enable maintenance of water potential by osmotic adjustment. The accumulation of organic solutes within the cell reduces cellular water potential below the external water potential, while avoiding deleteriously high ionic strength. This enables water to move into the cell and to be maintained there (Turner and Jones, 1981). Osmotic adjustment is an important mechanism in tolerance of hyperosmotic stress because it enables continuation of cell expansion (Wyn Jones and Gorham, 1983; Hsiao *et al.*, 1976), stomatal and photosynthetic adjustments (Ludlow, 1980) and continuation of plant growth even under the adverse environmental conditions.

A number of organic osmolytes have been identified in plant cells. Besides proline, they include betaines and other zwitterionic quaternary ammonium compounds (Rhodes and Hanson, 1993; Hanson *et al.*; 1994), polyols such as glycerol, mannitol, sucrose, sorbitol and pinitol (Brown and Hellebust, 1978; Handa *et al.*, 1983; Le Rudulier *et al.*, 1984; Binzel *et al.*, 1987; Adams *et al.*, 1992), protein amino acids including asparagine (Barnett and Naylor, 1966; Venekamp and Koot, 1988) and arginine (Boggess and Stewart, 1976; Boggess *et al.*, 1976b; Stewart and Boggess, 1977) as well as non-protein amino acids such as ornithine (Boggess and Stewart, 1976; Boggess *et al.*, 1976a) and γ-aminobutyrate (Handa *et al.*, 1986). The importance of osmolyte accumulation in plants exposed to hyperosmotic stress has been demonstrated recently by the enhanced stress tolerance of transgenic tobacco plants expressing a bacterial gene causing mannitol accumulation (Tarczynski *et al.*, 1993).

Unlike cellular inorganic solutes such as K^+, these osmoprotectants stabilise proteins and membranes when present at sufficiently high concentrations and may thus raise the cytoplasmic osmotic pressure in stressed cells without deleterious effects (Yancey *et al.*, 1982). This compatibility with cellular metabolism has resulted in them being referred to as compatible solutes (Brown and Simpson, 1972). Excessive levels of inorganic ions would not only inactivate enzymes, but also alter the ionic composition of the cytosol. This would in turn, most likely affect cellular electroneutrality and membrane-associated phenomena and thereby negate the ultimate goal of osmotic regulation, which is the preservation of macromolecular activity. Therefore, although accumulation of inorganic ions is important in osmotic adjustment of plant cells, it is confined to the metabolically inactive vacuolar compartment. Non-toxic compatible solutes accumulate in the cell cytoplasm (Stewart and Lee, 1974; Hall *et al.*, 1978; Paul and Cockburn, 1989; Ketchum *et al.*, 1991). Together, increased levels of organic solutes in the cytoplasm and accumulation of electrolytes in the vacuole restores low water potential in cells, thus allowing water absorption and the appropriate turgor needed for growth.

However, in metabolic terms, osmolyte synthesis is far more costly than electrolyte accumulation. Analysis of the cost benefit of turgor regulation with different solutes (Raven, 1985) indicates that only 2-4 mol photons of light energy is needed for the accumulation of 1 Osmol KCl or NaCl, whereas 68-78 mol photons is needed for the synthesis of 1 Osmol sorbitol or mannitol, 70-93 mol photons for 1 Osmol proline, and 78-101 photons for 1 Osmol glycine betaine. The exact amount of energy needed in each case depends on whether the solutes are accumulated in the roots or shoots and, for proline and glycine betaine, also on whether NO_3^- or NH_4^+ is used as nitrogen source (Raven, 1985). Therefore, although metabolically expensive to produce, compatible solutes are preferable to inorganic ions for cytoplasmic osmotic adjustment.

Despite the range of cellular osmolytes identified in a variety of plants, proline appears to be the most abundant compatible solute accumulated during hyperosmotic stress in a wide spectrum of plant species. Proline accumulation in response to hyperosmotic stress has been demonstrated in a large number of plants, some of which are indicated in Table 2.1.

Cellular accumulation of proline as a response to dehydration is a widespread phenomenon, occurring almost universally in living cells. Besides algae and higher plants, proline

Table 2.1 : Induction of proline accumulation during hyperosmotic stress in plants. Where more than one set of data has been obtained from a single species, the highest level of proline accumulation reported is given. Adapted from Delauney and Verma (1993).

Species	[Proline] fold increase / control	Stress condition	Reference
Algae			
Stichococcus bacillaris	140	1 234 mosmol.kg^{-1}	Brown and Hellebust (1978)
Dicotyledonous plants			
Arabidopsis thaliana (L.) Heynh.	8-20	120 mM NaCl, KCl 60 mM PEG	Chiang and Dandekar (1991)
Aster tripolium L.	8	333 mM NaCl	Gosu *et al.* (1982)
Glycine max L.			
leaves	11	200 mM NaCl	Kohl *et al.* (1991)
nodules	3		
Gossypium hirsutum L.	>100	withholding water	McMichael and Elmore (1977)
Helianthus tuberosum L.			
tubers	10.6	1.0 M sorbitol	Wrench *et al.* (1980)
Lycopersicon esculentum L.			
cell suspension culture	319	25% PEG	Handa *et al.* (1986)
Medicago sativa L.			
roots	8		
bacteroids	13	150 mM NaCl	Fou'gere *et al.* (1991)
cytosol	11		
Mesembryanthemum nodiflorum L.			
cell suspension culture	7	400 mM NaCl	Treichel (1986)
Mesembryanthemum crystallinum L.	10	400 mM NaCl	Thomas *et al.* (1992)
Nicotiana tabacum L.			
cell suspension culture	4.4	428 mM NaCl	Binzel *et al.* (1988)
leaves	20	200 mM NaCl	Szoke *et al.* (1992)
Pisum sativum L.	3.6	120 mM NaCl	Bar-Nun and Poljakoff-Mayber (1977)
Solanum tuberosum L.			
cell suspension culture	9	10% PEG	Corcuera *et al.* (1989)
Spinacia oleraceae L.	11	- 2 mPa	Huang and Cavalieri (1979)
Tamarix tetragyna L.	4.8	120 mM NaCl	Bar-Nun and Poljakoff-Mayber (1977)
Vicia faba L.	9	2-day drought	Venekamp and Koot (1988)
Monocotyledonous plants			
Hordeum vulgare L.	3	-1.5 mPa	Boggess *et al.* (1976a)
Oryza sativa L.	4	50 mM KCl	Chou *et al.* (1991)
Pennisetum americanum L.	20	1% NaCl	Das *et al.* (1990)
Sorghum bicolor (L.) Moench			
Var. Feterita (leaves)	16.24	-21.1 bar	Blum and Ebercon (1976)
Var. BTx3197	42.4	-16.2 bar	Bhaskaran *et al.* (1985)
Triticum aestivum L.			
apex and leaves	195.4	-3.6 mPa	Munns *et al.* (1979)

accumulation in response to hyperosmotic stress has also been observed in eubacteria (Measures, 1975; Le Rudulier *et al.*, 1984; Whatmore *et al.*, 1990), diatoms (Schobert, 1977a), protozoa (Geoffrion and Larochelle, 1984; Poulin *et al.*, 1987), algae (Brown and Hellebust, 1978; Greenway and Setter, 1979; Brown and Hellebust, 1980; Laliberte and Hellebust, 1989a), invertebrates (Gilles, 1979; Burton, 1991) and vertebrates (Lambert and Hoffman, 1982). This supports the proposal that prior to the evolution of the homeoosmotic mechanisms in higher organisms, proline accumulation may have been a universal response to osmotic stress (Measures, 1975). It is interesting to note that despite the great evolutionary distance between bacteria, animals and plants, only a few classes of osmolytes are found (Yancey *et al.*, 1982). This convergence probably reflects the need for common properties, especially the demand for compatibility with subcellular structure and function. Apparently, proline has many of these properties.

Although a number of free amino acids accumulate in many organisms in response to hyperosmotic stress, the preferential use of proline is most likely due to a combination of metabolic reasons and the unique biophysical and biochemical properties of this imino acid.

2.1.2.3 Biophysical and biochemical advantages of proline as an osmoticum

Throughout the literature it is generally assumed that, at least in the case of hyperosmotic stress, proline acts primarily as a compatible solute which mediates osmotic adjustment. Focus has therefore centred on the osmotic and biophysical phenomena associated with proline accumulation, in particular its protective effect on cytosolic enzymes and various cellular structures during dehydration. The inclusion of the α-amino group within the pyrrolidone ring structure bestows a number of special properties on proline relative to other amino acids (Adams and Frank, 1980).

Two alternative hypotheses have been proposed for the mechanism whereby proline stabilises enzymes during osmotic stress. The first hypothesis, proposed by Schobert (1977b) is that proline interacts with hydrophobic surface residues of proteins. Binding of the aliphatic portion of the proline ring to nonpolar residues on proteins via hydrophobic interactions results in exposure of the highly charged carboxyl and imino moieties of the

proline molecule to the aqueous environment. This interaction therefore results in coating of the protein with a hydrophilic shell (Schobert, 1977b). The increase in the total hydrophilic area of the protein stabilises it by increasing its solubility in an environment of low water availability and high ionic strength. Support for the conclusion that proline has unusual interactions with proteins was provided by the observation of Schobert and Tschesche (1978) that proline at concentrations of 5-6 M could enhance the solubility of insulin, β-lactoglobulin and bovine serum albumin.

However, the mechanism for the mode of action of proline on protein stability proposed by Schobert (1977b) has been contradicted by the observations of other workers (Pollard and Wyn Jones, 1979; Paleg *et al.*, 1981, 1984; Nash *et al.*, 1982; Arakawa and Timasheff, 1983, 1985) who have concluded that the special protective properties of proline and other organic osmoprotectants on protein structure derive precisely from the fact that they tend to avoid protein surfaces. Arakawa and Timasheff (1985) demonstrated that proline does not affect the partial molar volumes of proteins in aqueous solutions. These workers suggested that proteins are stabilised by solutes which are excluded from their surfaces because the surface area of denatured proteins is generally greater than that of native proteins (Arakawa and Timasheff, 1985). Therefore, solutes that are excluded from protein surfaces tend to favour the native conformation. Arakawa and Timasheff (1985) also noted that organic compounds that are excluded from protein surfaces are generally uncharged at physiological pH and thus have minimum electrostatic interactions with proteins. In this context it is interesting to note that proline does not carry a net electrical charge at pH 7 (Csonka, 1989). Proline can therefore be accumulated to high intracellular concentrations without greatly disturbing the structures of cellular macromolecules. Solutes that carry a net electrical charge are generally more deleterious to protein stability than nonpolar or zwitterionic solutes (Arakawa and Timasheff, 1985).

Besides having a stabilising effect on proteins, it has also been proposed that proline action may involve effects on the hydration layer surrounding phospholipids and possibly also its intercalation between phospholipid head groups (Rudolf *et al.*, 1986). Several studies have identified membranes as being a primary target of damage during hyperosmotic stress (Heber, 1967; Heber *et al.*, 1973; Santarius, 1973; Steponkus, 1984). This suggests that maintenance of membrane integrity might be an important adaptation to conditions of reduced water availability.

Proline may also stabilise other cellular structures such as polyribosomes. In seedlings of millet, proline enhances the incorporation of radioactive precursors into proteins, causing an increased translatability of mRNAs (Kandpal and Rao, 1985). Stability of the machinery of protein biosynthesis during stress is likely to be important in order to ensure continued translation of mRNAs involved in acclimation to stress, and production of proteins involved in post-stress recovery.

Besides its stabilising effects on cellular structure, proline is also capable of detoxifying free radicals by forming long-lived adducts with them (Floyd and Nagy, 1984; Smirnoff and Cumbes, 1989). This is a feature of most, but not all compatible solutes (Smirnoff and Cumbes, 1989). The extensive accumulation of active oxygen species during hyperosmotic stress and their contribution to damage induced by water deficit has recently been reviewed by Smirnoff (1993). The most likely explanation for the accumulation of active oxygen during water deficit is the concomitant reduction in photosynthetic rate (Kaiser, 1987; Chaves, 1991). Water deficit causes limitation to CO_2 uptake because of stomatal closure. Limitation in CO_2 fixation results in exposure of chloroplasts to excess excitation energy. This increases the rate of formation of active oxygen species (Smirnoff, 1993).

It has long been recognised that oxygen toxicity is a far more threatening challenge to plants than to most other aerobic organisms. Green plant tissues are particularly prone to oxygen toxicity effects since their chloroplasts have higher internal oxygen concentration than the surrounding atmosphere (Smirnoff, 1993). Furthermore, chloroplasts contain large amounts of polyunsaturated lipids in the thylakoid membranes and chlorophyll can absorb light and use this energy to form various oxygen species capable of causing damage (Smirnoff, 1993).

Oxygen radicals are among the most reactive chemical species known and are therefore potentially damaging to cells. In particular, they initiate lipid peroxidation, inactivate proteins and damage nucleic acids (Farr and Kogoma, 1991). In order to deal with the toxic and potentially lethal effects of active oxygen, plants like all aerobic organisms, have evolved a number of protective scavenging or antioxidant defence systems. These are both enzymatic or nonenzymatic. The most important enzymatic protective defences capable of removing, neutralising or scavenging free radical and oxy intermediates are the peroxidases, catalases and superoxide dismutases (Bowler *et al.*, 1992). In addition, the best known nonenzymatic defences include the low molecular weight nonprotein sulfhydryls glutathione,

cysteine and cysteinyl glycine, the vitamins ascorbate and α-tocopherol as well as carotenoids such as β-carotene. It is likely that crowding of the cytoplasm and chloroplastic stroma with these free radical scavengers might intercept active oxygen species before they can destroy cellular function.

The comparatively recent identification of proline as a free radical scavenger (Floyd and Nagy, 1984; Smirnoff and Cumbes, 1989) suggests that its accumulation might also contribute to the scavenging of active oxygen species. In support of this hypothesis, Alia *et al.* (1991) have demonstrated that proline enhances the photochemical electron transport activities of isolated thylakoid membranes by arresting photoinhibitory damage. The extent of the stimulation was greater in thylakoids of salt-stressed plants in comparison with unstressed plants (Alia *et al.*, 1991). Proline reduced lipid peroxidation during strong illumination. This suggests that its protective action was mediated via its ability to scavenge free radicals (Alia *et al.*, 1991).

In keeping with the findings of Alia *et al.* (1991), Van Rensburg *et al.* (1993) provided indirect evidence that proline levels in tobacco cultivars displaying different levels of drought tolerance were positively correlated with the membrane integrity of their chloroplasts. This effect may have arisen from either or both a direct protective effect of proline on the membranes or a reduction in free radical activity which would have prevented extensive lipid peroxidation.

Nevertheless, the functionality of proline in plants exposed to water deficit may have more aspects than those simply associated with its biophysical and biochemical properties. Several amino acids are excluded from protein surfaces to the same extent as proline (Arakawa and Timasheff, 1983, 1985) and are not toxic to enzyme activity at high concentrations (Yancey *et al.*, 1982). Furthermore, owing to their high reactivity, free radicals react with many compatible solutes other than proline (Smirnoff and Cumbes, 1989). Therefore, the arguments presented above cannot disprove the possibility that the protective action of proline on subcellular structure and its ability to scavenge free radicals is merely incidental and not of physiological importance.

Furthermore, the validity of many of these arguments rests on the presence of a sufficiently large pool of free proline. This is not always found in many glycophytes in which proline

accumulation has been reported. The nature of proline metabolism and its unique position in relation to the rest of intermediary metabolism are also likely to have acted in favour of proline as the preferred osmoticum in many plants.

2.1.2.4 Metabolic reasons for the suitability of proline accumulation as a stress response

In establishing why of all the amino acids, proline appears to have been selected as the choice osmoticum, it may be of value to compare its metabolism with that of other amino acids. In comparison with most other amino acids, proline has the metabolic advantage of being the terminal product of a relatively short and highly regulated pathway. Proline accumulation therefore affects fewer metabolic reactions than the buildup of multi-use substrates (e.g. glutamate) which are connected by several equilibrium enzymes to major metabolic intermediates. Furthermore, as pointed out by Phang (1985), proline and its immediate precursor Δ^1-pyrroline-5-carboxylate (P5C) are not interconverted by a single reversible enzyme, but by two distinct enzymes with different mechanisms and in different subcellular compartments (Section 2.2). Therefore, proline and P5C are not linked by a single equilibrium reaction and the final product of the proline biosynthetic pathway is not necessarily in equilibrium with its immediate precursor. Furthermore, because its α-nitrogen is a secondary amine, proline cannot participate in the transamination or decarboxylation reactions common to other amino acids (Phang, 1985). As outlined in Section 2.2, proline biosynthesis from both glutamate and ornithine occurs via the intermediate Δ^1-pyrroline-5-carboxylate (P5C). The strategic location of Δ^1-pyrroline-5-carboxylate (P5C) in intermediary metabolism suggests that it may mediate carbon transfer between the urea and tricarboxylic acid (TCA) cycles (Phang, 1985; LaRosa *et al.*, 1991).

In addition, the proline biosynthetic pathway from glutamate, although short, involves an extremely high rate of consumption of reductants (Section 2.2.1). Furthermore, proline degradation is capable of high energy output (Atkinson, 1977). These two features are likely to have contributed substantially to a role for proline in plants as a resource of value either in the acclimation to stress or in recovery upon relief from stress.

In this respect, several workers (Stewart and Voetberg, 1985; Bellinger and Larher, 1987; Kohl *et al.*, 1988; Venekamp, 1989; Shetty *et al.*, 1992) have proposed that selective preservation of the proline response in plants may relate to endpoints other than simply supplying proline. Indeed, it may be the metabolic processes of proline synthesis and degradation themselves that have been selected for by evolution as a means to tolerate stress. This line of thought is in keeping with what is known concerning the functionality of several other pathways involving amino acid metabolism. For example, the interconversions of aspartate and malate in the malate-aspartate shuttle play a critical role in transferring reducing equivalents from the cytoplasm into the mitochondria (Douce and Neuburger, 1990). In such cases it is the metabolic process itself which is of physiological value and not the products of the reaction. Therefore, the benefit of proline accumulation may derive more from its metabolism than from its contribution to the intracellular pool of free proline.

In this respect, it is interesting to note that certain workers (Blum and Ebercon, 1976; Itai and Paleg, 1982) have reported that the positive effect of proline accumulation is that it augments growth upon recovery from stress, rather than serving any direct function in enduring the stress itself. Bengston *et al.* (1978) suggested that proline may serve as a reserve substance for the synthesis of chlorophyll upon relief of stress. An adaptive role for proline acting as a sink for reduced nitrogen and carbon was suggested by Tully *et al.* (1979). Ahmad and Hellebust (1988) also proposed that proline may be a nitrogen storage compound during stress. Ericson and Alfinitos (1984) suggested that proline may be an important component of stress proteins. However, Stewart *et al.* (1977) found no evidence that proline levels affected rates of protein synthesis in leaves. While proline may be directly incorporated into protein during post-stress recovery, this does not account for why it should be accumulated during stress in preference to any of the other nineteen amino acids found in proteins.

A possible exception to this argument is the synthesis of proline- and hydroxyproline-rich cell wall glycoproteins. These have been collectively referred to as "extensin" - a somewhat misleading term since extensin actually promotes cell wall rigidity (Cassab and Varner, 1988). A restriction in cell expansion has been suggested to be associated with osmotic adjustments in plant cells (Van Volkenburgh and Boyer, 1985). Plant cell wall glycoproteins are believed to regulate cell expansion (Cassab and Varner, 1988). As in other structural

proteins such as collagen, hydroxyproline residues in extensin arise from hydroxylation of peptide-bound proline (Adams and Frank, 1980). It is thus possible that changes in cell wall proteins induced by osmotic stress might depend on changes in intracellular proline concentrations. However, Golan-Goldhirsh *et al.* (1990) found no significant effect of stress on proline and hydroxyproline contents of a purified cell wall fraction of sunflower. It appears as though changes in the physiochemical properties of the cell wall accompanying osmotic adjustment lie in other posttranslational modifications of extensin that are independent of the pool of free proline that accumulates during osmotic adjustment. Proline accumulated during hyperosmotic stress therefore does not appear to affect the synthesis of proteins important in stress tolerance.

It is more likely that proline constitutes a readily accessible source of energy or carbon after relief of stress or provides reducing power to support metabolic responses critical to recovery from hyperosmotic stress. Possible metabolic functions of proline biosynthesis and degradation in plants during and after relief from stress are represented diagrammatically in Figure 2.1.

A role for proline in recovery from stress is consistent with the observation that the extensive accumulation of proline in stressed tissues is usually followed by its rapid disappearance when the stress is removed. For example, in the alga *Chlorella emersonii*, a visible decrease in proline levels is evident within ten minutes after relief of hyperosmotic stress (Greenway and Setter, 1979). This is primarily by oxidation via P5C to glutamate and 2-oxoglutarate (Section 2.2.3). In this respect, proline differs from glycine betaine which is a metabolically inert end product (Hanson and Nelsen, 1978) and is maintained at stable levels long after relief from stress (Goas *et al.*, 1982; Naidu *et al.*, 1990).

Whereas proline biosynthesis is cytosolic, proline oxidation to 2-oxoglutarate occurs exclusively in the mitochondria (Figure 2.1; Section 2.2.3). The mitochondrial location of proline degradation suggests that this process may make a significant contribution to the TCA cycle (Figure 2.1). Blum and Ebercon (1976) appear to have been the first to provide convincing evidence of a correlation between levels of proline and the ability of a plant to recover upon relief of stress. In barley leaves recovering from drought, consumption of the proline accumulated in response to the stress takes approximately eight hours and can theoretically contribute carbon to the TCA cycle at a rate sufficient to account for 20% of

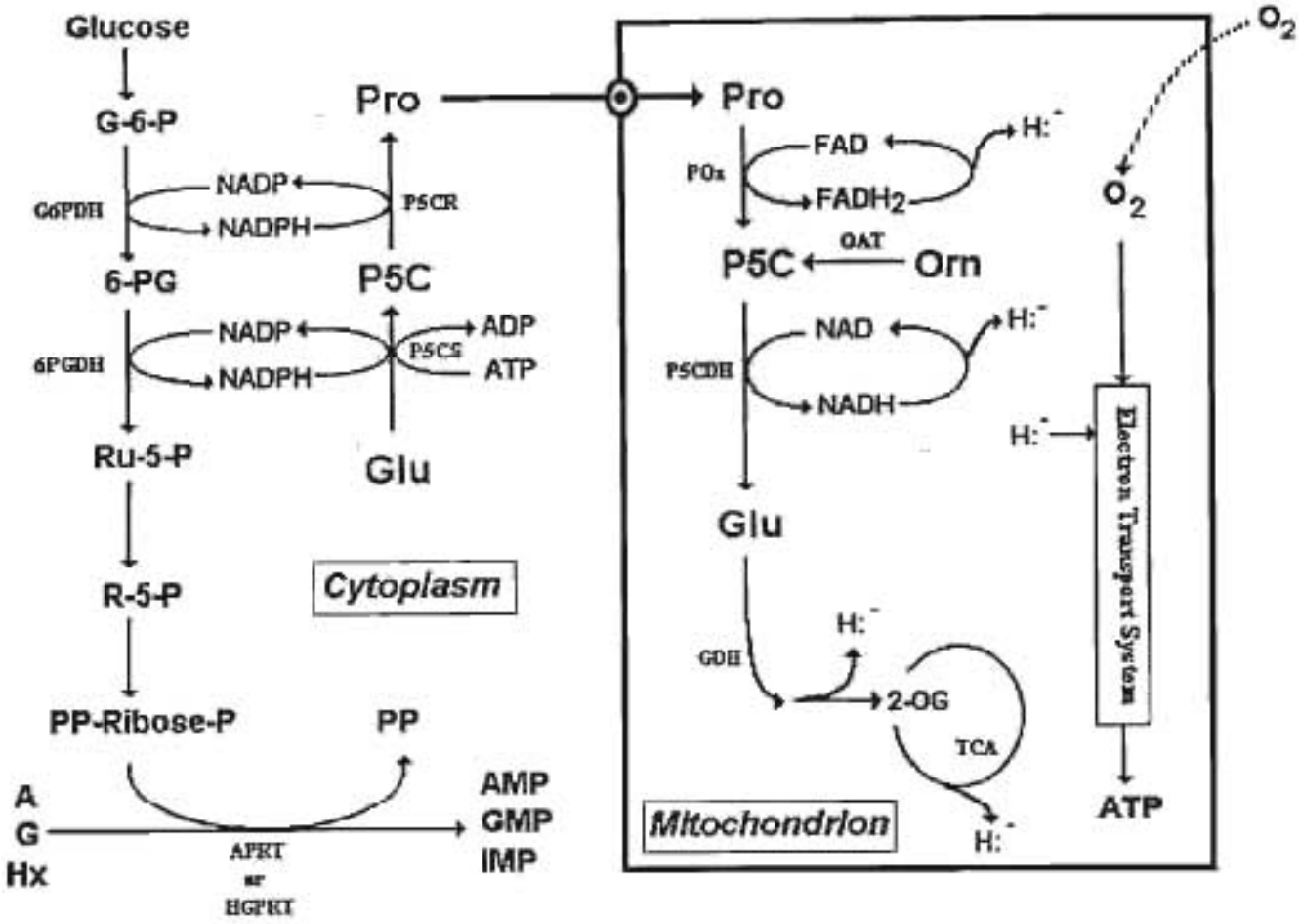


Figure 2.1 : Possible metabolic functions of proline biosynthesis and degradation in plants during and after relief from stress. Processes occurring in the mitochondrion are likely to be of importance only in recovery from stress. Abbreviations of intermediates are : A, adenine; AMP, adenosine monophosphate; G, guanine; G-6-P, glucose-6-phosphate; Glu, glutamate; GMP, guanosine monophosphate; Hx, hypoxanthine; IMP, inosine monophosphate; 2-OG, 2-oxoglutarate; Orn, ornithine; 6-PG, 6-phosphogluconate; P5C, Δ^1-pyrroline-5-carboxylate; PP-ribose-PP, phosphoribosylpyrophosphate; Pro, proline; R-5-P, ribose-5-phosphate; Ru-5-P, ribulose-5-phosphate. Abbreviations of enzymes are : AGPRT, adenine phosphoribosyltransferase; G6PGH, glucose-6-phosphate dehydrogenase; GDH, glutamate dehydrogenase; HGPRT, hypoxanthine, guanine phosphoribosyl transferase; OAT, ornithine δ-aminotransferase; 6PGDH, 6-phosphogluconate dehydrogenase; P5CRDH, P5C dehydrogenase; P5CR, P5C reductase; P5CS, P5C synthase; POx, proline oxidase. Also indicated is the tricarboxylic acid (TCA) cycle.

Reducing equivalents produced in the oxidation of Pro, P5C, Glu and 2-OG are represented by H:. The presence of a putative mitochondrial membrane-bound proline transport system (Cavalieri and Huang, 1980) is also indicated. The enzymes of the oxidative pentose phosphate pathway and P5CR are present in both the cytoplasm and plastids.

Adapted from Phang (1985) and Kohl *et al.* (1988).

the total respiratory activity (Stewart and Voetberg, 1985). The accumulation of proline appears to be an excellent means of storing energy since the oxidation of one molecule of proline can yield 30 ATP equivalents (Atkinson, 1977).

A number of observations in different biological systems support a role for proline as a primer for TCA cycle activity in plants recovering from osmotic stress. For example, proline is used as the primary mitochondrial fuel in energy-intensive processes such as insect flight (Weeda *et al.*, 1979, Gäde, 1992) and thermogenesis in the voodoo lily (Skubatz *et al.*, 1989). High levels of proline oxidation in the bacterioids of nitrogen-fixing root nodules of ureide-producing legumes suggest that proline is the primary energy source used by the bacterioids in fuelling the energy-intensive process of nitrogen fixation (Kohl *et al.*, 1988). Oxidation of proline is also believed to provide most of the energy required for pollen germination (Hong-qi *et al.*, 1982; Palfi and Palfi, 1982). Proline is the most abundant free amino acid in pollen grains and its amount in pollen grains is correlated with pollen viability. Additional evidence for a role for proline in priming oxidative respiration comes from studies involving animals deficient in proline oxidase activity. Both the PRO/Re mouse mutant (Blake, 1972) and sluggish-A mutant of *Drosophila* (Hayward *et al.*, 1993) are incapable of proline oxidation and exhibit markedly reduced mobility, presumably as a result of an impaired respiratory rate.

Besides contributing carbon to the TCA cycle, the mitochondrial degradation of proline to glutamate may also provide reducing equivalents needed to support mitochondrial electron transport and the generation of ATP for recovery from stress and for the repair of stress-induced damage (Figure 2.1). Nicotinamide nucleotides cannot cross the inner mitochondrial membrane, and consequently various shuttle mechanisms have been proposed for the transport of hydrogen from cytoplasmically generated NADH to the respiratory chain located inside the permeability barrier. A proline shuttle capable of such hydrogen transfer has been proposed in insect flight muscle (Balboni, 1978). This shuttle depends not only on the irreversibility of the reactions catalysed by Δ^1-pyrroline-5-carboxylate reductase (P5CR) and proline oxidase, but also on the cytosolic location of the former and the mitochondrial location of the latter (Figure 2.1). Interconversion of P5C and proline might therefore mediate transfer of redox potential between the cytosol and mitochondrion, and thereby stimulate oxidative respiration.

Proline synthesis in plants has also been implicated in regulation of cytosolic pH (Venekamp, 1989) and cellular redox potential (Bellinger and Larher, 1987; Kohl *et al.*, 1988, 1991). These hypotheses are based on the high demand for reductant in the proline biosynthetic pathway from glutamate. Although NADPH appears to be the preferred reductant in proline biosynthesis in plants, the proline biosynthetic enzymes appear capable of using its non-phosphorylated analogue NADH (Section 2.2.1). Depending on the availability of NADPH, NADH which is normally more ubiquitous in the cell (Stryer, 1988) may be the reductant used, especially under conditions of NADPH limitation.

As a result of the ability of proline biosynthetic enzymes to recognise both of the nicotinamide nucleotides, proline biosynthesis may directly impinge on the redox potential of the cell by altering the level of oxidation of the cellular pools of both NADH and NADPH. This most likely has widespread physiological consequences. The possible role of P5CR and other enzymes involved in P5C and proline metabolism in influencing the ratios of oxidised/reduced pyridine nucleotide cofactors in animal tissues has been reviewed by Phang (1985). Phang and co-workers have established an extensive hypothesis based on the assumption that proline and P5C constitute a redox couple and provide a mechanism for the intercompartmental and intercellular transfer of redox potential in animal cells. The transfer of redox potential alters the ratio of NAD(P)/NAD(P)H, thereby activating certain metabolic pathways dependent on the level of oxidation of the pyridine nucleotide pools. The rationale behind the extensive studies conducted by this group (Yeh and Phang, 1981; Mixson and Phang, 1988; Yeh and Phang, 1988; Merrill *et al.*, 1989) is the observation that in certain cancerous cell-types, regulation of P5C-metabolism often bears no apparent relationship to the accumulation of proline as the end product. These workers have proposed that in certain cell types, under certain physiological conditions or during particular developmental stages, the generation of NADPH during proline synthesis occurs at the expense of proline production. In this way, proline biosynthesis is primarily of importance in the regulation of carbohydrate metabolism and proline is merely a byproduct of the biosynthetic pathway (Phang, 1985).

Both NADH and NADPH are ancillary electron carriers for a wide variety of enzymes involved in metabolism (Stryer, 1988). The oxidation of NADH by the cytochromes of the electron transport chain constitutes the principal source of cellular ATP. The pyridine nucleotide cofactor NADH is thus a major currency in cellular energetics (Stryer, 1988). In

contrast, the phosphorylated form of NADH, namely NADPH, is important in providing reducing power for reductive biosyntheses (Stryer, 1988). An alteration in the ratio of the relative amounts of the reduced form of these coenzymes to their oxidised form is therefore likely to have widespread influences on cellular metabolism. In plants, NADPH is generated both by the oxidative pentose phosphate pathway and by the light reactions of photosynthesis (Stryer, 1988). Proline synthesis may be a means of regulating the level of reduction of the two pools of nicotinamide nucleotide coenzymes. Alteration of the NAD/NADH ratio impacts directly on cellular energetics, while the ratio NADP/NADPH is important in regulation of photosynthesis and reductive biosyntheses.

Under conditions of osmotic stress, the redox potential of plant cells is adversely affected in several ways. Activity of the mitochondrial electron transport system is decreased (Sells and Koepe, 1981; Schmitt and Dizengremel, 1989). This results in an accumulation of NADH and H^+, which is likely to inhibit several important metabolic reactions that require NAD and cause an increase in cellular acidity. Furthermore, inhibition of coupling of ATP synthesis to plastidial electron transport during water stress (Kaiser, 1987; Chaves, 1991) decreases the ratio of NADP/NADPH. This may have serious implications for regulation of many important biosynthetic reactions. Theoretically, proline biosynthesis is capable of oxidising reduced forms of both coenzymes and thereby maintaining a redox potential compatible with metabolism under normal conditions. In this way, proline accumulated under stress conditions may act as a sink for excess reductants by providing the NADP needed for the resumption of photosynthesis or increased activity of the oxidative pentose phosphate pathway (Figure 2.1). Alternatively, if NADH is used as an electron donor in proline biosynthesis, the NAD generated may drive oxidative respiration upon relief from stress.

In light of this role for proline in acting as a non-toxic store of excess reductant, it is interesting to note that most other common cellular osmolytes, including betaines, sucrose, hexitols and cyclitols are also highly reduced and might also serve a similar function during stress (Bellinger and Larher, 1987). Apparently, the production of most osmolytes under conditions of hyperosmotic stress may be an important mechanism of eliminating excess reductant, which may cause redox imbalance (Bellinger and Larher, 1987).

Proline biosynthesis from glutamate might be an adaptive mechanism to reduce the accumulation of NADH during osmotic stress. The decline in oxidative respiration during osmotic stress (Schmitt and Dizengremel, 1989) occurs at a time when even higher energy requirements are to be expected in order to deal with osmoregulation. Such adaptational processes might include ion pumping or the synthesis of compatible organic solutes. It therefore seems feasible to expect that the basal rate of oxidative respiration should be maintained, if not increased during hyperosmotic stress. Proline biosynthesis during hyperosmotic stress may be a means of activating the turnover of ATP, pyridine nucleotides and carbon skeletons in order to ensure continued respiratory activity. Proline biosynthesis may restore glucose catabolism, either via the glycolytic pathway and TCA cycle (when NADH serves as reductant in proline biosynthesis) or the oxidative pentose phosphate pathway (when NADPH is oxidised) and prevent extensive reduction of the pyridine nucleotide pools. Overall protein synthesis declines during water stress (Dungey and Davies, 1982a). Normally, protein synthesis consumes ATP, reduced pyridine nucleotides and carbon skeletons (Bellinger and Larher, 1987). Under conditions of water deficit, proline biosynthesis may substitute for protein synthesis in the turnover of ATP and the oxidation of the NAD(P)H pool.

Therefore, elevated levels of proline biosynthesis under conditions of hyperosmotic stress might maintain the NAD/NADH or NADP/NADPH ratios at values compatible with normal metabolism even in the presence of severe water deficit. The decline in levels of proline oxidation (Section 2.2.3) and P5C synthesis from ornithine (Section 2.2.2), both of which occur in the mitochondria is consistent with the need to prevent further reduction of the NAD pool in mitochondria. Proline biosynthesis within the chloroplast has been reported (Stewart, 1981). Activity of P5CR has been demonstrated in chloroplasts of tobacco (Noguchi *et al.*, 1966) and pea (Rayapati *et al.*, 1989). It is tempting to speculate that enhanced proline biosynthesis from glutamate in this subcellular compartment may provide NADP used in photophosphorylation. Alternatively, chloroplastic proline biosynthesis may be of importance in stimulating activity of the oxidative pentose phosphate pathway in the plastid. The entire sequence of oxidative pentose phosphate pathway enzymes are located in root plastids (Emes and Fowler, 1979). Furthermore, pentose phosphate pathway activity has been demonstrated in chloroplasts (Stitt and ap Rees, 1980) and plastids from cauliflower buds (Journet and Douce, 1985) and sycamore suspension cultures (Frehner *et*

The lower K_m of plant P5CR for NADPH in comparison with NADH (Section 2.2.1.2) suggests that reductant turnover by the proline biosynthetic pathway during stress may be more important in metabolic regulation than in maintaining oxidative respiration upon relief from stress. In most cellular systems, analysis of relative redox ratios reveals that the ratio NAD/NADH is relatively high, while the NADP/NADPH ratio is very low (Phang, 1985). This suggests that the NADP pool is mostly reduced (Phang, 1985). Therefore, a small change in NADP/NADPH ratios may have a large effect on flux through a redox-sensitive pathway dependent on NADPH. The oxidation of glucose-6-phosphate via the oxidative pentose phosphate pathway (Figure 2.1) is such a pathway.

It has long been considered that the ratio of carbon flux through glycolysis relative to the flux through the oxidative pentose phosphate pathway is an indicator of the physiological state of a tissue. A relatively high rate of carbon flux through the pentose phosphate pathway is considered to support a high rate of biosynthesis in an actively growing or differentiating tissue (Miernyk, 1990). Control of the partitioning of carbon through the two pathways is mediated primarily through regulation of the early reactions. If the activity of the glycolytic enzyme phosphofructokinase is inhibited, then more carbon is diverted through the oxidative pentose phosphate pathway (Miernyk, 1988). Conversely, inhibition of glucose-6-phosphate dehydrogenase and 6-phosphogluconate dehydrogenase increases carbon flux through glycolysis.

The enzyme glucose-6-phosphate dehydrogenase, which catalyses the rate-limiting step in the oxidative pentose phosphate pathway is not only dependent on NADP availability, but is also inhibited by NADPH (Miernyk, 1990). This extreme sensitivity to the level of oxidation of the coenzyme suggests that even a small rise in the NADP/NADPH ratio would open such a metabolic gate and thereby have important consequences for overall cellular metabolism. In particular, ribose-5-phosphate production would increase and become available as a substrate for nucleotide biosynthesis (Figure 2.1). Considering the *in vivo* concentrations of NAD/NADH and NADP/NADPH, the respective redox ratios of the two pyridine nucleotides and the preferential use of NADPH in proline biosynthesis, the conversion of P5C to proline is likely to affect the ratio of NADP/NADPH more than the NAD/NADH ratio (Phang, 1985).

In human erythrocytes, proline synthesis from glutamate is considered to be coupled to ribose-5-phosphate production. It has been shown that addition of P5C stimulates synthesis of phosphoribosyl pyrophosphate and therefore increases purine biosynthesis (Yeh *et al.*, 1981; Yeh and Phang, 1988). Based on these findings in an animal system, Kohl *et al.* (1988) postulated that extremely high levels of proline biosynthesis in nitrogen-fixing nodules of soybean function primarily to regenerate NADP needed to support the oxidative pentose phosphate pathway in the plant cytosol. High levels of activity of the pentose phosphate pathway are needed in root nodules of tropical legumes to supply ribose-5-phosphate for the synthesis of purines and ureides.

Although no work has been reported on the possible functional significance of the alteration of NADP/NADPH ratios by proline biosynthesis in response to water deprivation, it is possible that this may be of considerable importance. Proline biosynthesis may be important in increasing the NADP/NADPH ratio during stress to augment the synthesis of purine ribonucleotides by both salvage and *de novo* pathways upon the relief of stress (Figure 2.1). Although DNA biosynthesis is not affected by water stress, rates of RNA synthesis are enhanced in water stressed millet seedlings (Kandpal and Rao, 1985). Many of these transcripts are likely to encode products which are important in acclimation to stress. Alternatively, enhanced nucleotide synthesis might be essential to support the re-initiation of growth upon relief from stress. Generation of pentose phosphates is also essential for the synthesis of cell wall polymers (Miernyk, 1990).

A number of studies using both plant and animal systems have provided evidence that proline metabolism may be associated with enhanced rates of cell division. Proline stimulates cytokinin-induced shoot organogenesis in *Cucumis melo* (Shetty *et al.*, 1992). These workers proposed that the basis for this response was an increase in the supply of NADP for purine biosynthesis, which would be needed to maintain rapid cell division. The most important observation arising from this study was that thioproline (L-thiazolidine-4-carboxylic acid), a proline analogue known to inhibit proline oxidation (Elthon and Stewart, 1984), markedly reduced the extent of proline-mediated shoot formation. Thioproline has also been reported to stimulate proline synthesis in barley (Beffagna *et al.*, 1986) and *Vigna radiata* (Kumar and Sharma, 1989). This is consistent with an enhancing effect of thioproline on embryogenesis in *C. melo* (Shetty *et al.*, 1992). Since the process of proline synthesis, rather than the presence of proline *per se*, is important in enhancing activity of

the oxidative pentose phosphate pathway (Figure 2.1), these findings are in keeping with the proposal that proline metabolism is intimately linked with nucleotide biosynthesis.

Supplementation of tissue culture media with proline is also known to stimulate auxin-induced somatic embryogenesis in several plant systems (Nuti-Ronchi *et al.*, 1984; Armstrong and Green, 1985; Trigiano and Conger, 1987; Radojevic, 1988; Shetty and Asano, 1991; Radojevic and Subotic, 1992; Shetty and McKersie, 1993). Proline also stimulates elongation of alfalfa somatic embryos on a hormone-free medium (Stuart and Strickland, 1984). A role for proline metabolism in directing cell differentiation in certain mammalian cancer systems has also been proposed (Phang *et al.*, 1982). Proline is capable of reversing the inhibition of differentiation by L-azetidine-2-carboxylic acid (a proline analogue) of Leydig cells in the rat fetal testis (Jost *et al.*, 1988). Furthermore, thioproline is a component of plant biostimulants such as Ergostim (Beffagna *et al.*, 1986). This provides further evidence of the involvement of proline metabolism in the overall regulation of plant growth. It is also consistent with the observation that thioproline can reduce certain types of animal tumours (Brugarolas and Gosalvez, 1980; Grier *et al.*, 1983). In this context, it is also interesting to note that following injection of *Agrobacterium tumefaciens* Conn. into stems of tobacco and tomato, levels of proline in the resultant crown galls are 70 and 22 times higher than those found in normal stem tissues of the respective plants (Seitz and Hochster, 1964). While this may reflect a response to the stress associated with pathogen infection (Section 2.1.3), it is also likely to be associated with the rapid proliferation of cells within the crown gall tumours.

An oxidative component is a feature of virtually all stresses known to induce proline accumulation (Section 2.1.3). Defense mechanisms employed by plant cells exposed to oxidative stress have already been described briefly in Section 2.1.2.3. Protective mechanisms against active oxygen species, such as the peroxidase and glutathione system activities require reduced substrates and hence efficient endogenous mechanisms for the regeneration of these substrates. It is possible that proline biosynthesis may be important in turnover of the reduced and oxidised forms of NADP in order to provide such a source of reducing power. The enzymes of the Halliwell-Asada pathway involved in free radical scavenger metabolism are dependent on NADPH (Bowler *et al.*, 1992). In particular, thioredoxin reductase and glutathione reductase use NADPH as an electron source. Repair of protein damage induced by active oxygen by methionine sulphoxide reductase also

requires NADPH (Farr and Kogoma, 1991). Likewise, the reduction of the ascorbate free radical monodehydroascorbate back to ascorbate, catalysed by monodehydroascorbate reductase, is also dependent on the concomitant oxidation of NADPH (Dalton *et al.*, 1992).

In this respect, it is interesting to note that glucose-6-phosphate dehydrogenase is induced by oxidative stress in *Escherichia coli* (Greenberg and Demple, 1989; Kao and Hassan, 1985; Tsaneva and Weiss, 1990). This suggests that enhanced activity of the oxidative pentose phosphate pathway activity may be important in the development of tolerance to high levels of active oxygen species within the cell. Analogous studies in plants systems are limited. However, Argandona and Pahlich (1991) demonstrated a four-fold increase in activity of glucose-6-phosphate dehydrogenase in water stressed epidermal tissue and in primary leaves of barley. Activation of glucose-6-phosphate dehydrogenase has also been observed in pea root tips which were salt stressed (Porath and Poljakoff-Mayber, 1964). These workers reported that the stress treatment decreased glucose consumption via glycolysis. In light of the accepted functioning of the reaction catalysed by glucose-6-phosphate dehydrogenase as the rate-limiting reaction in the oxidative pentose phosphate pathway, this increased activity suggests that in plants, the oxidative pentose phosphate pathway may be activated by hyperosmotic stress. Enhanced proline biosynthesis during hyperosmotic stress may stimulate activity of the oxidative pentose phosphate pathway by generating NADP (Figure 2.1). This is likely to be of particular importance in chloroplasts, which are particularly prone to oxygen toxicity (Smirnoff, 1993; Section 2.1.2.3). As already described, both proline biosynthesis (Stewart, 1981) and pentose phosphate pathway activity (Stitt and ap Rees, 1980) have been reported in chloroplasts.

A decrease in intracellular pH has been implicated as one factor capable of eliciting proline accumulation (Pesci and Beffagna, 1985; Göring and Plescher, 1986; Chou *et al.*, 1991). Weak acids such as acetate and isobutyrate induce proline accumulation in leaf segments of barley seedlings, wheat seedlings and rice. However, there is no data indicating at what level this effect is mediated. It is likely that the increase in H^+ could stimulate NAD(P)H-mediated reduction of organic acids by dehydrogenases. Accumulation of proline could be the consequence of such reactions.

An increase in concentration of organic acids is a significant consequence of water stress in all plants, especially those capable of Crassulacean acid metabolism (Osmond, 1978;

Timpa *et al.*, 1986; Hubac *et al.*, 1986; Venekamp *et al.*, 1989). Citrate accumulates due to reduced availability of NAD resulting from the impairment of electron transport. Malate and lactate accumulation may be ascribed to the oxidation of NADH in their synthesis (Saradhi and Saradhi, 1991). Furthermore, the drop in ATP level accompanying a decline in respiration under water stress might contribute to an increase in acidity as a result of the inhibition of proton pumping at low ATP concentrations.

This increase in organic acid concentration results in a concurrent increase in the concentration of H^+ ions. Under normal conditions of water supply, these H^+ ions are exchanged against K^+ in the root system (Dijkshoorn *et al.*, 1968; Allen and Raven, 1987). Phosphoenolpyruvate carboxylase and malic enzyme have been implicated in the control of cytosolic pH under normal conditions of water supply (Smith and Brown, 1981; Davies, 1986). In the case of drought, water uptake is blocked, sometimes completely. As a result, this mechanism of K^+ exchange is reduced because ion uptake from the rooting medium is inhibited. The resulting decline in cytosolic pH could have disastrous consequences for several physiological processes. In particular, with only two exceptions, all the pH optima of the enzymes of the TCA cycle fall between pH 6.8 and pH 7.8 (Venekamp, 1989). Therefore, a drop in the cytosolic pH towards the acid range would probably diminish the activity of oxidative respiration. Maintenance of oxidative respiration depends on the adequate removal of excess H^+. This might be mediated by the accumulation of proline.

Under such conditions, Davies (1986) proposed that an important means of regulating cytosolic pH might be the conversion

$$X + NAD(P)H + H^+ \longrightarrow XH_2 + NAD(P)^+$$

,catalysed by a dehydrogenating enzyme. Removal of H^+ ions occurs in both formation and reduction of the intermediate P5C during proline biosynthesis from glutamate (Section 2.2.1). If the precursor glutamate is derived from 2-oxoglutarate via glutamate dehydrogenase, this involves another dehydrogenation (Layzell, 1990). This has led to the proposal that proline synthesis may be an efficient metabolic mechanism to counteract the increase in cytosolic acidity during stress (Venekamp, 1989).

Furthermore, in studies conducted with *Vicia faba* plants, Venekamp and Koot (1988) and Venekamp *et al.* (1989) demonstrated that the organic acids of oxidative respiration, namely citrate, malate and lactate, are the direct source of the carbon skeletons required for proline synthesis under conditions of water deficit. Therefore, proline accumulation appears to combat the problem of cytosolic acidity at a second level too, namely depletion of the organic acid pool itself.

Of course, continued removal of H^+ induces a concurrent formation of NAD(P), which needs to be reduced back to NAD(P)H in order for proline synthesis to continue. Normally, this would occur by means of the metabolism of glucose via glycolysis and the TCA cycle or via the oxidative pentose phosphate pathway. However, besides NAD(P)H, H^+ ions would be regenerated. Venekamp (1989) has suggested that under conditions of water deficit, generation of NAD(P)H might be possible without the concurrent production of H^+ by the mitochondrial oxidation of glycine. This yields NH_4^+, CO_2 and a methylene tetrahydrofolate derivative and involves the concurrent reduction of NAD(P) without liberation of H^+ (Tolbert, 1981). This is a general reaction occurring in the mitochondria of leaves in both C_3 and C_4 plants.

Therefore, several features of the metabolic reactions involved in proline biosynthesis and oxidation suggest that the metabolism of proline, rather than the accumulation of the free imino acid itself, is important in stress tolerance. Whereas a strong case has been presented for the regulatory role of P5C-proline interconversions in metabolism in animal systems (Phang, 1985), analogous investigations in plants are limited. The nodules of certain nitrogen-fixing legumes are characterised by extremely high rates of activity of proline metabolic pathways (Kohl *et al.*, 1988). Proline biosynthesis in root nodules is believed to generate NADP, and thereby stimulate activity of the oxidative pentose phosphate pathway. This generates phosphoribosyl pyrophosphate required for the synthesis of ureides. However, a comparison between ureide- and amide-exporting nodules revealed similar high P5CR activities, although only the ureide-forming type of nodule would require stimulated purine nucleotide synthesis (Kohl *et al.*, 1991). The only significant difference between ureide- and amide-exporting nodules was the considerably higher activities of proline oxidase in the bacteroids of the ureide-exporting species. However, the importance of this finding is not clear.

To date in stress physiology research, most work on the metabolic importance of proline biosynthesis from glutamate has centred on the possible role of the pathway in reducing acidity. Little attention has been paid to the metabolic regulatory implications of proline metabolism. Proline accumulation may be closely linked with the activity of oxidative respiration. Although proline may constitute a source of 2-oxoglutarate to serve either as substrate for the TCA cycle, this is only likely to occur upon relief from stress. During stress, it is more likely that proline synthesis is involved in redox transfer. Redox potential is likely to be severely altered by water deprivation owing to the concomitant decrease in electron transport in both mitochondria and chloroplasts. During the period of exposure to stress, proline synthesis may be a mechanism of preventing lethal over-reduction of the pyridine nucleotide pools and maintain NAD(P)/NAD(P)H ratios at values compatible with normal cell metabolism. Upon relief of stress, mitochondrial oxidation of the accumulated proline is likely to prime TCA cycle activity and thereby provide energy needed for recovery.

2.1.2.5 Arguments for and against proline accumulation as an adaptive response to hyperosmotic stress

Despite the multitude of possible roles proline may serve in plants during osmotic stress, none of the evidence currently available directly proves a function for proline in the process of acclimation to osmotic stress. It still remains necessary to establish that the apparent relationship between proline accumulation and stress tolerance is more than simply a correlation. Increases in proline levels in stressed tissues do not necessarily indicate an adaptation to stress. Rather, proline accumulation may simply result from a disruption in cellular metabolism as a result of severe water stress.

Whereas a large body of evidence supports a positive correlation between proline accumulation and adaptation to osmotic stress (Singh *et al.*, 1972; Bar-Nun and Poljakoff-Mayber, 1977; Handa *et al.*, 1986; Le Dily *et al.*, 1991; Voetberg and Sharp, 1991), this has not been corroborated by other workers (Hanson *et al.*, 1977, 1979; Ferreira *et al.*, 1979; Tal *et al.*, 1979; Hassan and Wilkins, 1988; Ashraf, 1989). Results of certain workers (Chandler and Thorpe, 1987a, 1987b; Rodriguez and Heyser, 1988) have even indicated proline to be negatively correlated with osmoregulation in stressed plant tissues.

Although a number of desirable physiochemical properties are exhibited by proline at molar concentrations (Schobert and Tschesche, 1978), these may not play a decisive role when proline is found at the relatively modest concentrations generally observed in plant cells adapted to hyperosmotic conditions. Schobert and Tschesche (1978) reported that proline concentrations above 5 M were capable of enhancing the solubility of insulin by approximately 170-fold. However, at concentrations below this, proline had only negligible effects. Furthermore, many studies which have implicated a protective role for proline on subcellular structures (Pollard and Wyn Jones, 1979; Paleg *et al.*, 1981, 1984; Arakawa and Timasheff, 1983, 1985; Rudolf *et al.*, 1986; Schwab and Gaff, 1990) have, of necessity, been performed *in vitro*, with enzymes and membrane systems removed from their natural micro-environment and in the case of proteins, usually at low non-physiological levels.

However, the argument that proline levels in stressed plants may be inadequate to protect subcellular structure assumes that the imino acid is not compartmentalised within the cell. Some evidence suggests that compartmentation of proline within the cell may occur. Several workers (Stewart and Lee, 1974; Hasson and Poljakoff-Mayber, 1983; Ketchum *et al.*, 1991) have reported that proline accumulation in vacuolated plant cells in response to hyperosmotic stress is primarily restricted to the cytoplasm and cytoplasmic organelles. This might ensure relatively high localised concentrations of cellular proline. The cytoplasm constitutes only 5-10% of the osmotic volume (Briggs and Robertson, 1957) of vacuolated cells (Flowers and Yeo, 1986). Therefore, relatively small amounts of solute in the cytoplasm may account for cytosolic osmotic adjustment to low external water potential. Consequently, localised concentrations of proline within the cell may be sufficient in certain cases to account for the protective action of proline on protein-structure as proposed by Schobert and Tschesche (1978). Furthermore, the actual concentration of proline within the cytoplasm is likely to be higher than what is apparent because of the reduction in cell volume that accompanies dehydration (Turner and Jones, 1981).

Nevertheless, Pahlich *et al.* (1983) demonstrated that in polyethylene glycol-treated *Nicotiana rustica* protoplasts, most proline is sequestered in the vacuole. This implies that increases in proline accumulation in this species probably has no pronounced osmotic or biophysical influence in the cytoplasm, but the value of proline accumulation may be mediated via its metabolism (Section 2.1.2.4). This could be mediated at either or both the level of proline biosynthesis during stress, or by proline degradation upon relief from stress.

In contrast, most of the accumulated proline effective in osmoregulation of cultured cells of the salt marsh grass *Distichlis spicata* was found in the cytoplasm (Ketchum *et al.*, 1991). Cells maintained a cytoplasmic proline concentration at least one order of magnitude greater than that of the vacuole. This would be important in maintenance of osmotic balance between the vacuole and cytosol, especially in halophytes, which absorb much NaCl and store it in the vacuole. It is generally accepted that osmotic adjustment with organic compatible solutes in the cytoplasm balances inorganic ion storage in the vacuole (Stewart and Lee, 1974). Confinement of the compatible solute pinitol to the cytoplasm when *Mesembryanthemum crystallinum* is salt stressed has also been shown. No pinitol accumulated in the vacuole (Paul and Cockburn, 1989). Similarly, Hall *et al.* (1978) demonstrated that glycine betaine accumulated by *Suaeda maritima* was restricted exclusively to the cytoplasm. Therefore, vacuolar accumulation of proline in osmotically-stressed *Nicotiana rustica* may be a feature unique to this species.

Bellinger and Larher (1987) have argued that frequent use of the term "accumulation" in relation to increases in proline levels is misleading, since it has even been used to describe an increase of less than 100% in an initially low proline pool (approximately 0.1 mM, Pesci and Beffagna, 1984). Another argument against functionality for proline accumulation is that this response is not a feature of the response of all higher plants to hyperosmotic stress. For example, increased concentrations of chloride or sulphate salts could not effectively stimulate proline accumulation in sugarcane leaves of a salt-sensitive variety (Naik and Joshi, 1983). Similarly, pigeon pea plants fail to accumulate proline at high salinity levels (Joshi, 1984). In *Andropogon glomeratus*, a C_4 nonhalophytic salt marsh grass, proline probably plays no role in osmotic adjustment, since very high levels of salinity are required to increase its concentrations (Bowman, 1988). Cucumber, which is particularly sensitive to water deprivation, does not accumulate proline as a result of water stress. Even exogenous application of proline, which has shown to be taken up by the tissue, does not increase stress tolerance of cucumber (Itai and Paleg, 1982).

Contradictory reports for the stabilising effect of proline on enzymes have been made. Schwab and Gaff (1990) reported a protective role for proline on the activities of NAD-malate dehydrogenase, glucose-6-phosphate dehydrogenase, NADP-isocitrate dehydrogenase and glyceraldehyde phosphate dehydrogenase from two grasses, one dessication-tolerant and the other dessication-sensitive. Almost full protection of enzyme activities was obtained

when proline was added at a molar ratio of 2:1 (protectant to salt) with the simultaneous addition of salt to the reaction media. No protection was found when proline was added after a 1 h preincubation of enzyme extracts with high salt concentrations (above 0.5 M). Extracts from air-dry leaves of the dessication tolerant species *Sporobolus stapfianus* recovered almost fully after 5 h incubation in 1 M proline (Schwab and Gaff, 1990). Although Manetas *et al.* (1986) also reported full protection of phosphoenolpyruvate carboxylase against NaCl inhibition in two species of Poaceae with a proline concentration between 200-800 mM, proline acted as a competitive inhibitor of the same enzyme from members of the Chenopodiaceae (Manetas *et al.*, 1986).

Such issues have led to much debate as to the exact function of proline, if any, in hyperosmotic stress. In particular, controversy continues as to whether proline is a cause or the consequence of metabolic adaptation to stress conditions. For example, it might be argued that proline biosynthesis may simply be extremely sensitive to elevated levels of reductants and organic acids that accumulate during osmotic stress. Alternatively, if proline is a redox shuttle molecule (Section 2.1.2.4), as has been established in certain mammalian tissues (Phang, 1985) and proposed in nitrogen fixing nodules (Kohl *et al.*, 1988), uncoupling of the redox shuttle mechanism might cause proline accumulation. A decline in proline oxidation concomitant with water stress (Stewart *et al.*, 1977; Section 2.2.3) might cause such uncoupling. Increased proline synthesis in response to osmotic stress would exacerbate this situation. Proline accumulation may then merely be a symptom of metabolic dysfunction. The biocompatible features of proline and its apparent lack of toxicity to cells suggests that such accumulation might be tolerated until homeostasis is regained.

Attempts to correlate the ability of different varieties of the same species to accumulate proline with their drought resistance ratings have met with different results. Positive correlations between capacity for proline accumulation and drought tolerance have been found in ten barley cultivars (Singh *et al.*, 1972), two maize cultivars (O'Regan *et al.*, 1993) and four tobacco cultivars (Van Rensburg *et al.*, 1993). The last study reported that drought tolerant cultivars were able to accumulate proline both earlier and to much higher end concentrations. In twelve rice varieties, the salinity index of yield showed a strong positive correlation with proline accumulation (Pandey and Srivastava, 1990).

However, Hanson *et al.* (1977) have challenged the value of proline-accumulating potential as a positive index for drought resistance in different strains of barley. Furthermore, Richards and Thurling (1979) observed only a weak correlation between yield under drought and proline accumulation in *Brassica napus*. Proline accumulation appears to be a poor indicator of salt tolerance in black gram (Ashraf, 1989) and soybean (Moftah and Michel, 1987). The latter study found proline levels to be inversely correlated with the tolerance of soybean to salinity stress. In black gram, a negative correlation of proline accumulation with salt tolerance was observed (Ashraf, 1989). Levy *et al.* (1988) found no relation between proline content in leaves or tubers of potatoes and their relevant tolerance or susceptibility to salinity. Both Blum and Ebercon (1976) and Bhaskaran *et al.* (1985) respectively failed to correlate drought tolerance with the capacity of eight and ten cultivars of sorghum to accumulate proline.

Furthermore, although stress-induced proline accumulation is usually rapid, in certain species it begins only when cell injury is already evident (Hanson *et al.*, 1977; Moftah and Michel, 1987). Moreover, in certain species such as *Brassica napus*, elevated levels of proline may persist for up to one month after stressed cells have been returned to normal osmotic conditions (Chandler and Thorpe, 1987b). This suggests that it is unlikely that proline makes any contribution to recovery from stress in callus cultures of *Brassica napus*.

Observations such as these have resulted in the establishment of a school of thought supporting the conclusion of Hanson *et al.* (1979) that proline accumulation is merely a symptom of profound metabolic disturbance induced by a reduction in water potential. To adherents of this philosophy, the sensitivity of proline biosynthesis to expanded pools of organic acids or reductants following hyperosmotic stress does not provide adequate evidence that the process itself is of any functional importance in acclimation to osmotic stress.

However, Delauney and Verma (1993) point out that the absence of a positive correlation between proline accumulation and osmotolerance in some species does not negate an adaptive role for proline *per se*. Rather, it may reflect the predominance in these species of osmoregulatory mechanisms other than osmotic adjustment. Many alternative responses to hyperosmotic stress other than proline accumulation are possible. The choice of a response is likely to be dependent on the carbon and nitrogen allocation patterns found to be optimal

by the species in question. It is also likely to be influenced by plant size and total surface area as well as by stomatal physiology. Proline accumulation is dependent on the availability of sufficient carbohydrates (Hsiao, 1973). Differences between species or even varieties in their capacities for starch mobilisation may also account for differences in their ability to accumulate proline.

The mechanism of drought stress tolerance in many plant species is through avoidance of the stress. Such adaptations are often at the whole plant level. They may be morphological, for example increased leaf waxiness, leaf pubescence or development of a more extensive and deeper rooting system (Malik *et al.*, 1979; Sharp and Davies, 1985). Alternatively, adaptations to stress tolerance are often developmental. For example, the time of flowering of a species is often a determinant of stress avoidance. A shorter life cycle may enable the plant to mature safely during a rainfall period. Annual plants characteristically spend the driest periods of the year in a dormant state. Yet other adaptations to hyperosmotic stress may be physiological. These might include active exclusion of salt or sequestration of ions within the vacuole.

Alternatively, poor proline accumulators may have evolved metabolic responses other than proline accumulation. A number of changes in the protein and mRNA populations of plant cells have been observed in response to hyperosmotic stress (Singh *et al.*, 1985; Winicov *et al.*, 1989). Despite its apparent eminence, it would be a gross oversimplification to examine proline accumulation in isolation, when examining plant stress responses. One such metabolic adaptation, common to many drought tolerant species, might be a shift in photosynthesis from the C_3 photosynthetic pathway to Crassulacean acid metabolism (Cushman *et al.*, 1992). Another might be the accumulation of alternative osmolytes. It is well known that the extent to which any of the common intracellular solutes play a role in osmotic adjustment varies between different plant species. Different osmolyte preferences also exist among species of bacteria (Csonka, 1989, Madkour *et al.*, 1990).

Although the pathways of proline biosynthesis and degradation are obviously present in all cells, several workers (Itai and Paleg, 1981; Naik and Joshi, 1983; Joshi, 1984; Bowman, 1988) have reported proline accumulation to be minimal in certain plants exposed to osmotic stress. However, in species where proline is not accumulated to substantial levels, other osmolytes tend to accumulate. Presumably this is to cope with the osmotic load by

compensating for a low level of proline. In most species studied, a reciprocal relationship is evident between the levels of free proline and the preferred osmoprotectant. In species where both proline and glycine betaine accumulate, for example *Aster tripolium*, the metabolic pathways involved in accumulation of both compatible solutes appear to be highly coordinated (Goas *et al.*, 1982).

Many glycophytes modulate cellular osmotic balance primarily by using low molecular weight carbohydrates (Greenway and Munns, 1980; Weimberg *et al.*, 1982). In these plants, proline is most probably only of secondary importance. Some plants are almost exclusively sorbitol or mannitol accumulators. Such species include members of the Plantaginaceae (Ahmad *et al.*, 1979), brown algae (Reed *et al.*, 1985) and *Apium graveolens* (Larher, 1987). They do not accumulate proline. Furthermore, euhalophytes are typically not proline accumulators. Members of the Chenopodiaceae, which exhibit among the highest salt tolerances reported, accumulate glycine betaine in preference to proline (Coughlan and Wyn Jones, 1980). The stabilising effect of glycine betaine on enzymes isolated from members of the Chenopodiaceae, but inhibitory effect of proline on the same enzyme activities has led to the proposal that enzymes from certain species may have coevolved with the osmolyte selected for by that species (Manetas *et al.*, 1986; Nikopoulos and Manetas, 1991). This may explain the strong taxonomic preference towards a particular osmolyte in certain plant families.

Proline accumulation is not believed to play a major role in the osmoregulation of most halophytes (Storey and Wyn Jones, 1977; Doddema *et al.*, 1986; Demmig and Winter, 1986). Plants growing in saline environments usually accumulate large amounts of NaCl in the tissue. Owing to the inhibitory effect of sodium and chloride ions on many enzymes, their presence in the cytoplasm should be minimal. Evidence for the compartmentation of electrolytes between the cytosol and vacuole in halophytes has been presented by Stewart and Lee (1974) and Ketchum *et al.* (1991). The necessary osmotic balance between the two compartments is achieved through accumulation of organic solutes in the cytoplasm. Besides playing an osmotic role, compatible solutes are likely to protect enzymes against denaturation or inhibition of activity.

However, some proline accumulation in halophytes is frequently observed as an initial response to an osmotic stress imposed very rapidly or to levels much higher than those

promoting normal growth (Storey and Wyn Jones, 1977; Doddema *et al.*, 1986). However, the levels of proline accumulated are unlikely to be osmotically effective. Presumably this provides time to activate more permanent adaptive responses. Furthermore, exceptions to this generalisation for halophytes do exist. For example, the halophyte *Triglochin maritima* is a typical proline accumulator and does not produce a significant amount of glycine betaine (Stewart and Lee, 1974).

Instead of detracting from evidence for an osmoregulatory role for proline, these findings provide strong support for functionality of proline in plants that are proline accumulators. Species which have selected for other osmolytes, probably on the basis of differences in their carbon and nitrogen allocation patterns, have apparently dispensed of this otherwise metabolically expensive pathway. This provides evidence against the proposal originally presented by Hanson *et al.* (1979) that proline accumulation is nothing more than a symptom of metabolic modifications induced by stress.

Direct regulation of proline accumulation by other osmoprotectants has recently been reported by Larher *et al.* (1993). These workers showed that in leaf discs of *Brassica napus*, proline accumulation induced by mannitol, polyethylene glycol, sodium chloride or sodium nitrate could be inhibited by glycine betaine and two other betaines, pipecolic acid and dimethylsulphoniopropanoic acid. Although it was not established whether the inhibition operated at the level of restoring feedback inhibition of proline biosynthesis or enhancing degradation, this finding suggests the existence of a finely-controlled mechanism of attenuating stress-induced proline accumulation in the presence of sufficient amounts of an osmolyte of equal, if not greater efficacy. Such regulation suggests functionality for proline in counteracting the effects of hyperosmotic stress.

Drought tolerance in plants is almost certainly dependent on a complex of often unrelated properties. Failure to correlate drought tolerance with only one of several possible adaptations does not necessarily implicate a lack of function for the response in question. Considering the complexity of drought tolerance mechanisms, it is perhaps not surprising that in many species, no straightforward relationship between proline accumulation and drought tolerance is evident. Restrictions in plant growth cannot be attributed to a single process. Instead, plant growth is the result of many integrated and regulated physiological and biochemical processes.

Although much evidence for functionality of proline accumulation in plant systems is indirect, more compelling evidence in favour of a direct role for proline in counteracting the effects of hyperosmotic stress comes from studies in which application of exogenous proline has been shown to have a protective effect on bacteria, as well as whole plants, individual plant tissues and isolated plant organs.

The osmoprotective properties of proline were first reported in a prokaryotic system when Christian (1955) observed that exogenous proline could alleviate the growth inhibitory effects of osmotic stress in *Salmonella oranienburg*. A wide variety of osmotically-stressed bacteria have been shown to accumulate proline (Measures, 1975) and addition of proline to hyperosmotic media stimulates the growth rate of many bacterial species (Kieft and Spence, 1988; Gloux and Le Rudulier, 1989). Proline-overproducing mutants of *Escherichia coli* (Dandekar and Uratsu, 1988) and *Serratia marcescens* (Sugiura and Kisumi, 1985) clearly exhibit increased osmotolerance. Transfer of a plasmid-borne gene encoding a feedback-resistant γ-glutamyl kinase from *Salmonella typhimurium* (Csonka, 1981; Mahan and Csonka, 1983) to *E. coli* and *Klebsiella pneumoniae* confers proline overproduction and associated osmotolerance (Jakowec *et al.*, 1985).

Furthermore, some evidence in favour of a functional role of proline as a compatible solute has been obtained by exogenous application of proline to whole plants or to isolated plant tissues. Addition of 100 mM proline to a Hoagland solution containing 120 mM NaCl neutralised the effect of salinity stress on peas and *Tamarix tetragyna* (Bar-Nun and Poljakoff-Mayber, 1977). In the same study, proline was also found to counteract the NaCl-induced inhibition of pea seed germination. Handa *et al.* (1986) demonstrated that addition of exogenous proline to culture medium during water stress and osmotic downshock alleviated the normally resulting inhibition of growth of tomato cell suspension cultures. Such a protective role was not observed for any of the other amino acids tested (Handa *et al.*, 1986). Addition of 10 mM proline to cultured barley embryos increased shoot elongation under saline conditions (Lone *et al.*, 1987). This effect was attributed to the ability of proline to decrease the leaf salt load. Callus lines of *Cicer arietinum* grown in a medium containing 100 mM NaCl and 10 mM proline increased their fresh and dry weights (Pandey and Ganapathy, 1985).

These data implicate a direct role for proline itself, and do not implicate the process of proline biosynthesis as being of importance in stress tolerance. The process of proline synthesis, as opposed to the presence of proline is important in coupling the oxidative pentose phosphate pathway to regulate purine metabolism and nucleotide turnover (Figure 2.1). In contrast, Itai and Paleg (1982) reported beneficial effects of exogenous proline during recovery of barley plants from water stress. However, no effect on growth during stress was noted (Itai and Paleg, 1982). This is consistent with the hypothesis that proline confers its adaptive advantage via its metabolism upon relief from stress (Section 2.1.2.4).

Nevertheless, as with much of the work done concerning the beneficial effects of proline in stressed plants, contradictory findings have also been reported in which exogenous proline does not confer any advantage on stressed plants. For example, the presence of 1 mM or 10 mM proline in media containing 100 mM or 200 mM NaCl had little effect on the growth of salt-adapted callus of rice (Kishor, 1989). However, the same study indicated that some concentrations of proline significantly increased the growth of salt-unadapted rice callus. Proline (10 mM) inhibited the growth of salt grass suspension cultures in the presence of 260 mM NaCl (Rodriguez and Heyser, 1988). Nevertheless, it is worth noting that in this study, exogenous [^{13}C]-proline inhibited the normal biosynthesis of proline that would have occurred in suspensions grown at this salinity level. This may account for this result.The findings of Rodriguez and Heyser (1988) bolster the argument presented in Section 2.1.2.4 that the synthesis of proline and not merely its presence, is of importance in counteracting the effects of hyperosmotic stress.

Another approach to resolving the controversy surrounding the role of proline in adaptation to stress has been the isolation of mutants which over-produce proline under non-stressed conditions. Cell lines of *Daucus carota* with a twelve times increase in proline content demonstrate an enhanced resistance to high salinity (Riccardi *et al.*, 1983). A barley mutant resistant to hydroxyproline accumulated proline in the soluble fraction of the leaf to three times the normal amount (Kueh and Bright, 1982). The growth of this mutant was less inhibited by concentrations of NaCl below 100 mM than wild-type plants and also under conditions in which NaCl concentration was raised to 200 mM over a period of three days. However, at constant concentrations of NaCl above 100 mM and in polyethylene glycol (0-40%), the growth of the mutant and wild-type plants was equally inhibited (Kueh and Bright, 1982). This implied that a three-fold increase of proline content in the mutant did

not confer an increased ability to grow under water stress.

More convincing evidence comes from studies of barley mutants lacking ferridoxin-dependent glutamate synthase, and possessing low levels of glutamate, the primary precursor for proline under conditions of stress (Section 2.2). When wild-type barley plants were subject to a gradual increase in water deficit by soil drying, there was a rapid accumulation of proline (Al-Sulaiti *et al.*, 1990). However, the glutamate synthase deficient mutant lacked the ability to accumulate proline and showed premature symptoms of stress. Furthermore, unlike the wild-type plants, the glutamate synthase deficient mutants exhibited a decrease in root to shoot ratio following the soil drying treatment (Al-Sulaiti *et al.*, 1990).

Three lines of evidence intuitively suggest a functionality for proline in osmotic stress. Firstly, as pointed out by Handa *et al.* (1986) the accumulation of proline in a diverse array of organisms spanning the biological kingdoms is unlikely to be coincidental. Furthermore, proline concentrations are directly proportional to the salinity level or the intensity of water stress. Also, besides the rapidity of accumulation following the onset of stress, proline is usually lost rapidly upon restoring plants to optimal growth conditions, principally by oxidation (Paleg and Aspinall, 1981). This reinforces the idea that proline accumulation under stressful conditions has some functional significance. Many other organic solutes which are accumulated during stress (e.g. glycine betaine) do not show a similar decline (Goas *et al.*, 1982; Naidu *et al.*, 1990). Asparagine synthesis has been reported to increase upon relief from stress (Venekamp and Koot, 1988).

2.1.2.6 Conclusion

Proline accumulation in response to hyperosmotic stress is a highly complex and still incompletely understood process. Although presently there is no unified model available for a comprehensive view of the functional significance of proline during hyperosmotic stress, it is most likely that the effects of an accumulation of free proline are multifaceted. Proline accumulation may well have been strongly selected for as a response to hyperosmotic stress because it fortuitously serves several adaptive functions other than simply adjusting osmotic potential.

In particular, proline may act as a free-radical scavenger (Smirnoff and Cumbes, 1989), protect subcellular structure by an interaction with enzymes (Schobert, 1977b; Arakawa and Timasheff, 1985), membranes (Rudolf *et al.*, 1986) or ribosomes (Kandpal and Rao, 1985) or serve as a carbon or nitrogen source for recovery upon relief from hyperosmotic stress (Dashek and Erickson, 1981). During stress, proline biosynthesis may assume a primary role in reducing cytosolic acidity following a breakdown of other mechanisms capable of regulating cytosolic pH (Venekamp, 1989). To date, comparatively little attention has been devoted to the possibility that proline synthesised during stress may act as a non-lethal store of reductant for use in repair of stress-induced damage and the resumption of growth upon return to normal conditions. Coupling of proline synthesis to the oxidative pentose phosphate pathway may be important in increasing ribose-5-phosphate production for the synthesis of nucleotides needed for enhanced transcription of genes which encode products important for stress tolerance. Nucleotide synthesis may also be required for the initiation of cell division upon relief from stress. Alternatively, pentose phosphates produced by the oxidative pentose phosphate pathway may be of importance in the synthesis of cell wall polymers during plant growth following relief from stress.

However, the debate whether proline is a symptom of drought susceptibility or a response with possible adaptive significance has yet to be resolved. Considering the importance of understanding and increasing plant resistance to drought, the controversy regarding the adaptive significance of proline is likely to continue until it is possible to distinguish conclusively between what may be a response due to altered metabolic balance and what is of adaptive significance. Appreciation of the functional significance of proline accumulation in response to hyperosmotic stress is not trivial. It is of vital importance to make a distinction between what is an adaptive response and what is merely an incidental consequence of stress before implementing plant-improvement programs aimed at enhancing stress resistance.

The abundance of literature accumulated over the past forty years concerning the physiology of proline accumulation in plants and its possible significance, is rife with contradiction. The conclusions drawn by different workers are often in direct dispute. This most probably arises from the range of possible adaptations that plants may choose in dealing with osmotic stress (Delauney and Verma, 1993). Proline accumulation is only one of a number of possible measures which plants may take to ameliorate the negative effects of hyperosmotic stress.

The accumulation of proline in cultivars is not always correlated with their stress tolerance. Much of the lack in reproducibility of results obtained from the same species probably arises from a failure to use standardised methods and conditions. For example, recent work (Naidu *et al.*, 1992) has reconciled the conflicting results of Singh *et al.* (1972) and Hanson *et al.* (1977) concerning proline accumulation in barley cultivars. Apparently, these two groups of workers used plants grown and stressed under different vapour pressure deficits, which accounts for their contradictory results (Naidu *et al.*, 1992).

Increasingly, evidence is emerging that proline production and metabolism is extremely sensitive to the growth environment and that the previous environmental history of the plant has a large impact on its response to any change in the environment. The extent of proline accumulation is affected by the rate of imposition of stress (Naidu *et al.*, 1990), stress preconditioning, organ type and age and genetic variation within and between species (Cramer *et al.*, 1990). Variations in the level of inadvertent acclimation to water deprivation prior to imposition of experimental stress conditions are also likely to account for much of the variability in data obtained.

Many workers who have reported a negative correlation between proline accumulation and stress tolerance have used cell culture systems (Bhaskaran *et al.*, 1985; Chandler and Thorpe, 1987a, 1987b; Hassan and Wilkins, 1988; Rodriguez and Heyser, 1988). The ability to extrapolate such results directly back to a whole plant system is questionable. Furthermore, experimental work with certain plants has involved wounding as a necessary step, especially for precursor uptake, protoplast isolation or simply handling of parts of larger plants. Wang *et al.* (1982) have demonstrated that proline accumulates in excised rice leaves. This suggests that such experimental approaches may introduce metabolic artifacts unrelated to osmotically-induced proline accumulation and thereby influence the outcome of experiments.

Furthermore, in many experiments, water stress is produced by incubation in polyols or sucrose solutions (osmotic stress) or by treatment with high concentrations of NaCl (salt stress). In the former case, it has been shown that absorbable osmotic agents partly contribute to proline accumulation via a metabolic effect unrelated to osmotic stress (Cress and Johnson, 1987). In the latter case, the general effects of water deficiency may be superimposed by specific effects of ion imbalances. Although salinity stress constitutes an

osmotic stress, it differs from water deprivation stress. This is reflected in the kinetics of proline accumulation and degradation (Voetberg and Stewart, 1984). In barley, proline is degraded rapidly as a result of recovery from water stress, but not as an immediate response to salt removal from media (Voetberg and Stewart, 1984). Stewart and Voetberg (1985) demonstrated a correlation between abscisic acid (ABA) and proline accumulations in water-stressed barley. When ABA levels exceeded a threshold level, they induced proline accumulation. However, in salt-treated plants, turgor was maintained and proline accumulation was not dependent on precursory ABA accumulation. This suggests that the mechanisms that initiate proline accumulation in response to water stress and salt stress may be different.

Proline accumulation is a complex function depending not only on the rate of synthesis, but also on the rate of utilisation of the imino acid in protein synthesis, rate of proline oxidation and of proline release via protein turnover. Differences between genotypes may derive from any of these processes.

However, the sustained interest in proline accumulation as a response to hyperosmotic stress suggests an overall consensus that the response of proline accumulation must be of some physiological importance. Failure of certain species to accumulate proline or demonstrate increased stress tolerance upon treatment with proline does not necessarily withdraw from a possible significance of proline accumulation in other species (Delauney and Verma, 1993).

The comparatively recent emergence of molecular techniques provides an effective means of analysing the importance of physiological events. It is likely that use of such an approach will indicate the relative significance of proline accumulation as a response to hyperosmotic stress. Understanding how proline accumulation is regulated at the biochemical and molecular level is likely to be of great value in defining precisely the significance of proline accumulation in plants exposed to hyperosmotic stress. In order to do this, the genes involved in proline accumulation need to be fully characterised. This may enable the use of transgenic methodologies to alter proline metabolism in plants. Subsequent to this, molecular genetic methods may be used to experimentally manipulate transgenic plants. This is likely to enable assessment of the functionality of proline during and after osmotic stress and to eliminate much of the controversy concerning the functional significance of the response.

2.1.3 Proline accumulation as a general stress response

Proline accumulation in plants has been reported in response to a range of other stresses other than water deficit induced by drought or salinity. These include temperature extremes, nutrient deficiency, heavy metal toxicity, exposure to atmospheric pollutants, infection by pathogens, anaerobiosis and nutrient deficiency (Table 2.2). This range of stresses capable of inducing proline accumulation suggests that the response may be part of a general adaptation to adverse environmental conditions. Engineering crops to overproduce proline may thus be a good objective in increasing their overall tolerance to environmental stress.

It seems likely that many of the features of proline and its metabolism that are likely to be of functional value during hyperosmotic stress (Sections 2.1.2.3 and 2.1.2.4) apply equally in response to other stresses. For example, Saradhi and Saradhi (1991) have suggested that during heavy metal stress, proline accumulation may play a role in the regulation of cellular redox potentials as outlined in Section 2.1.2.4.

Furthermore, if proline is accepted as a stabiliser of proteins (Schobert, 1977b; Pollard and Wyn Jones, 1979; Paleg *et al.*, 1984; Arakawa and Timasheff, 1985), membranes (Rudolf *et al.*, 1986) and polysomes (Kandpal and Rao, 1985) during hyperosmotic stress, then its accumulation is likely to be of significance in all stresses capable causing damage to subcellular structures. Nash *et al.* (1982) and Paleg *et al.* (1981) have demonstrated that proline stabilises enzyme structure against heat denaturation, without substantive loss in their catalytic activity. Krall *et al.* (1989) reported that proline protects maize pyruvate phosphate dikinase from cold denaturation. A role for proline in preventing protein degradation and/or inactivation of enzymes in plants subjected to nutrient deficiency has been proposed by Rogozinska and Flasinski (1987). Proline is capable of preventing freezing-induced inactivation of membrane activities (Heber *et al.*, 1973). Since plant membrane damage during chilling is related to the peroxidation of membrane lipid due to stress-induced accumulation of free radicals (Wise and Naylor, 1987), it is possible that proline may act as an antioxidant to counteract chilling-induced free radical accumulation. Furthermore, use of proline as a cryoprotectant for the storage of cultured plant cells at subzero temperatures has been reported by Hellergren and Li (1981) and Withers and King (1979). It is likely that this effect is mediated at the level of protection against solution effects caused by dehydration during freezing (Withers and King, 1979).

Table 2.2 : Stresses other than direct osmotic stress that cause proline accumulation

Stress		Species	Reference
High temperature	(39°C) (39°C)	*Hordeum vulgare* L. *Raphanus sativa* L.	Chu *et al.* (1974)
	(33°C)	*Lycopersicon esculentum* L.	Kuo *et al.* (1986)
Low temperature	(2°C)	*Lolium perenne* L.	Draper (1972)
	(4°C) (4°C)	*Hordeum vulgare* L. *Raphanus sativa* L.	Chu *et al.* (1974)
	(< 8 °C) (<12°C)	*Hordeum distichum* L. *Triticum aestivum* L.	Chu *et al.* (1978)
	(4°C)	*Triticum aestivum* L.	Naidu *et al.* (1991)
Heavy metal stress	(Cd, Co, Zn, Pb)	*Cajanus cajan* L. *Vigna mungo* L. *Triticum aestivum* L.	Saradhi and Saradhi (1991)
	(Cu, Zn)	*Lemna minor* L.	Bassi and Sharma (1993)
Pathogen infection	(tungro virus)	*Oryza sativa* L.	Mohanty and Sridhar (1982)
	(*A.tumefaciens*)	*Nicotiana tabacum* L.	Seitz and Hochster (1964)
		Lycopersicon esculentum L.	Meon *et al.* (1978)
		Citrus spp.	Labanauskas *et al.* (1974)
	(*Phytophthera* root rot)	*Citrus paradisi* L.	Hanks and Feldman (1963)
	(Nematode infestation)	*Lycopersicon esculentum* L.	Meon *et al.* (1978)
Atmospheric pollutants	(SO_2, NH_3, NO_2)	*Oryza sativa* L.	Anbazhagan *et al.* (1988)
Flooding-induced anaerobiosis		*Helianthus annus* L.	Wample and Bewley (1975)
		Lycopersicon esculentum L.	Aloni and Rosenshtein (1982) Kuo and Chen (1980)
		Citrus spp.	Labanauskas *et al.* (1974)
Nutrient deficiency	(low Cl)	***Brassica* spp.**	Freney *et al.* (1959)
	(low K, Mg)	*Theobroma cacao* L.	Machicado and Boynton (1961)
	(low Zn, Cu, Mg)	*Citrus* spp.	Stewart (1962)
	(low Mg)	*Pisum sativum* L.	Klein and Jäger (1978)
	(low P)	*Citrus aurantium* L. *Citrus limonia* L.	Nemec and Meredith (1981)

Evidence that proline may stabilise polysomes during water deficit has been provided by Kandpal and Rao (1985). Polysome degradation is a common symptom of several environmental stresses. Besides being a response to water deficiency (Hsiao, 1970; Bartels *et al.*, 1988), polysome instability has been reported in response to nutrient deficiency (Webster and Van't Hof, 1973; Webster, 1980; Walter and Hahlbrock, 1985), anaerobiosis (Lin and Key, 1967; Sachs *et al.*, 1980), wounding (Theillet *et al.*, 1982) and heat shock (Key *et al.*, 1981). It is tempting to speculate that proline may serve to stabilise the translational apparatus of cells during exposure to a number of different stresses. However, the stabilising effects of proline on subcellular structure cannot account for proline accumulation in response to the full range of stresses listed in Table 2.2.

Oxidative stress appears to be a common component to all of the stresses and physiological conditions associated with proline accumulation. Environmental extremes associated with free radical or oxidative damage in plants include water deficit (Smirnoff, 1993), temperature extremes (Kendall and McKersie, 1989; Schoner and Krause, 1990; Mishra and Singhal, 1992), metal toxicity (De Vos *et al.*, 1992; Moran *et al.*, 1994), exposure to air pollutants (Mehlhorn *et al.*, 1990) and nutrient deficiency (Cakmak and Marschner, 1988). Furthermore, the response of plants to infection by pathogens appears to have a strong oxidative component (Apostol *et al.*, 1989; Chen *et al.*, 1993). All of these stresses induce proline accumulation (Table 2.2). As outlined in Section 2.1.2.3, proline may act as a scavenger of free radicals which accumulate during oxidative stress. Alternatively, turnover of the NADPH pool by proline biosynthesis may increase activity of the oxidative pentose phosphate pathway (Section 2.1.2.4). The oxidative pentose phosphate pathway provides NADPH for the enzymes of the Halliwell-Asada pathway which are involved in free radical scavenger metabolism (Bowler *et al.*, 1992).

The range of stresses capable of inducing proline accumulation introduces the issue of elucidating the molecular basis of this response. If it is assumed that proline accumulation in response to this disparate range of stresses is triggered by the same fundamental mechanism, this necessitates examination of features common the suite of environmental conditions shown in Table 2.2. In the light of the involvement of oxidative stress in all of the environmental conditions capable of inducing proline accumulation, it seems feasible that the response may be triggered by elevated levels of active oxygen in the cell. In bacteria, a number of genes encoding defence and repair enzymes are directly controlled by active

oxygen or its products (Farr and Kogoma, 1991). It is currently not known whether similar regulation of gene expression by active oxygen species occurs in plants. However, in accordance with evidence that proline is an effective scavenger of free radicals (Alia *et al.*, 1991; Smirnoff and Cumbes, 1989), its *de novo* synthesis in response to oxidative stress would not be unexpected.

Alternatively, it has been suggested that many similar effects of different environmental challenges might be due to the existence of a general plant stress response mechanism capable of transducing environmental signals into physiological events (Hanson and Hitz, 1982; Chapin, 1991a; Vernon *et al.*, 1993). This is consistent with the natural coupling of most environmental stresses. Not only do environmental stresses frequently operate in concert, but they are often linked by common aspects of their effects, the signal systems whereby they are detected, or the response to them (Nover, 1989). It seems probable that during the course of evolution, plants have acquired a set of distinct but partially overlapping stress response systems to enable them to survive and propagate under a multitude of unfavourable environmental conditions. This centralised system of physiological responses might enable plants to respond to any physiological stress, regardless of its nature (Chapin, 1991a). However, the basic physiological framework that regulates plant growth in response to stress is likely to be extremely complex, involving changes in hormone balance, water relations, carbon balance and nutrient use (Chapin, 1991a; Vernon *et al.*, 1993).

The hypothesis of a central stress response system is particularly appealing to the plant molecular biologist. If many or all of the potential environmental stresses act directly on a common pathway, then the mapping of stress perception to responses at the cellular level is considerably simplified. The almost universal nature of proline accumulation throughout the plant kingdom (Table 2.1) suggests that the response may be induced at an early stage in a cascade phenomenon designed to produce a rapid highly amplified response irrespective of the nature of the particular stress encountered. It is possible that characterisation of the molecular basis of proline accumulation may serve as a paradigm for understanding signal transduction pathway(s) that link(s) environmental stress with changes in gene expression. At present, the mechanism of perception of osmotic stress by plants and the transformation of such a physical phenomenon into metabolic responses remains enigmatic (Bray, 1994).

Some evidence is available to support the existence of such a general stress response system in prokaryotes. A close connection of oxidative stress and the heat shock response is valid for bacteria. Morgan *et al.* (1986) and Van Bogelen *et al.* (1987) have defined overlapping stress domains, which have been proposed to be parts of a common regulatory entity (stimulon). Changes in gene expression related to nutrient starvation have been extensively studied in microorganisms. Metabolic reprogramming in *Escherichia coli* and *Salmonella typhimurium* under conditions of glucose, phosphate or nitrogen starvation is associated with synthesis of complex, partially overlapping sets of new starvation proteins. Some of these are also induced by heat shock (Jenkins *et al.*, 1988; Spector *et al.*, 1986). Furthermore, starvation of *E. coli* increases heat resistance (Jenkins *et al.*, 1988). These findings may be a guide for future investigations into whether higher plants possess an analogous general stress response system.

Conceivably, such a transduction pathway might be triggered by changes in the levels of plant growth regulators. Several physiological studies have indicated that individual plants respond to most environmental stresses by changing their hormone balance, frequently producing more abscisic acid (ABA) and less cytokinins (Chapin, 1991a). These hormonal changes may constitute the trigger responsible for directly eliciting reduced growth in response to stress (Chapin, 1991a). Although it is likely that there may be several modes of action of plant hormones, increasing evidence points to their primary actions at the level of gene expression (Parthier, 1989). Characterisation of the effects of stress modulators on gene activity is thus likely to be an important component in elucidating the physiological changes that accompany imposition of environmental stress.

ABA is often referred to as a generalised stress hormone, which may be the produced as a consequence of different stresses which affect water status in the plant (Chandler and Robertson, 1994). An increase in ABA is commonly observed in all situations of water deficit and most aspects of the response are mimicked by the application of exogenous ABA. Osmotic stress is a common component in many environmental factors capable of inducing proline accumulation. Elevated temperature and reduced water supply often occur at the same time. High levels of heavy metals in soils are likely to interfere with water uptake by roots (Chapin, 1991b). The reduced mobility of water associated with chilling can slow root water uptake and thereby cause water deficiency. More extreme cold causes internal osmotic stress because ice formation in the extracellular spaces removes water from

the intracellular environment (Hejala *et al.*, 1990).

Endogenous ABA content has been shown to rise in cold-stressed plant tissue (Daie and Campbell, 1981; Chen *et al.*, 1983; Taylor *et al.*, 1990). Exogenous ABA treatment also induces chilling and freezing tolerance in many plant species (Bornman and Jansson, 1980; Rikin *et al.*, 1975; Chen *et al.*, 1983; Reaney *et al.*, 1989; Lang *et al.*, 1989). The involvement of ABA in induced freezing tolerance has been further implicated by the observation that an ABA-deficient mutant of *Arabidopsis* is unable to undergo cold-acclimation; the ability to acclimate is restored by addition of exogenous ABA (Heino *et al.*, 1990). ABA also increases in response to insufficient nitrogen supply (Radin *et al.*, 1982) as well as in response to flood-induced anaerobiosis (Wadman van Schravendijk and van Andel, 1985). Levels of ABA have also been shown to rise in maize seedlings exposed to cadmium (Bonham-Smith *et al.*, 1988). This implicates ABA as part of a response mechanism to heavy metal toxicity.

A role for ABA in induction of proline accumulation has been investigated by several workers. Exogenous ABA treatments result in increased proline concentration in barley (Aspinall *et al.*, 1973; McDonnell *et al.*, 1983; Stewart, 1980; Stewart and Voetberg, 1985; Rajagopal and Anderson, 1978), *Lolium temulentum* (Aspinall *et al.*, 1973), pea (Hasson and Poljakoff-Mayber, 1983), *Arabidopsis thaliana* (Finkelstein and Somerville, 1990) and rice (Chou *et al.*, 1991). The effect of ABA on proline accumulation is attributable to neither acidification of the cell sap (Chou *et al.*, 1991) nor to a reduction in proline degradation (Stewart, 1980; Dallmier and Stewart, 1992). Stewart (1980) presented evidence that the main effect of ABA on proline metabolism is stimulated proline synthesis from glutamate.

However, Henson (1985) showed that injecting ABA into the midrib of *Pennisetum* increased the endogenous ABA concentration but had no effect on proline accumulation. Likewise, McDonnell *et al.* (1983) reported that exogenous ABA treatments had no effect on proline accumulation in *Spinacia* or *Pennisetum* seedlings. Wample and Bewley (1975) found that ABA was incapable of inducing proline accumulation in unstressed sunflower plants. Thomas *et al.* (1992) reported that exogenous ABA was a poor substitute for NaCl in inducing proline accumulation in *Mesembryanthemum*. Furthermore, proline accumulation in a wilty mutant of tomato that contains low levels of ABA and does not accumulate ABA in response to stress was as rapid as in wild-type plants that did accumulate ABA (Stewart

and Voetberg, 1987). This finding contrasts with those of Finkelstein and Somerville (1990) using three ABA-insensitive mutants of *Arabidopsis*. In two of these, proline accumulation in response to exogenous ABA was approximately half that found in wild-type plants. Therefore, ABA and proline accumulation are not causally linked in all species. Stewart and Voetberg (1985) have argued that some of the variability in data concerning ABA-induced proline accumulation in different species may be due to problems of penetration and differences in metabolism of applied ABA in different species.

However, induction of proline accumulation by exogenous application of ABA does not conclusively prove that ABA is involved in the regulation of proline concentration at low water potentials. Although Stewart and Voetberg (1987) found exogenous ABA to increase proline levels in barley leaves, dehydration-induced proline accumulation in barley was unaffected by treatment with fluoridone, an inhibitor of ABA synthesis. In barley (Stewart and Voetberg, 1985) and maize (O'Regan *et al.*, 1993), ABA accumulation in response to water stress precedes a rise in proline levels and upon relief from stress, a decline in ABA precedes the disappearance of proline. This observation is consistent with, but does not demonstrate, a role for ABA in causing stress-induced proline accumulation. Nevertheless, Ober and Sharp (1994) recently reported that accumulation of proline in the maize primary root tip at low water potentials is dependent on ABA, as the response was not observed in wild-type roots treated with fluoridone or in roots of a mutant deficient in carotenoid (and ABA) synthesis.

Together with ABA, cytokinins appear to play an important role in the regulation of growth under stress conditions. However, whereas ABA usually increases in response to environmental stress, cytokinin levels decline in water-stressed plants (Itai and Vaadia, 1971) and plants exposed to anaerobiosis (Burrows and Carr, 1969). Cytokinin levels return to normal upon relief from stress (Walker and Dumbroff, 1981). Cytokinin levels have been shown to decline under conditions of low nitrogen supply (Chapin, 1991a).

Wample and Bewley (1975) reported that benzyladenine inhibited proline accumulation induced by wilting or flooding. Proline levels in wilted sunflower plants treated with benzyladenine were less than 25% of the level found in wilted plants which were not treated with the cytokinin. Treatment of unstressed plants with benzyladenine caused at least a 50% decrease in proline levels throughout the plant (Wample and Bewley, 1975). In *Pennisetum*

(Eder and Huber, 1977), barley (Stewart *et al.*, 1986) and rice (Chou *et al.*, 1991), cytokinin reduces the extent of ABA-induced proline accumulation. Stewart *et al.* (1986) also demonstrated that benzyladenine inhibits proline accumulation in wilted and salt-shocked leaves of barley. However, the mechanism of action of cytokinin on proline levels is unclear. Stewart *et al.* (1986) showed that benzyladenine did not affect ABA accumulation or disappearance in salt-stressed barley. This eliminates the possibility that cytokinin might inhibit stress-induced ABA accumulation in barley. Cytokinin application both before and after imposition of salt stress was found to inhibit proline accumulation (Stewart *et al.*, 1986). In contrast, Chou *et al.* (1991) found no evidence of a cytokinin-mediated inhibition of isobutyric acid-induced accumulation of proline. These workers concluded that cytokinin does not have a direct effect on proline metabolism, but appears to affect the accumulation of proline via an interaction with ABA.

Therefore, of the classical growth regulators, ABA and cytokinin appear to be the most likely candidates for mediating proline accumulation in response to environmental stress. Not only does evidence for hormonal regulation of proline accumulation provide evidence for a functional role of proline in stress tolerance, but it also supports the hypothesis that proline metabolism may be affected by a general stress response system concerned with the acclimation of plants to commonly encountered environmental stresses.

Besides alterations in the levels of these hormones, evidence is currently available implicating events such as changes in jasmonate concentration, redistribution of intracellular Ca^{2+} and protein phosphorylation in linkage of the perception of environmental stress to changes in gene expression (Skriver and Mundy, 1990; Maslenkova *et al.*, 1992; Bray, 1994; Reinbothe *et al.*, 1994). It is probable that these might also play an important role in mediating the accumulation of proline during stress. Many hormones and neurological stimuli in animals regulate cell metabolism by inducing an alteration in the Ca^{2+} concentration in the cytoplasm. As in animal systems, it appears that plant growth regulators might alter cellular Ca^{2+} levels and/or protein phosphorylation (Raz and Fluhr, 1993). A role for Ca^{2+} as a secondary messenger important in modulation of hormone-responsive plant systems is therefore highly probable. In the light of a possible role for Ca^{2+} as a secondary messenger in the drought response of plants, it is interesting to note that Shah *et al.* (1990) demonstrated that NaCl-induced accumulation of proline in callus cultures of *Medicago sativa* was enhanced by Ca^{2+}. Hu *et al.* (1992) have demonstrated that the proline

biosynthetic enzyme Δ^1-pyrroline-5-carboxylate synthetase (P5CS) from *Vigna aconitifolia* contains a potential phosphorylation site. This may be of importance in the loss of feedback inhibition of proline biosynthesis during hyperosmotic stress (Hu *et al.*, 1992; Delauney and Verma, 1993). Urao *et al.* (1994) recently reported the isolation of two genes encoding Ca^{2+}-dependent protein kinases induced by drought and high salt stresses in *Arabidopsis*. These are likely to be of importance in modulation of signal transduction involved in drought-inducible gene expression.

In conclusion, the capacity of a multitude of environmental parameters to induce proline accumulation (Table 2.2) suggests that this may be one of a limited suite of mechanisms that natural selection has chosen for coping with any of several environmental stresses that plants may encounter. Analysis of stress responses unique to a species or group of closely related species is likely to be of only limited value in understanding stress responses in plants from different environments. An apparently convergent stress response such as proline accumulation may enable a synthesis of the mechanistic basis of stress responses. Currently, the molecular basis of proline accumulation is completely unknown. Isolation and characterisation of the genes involved in the process is an essential prerequisite to increasing our overall understanding of the process and how it may be regulated. It is therefore possible that characterisation of the molecular basis of proline accumulation may serve as a paradigm for understanding the signal transduction pathway(s) that link(s) environmental stress with changes in gene expression.

2.2 Proline biosynthesis and degradation in plants

A thorough understanding of the physiology and biochemistry of proline accumulation is an essential prerequisite for any attempt to understand the genetic regulation of proline accumulation in plants. There are at least four mechanisms whereby proline may accumulate in response to stress (Hanson and Hitz, 1982). These include increased synthesis from its precursors glutamate and ornithine, reduced oxidation, impaired incorporation into protein due to a decline in protein synthesis or accelerated protein breakdown.

Although a decline in incorporation of free proline into protein (Stewart *et al.*, 1977; Dungey and Davies, 1982a) and increase in protein degradation (Petrie and Wood, 1958; Shah and Loomis, 1965; Thompson *et al.*, 1966; Dungey and Davies, 1982b) commonly occur in response to osmotic stress in plants, quantities of free proline accumulated in response to osmotic stress are far greater than can be accounted for by such mechanisms (Stewart, 1981; Venekamp and Koot, 1988).

Metabolic labelling studies (Boggess and Stewart, 1976; Boggess *et al.*, 1976a; Rhodes *et al.*, 1986) indicate that most of the proline accumulated in plants in response to stress is the result of enhanced synthesis from glutamate. However, there is also a decrease in proline oxidation concomitant with the onset of stress (Stewart *et al.*, 1977; Rayapati and Stewart, 1991). Proline may also be synthesised from ornithine (Adams and Frank, 1980; Delauney *et al.*, 1993). However, this biosynthetic route is not believed to make a significant contribution to the vast proline pool accumulated in many plants during stress (Boggess and Stewart, 1976; Delauney *et al.*, 1993).

Use of translational and transcriptional inhibitors in studies on ten-day-old *Arabidopsis thaliana* plantlets indicated that both *de novo* transcription and translation are required during the first four hours of salt stress before proline begins to accumulate (Verbruggen *et al.*, 1993). The transcriptional and translational inhibitors cordycepin and cycloheximide also inhibit proline accumulation in barley leaves that have wilted (Stewart *et al.*, 1986) or been treated with ABA (Stewart *et al.*, 1986; Pesci, 1987) or isobutyric acid (Pesci, 1987).

In salt-shocked leaves of barley, cordycepin-mediated inhibition of transcription prevents proline accumulation when added after salinisation but before proline begins to accumulate, but not when added after the onset of proline accumulation (Stewart *et al.*, 1986). These results conflict with those of Verbruggen *et al.* (1993) who found that cordycepin adversely affected proline accumulation even twelve hours after imposition of salt stress in *Arabidopsis*. The translational inhibitor cycloheximide delays proline accumulation in salt-shocked barley leaves (Stewart *et al.*, 1986). However, with time proline accumulates in cycloheximide-treated leaves at rates comparable to salt-treated controls. This delay and subsequent accumulation is observed when cycloheximide is added before, during and after salt treatment. However, the earlier in the salt-treatment period that cycloheximide is applied, the longer is the observed delay (Stewart *et al.*, 1986). These data (Stewart *et al.*, 1986; Pesci, 1987; Verbruggen *et al.*, 1993) indicate that gene activation is involved in proline accumulation.

However, genetic regulation of proline accumulation may vary between species. For example, in the unicellular alga *Chlorella autotrophica*, proline accumulation begins immediately after osmotic shock and is not dependent on protein synthesis (Ahmad and Hellebust, 1984). In cultured cells of the salt marsh grass *Distichlis spicata*, salt-induced accumulation of proline was inhibited by cycloheximide but not by the transcriptional inhibitor actinomycin D (Ketchum *et al.*, 1991). This indicates that mRNA translation, not mRNA transcription, is required before proline production in this species. Proline accumulation may be regulated exclusively at the level of translation of transcripts encoding the biosynthetic enzymes. However, it is worth noting that actinomycin D had no effect on proline accumulation in wilted barley leaves, whereas cordycepin inhibited the process (Stewart *et al.*, 1986). These workers attributed this to a difficulty in ensuring penetration of the inhibitor into cells. The results of Ketchum *et al.* (1991) should therefore not be interpreted to necessarily indicate that proline accumulation is independent of gene transcription. It is likely that in all higher plants, proline accumulation is mediated by increases in both transcription of the proline biosynthetic genes and translation of the corresponding mRNA transcripts. Evidence for transcriptional regulation of proline degradative genes in response to stress is presently lacking, although an inhibition of proline oxidation occurs at the enzyme level (Rayapati and Stewart, 1991).

In any event, these data confirm that proline accumulation is mediated directly at the genomic level. This suggests that proline accumulation is a regulated response to osmotic stress and therefore likely to be of adaptational value. The genetic control of proline accumulation in plants therefore warrants investigation.

Pathways of proline biosynthesis and degradation in plants are shown in Figure 2.2.

2.2.1 Biosynthesis from glutamate

Numerous studies (Morris *et al.*, 1969; Boggess and Stewart, 1976; Boggess *et al.*, 1976a; Wang *et al.*, 1982; Badzinski Buhl and Stewart, 1983; Rhodes et al., 1986; Venekamp and Koot, 1988; Venekamp *et al.*, 1989) indicate that most of the proline accumulated in plants in response to stress is synthesised *de novo* from glutamate. To date, much of our understanding of proline biosynthesis from glutamate in plants is by analogy to the enzymes defined genetically and biochemically in *Escherichia coli* (Baich and Pierson, 1965; Baich, 1969, 1971; Rossi *et al.*, 1977a; Hayzer and Moses, 1978, Hayzer and Leisinger, 1980, 1981, 1982; Deutch *et al.*, 1982; Hayzer and Leisinger, 1983; Hayzer, 1983; Deutch *et al.*, 1984).

Proline biosynthesis from glutamate has been thoroughly characterised in *E. coli* (Leisinger, 1987). The biosynthetic pathway of proline from glutamate is shown in Figure 2.3. In *E. coli*, the proline biosynthetic pathway begins with the phosphorylation of glutamate by γ-glutamyl kinase (γ-GK; ATP:L-glutamate 5-phosphotransferase; EC 2.7.2.11; encoded by the *proB* gene), to form γ-glutamyl phosphate, which is reduced to glutamic-γ-semialdehyde (GSA) by GSA dehydrogenase (L-glutamate 5-semialdehyde:$NADP^+$ oxidoreductase [phosphorylating]; EC 1.2.1.41; encoded by the *proA* gene). The intermediate GSA spontaneously cyclises to Δ^1-pyrroline-5-carboxylate (P5C), which is reduced by P5C reductase (P5CR; L-proline:$NAD(P)^+$ 5-oxidoreductase; EC 1.5.1.2; encoded by the *proC* gene) to proline.

This pathway was first proposed by Vogel and Davis (1952). Subsequent isotope dilution experiments (Vogel and Kopac, 1959) revealed that unlabelled GSA decreased the

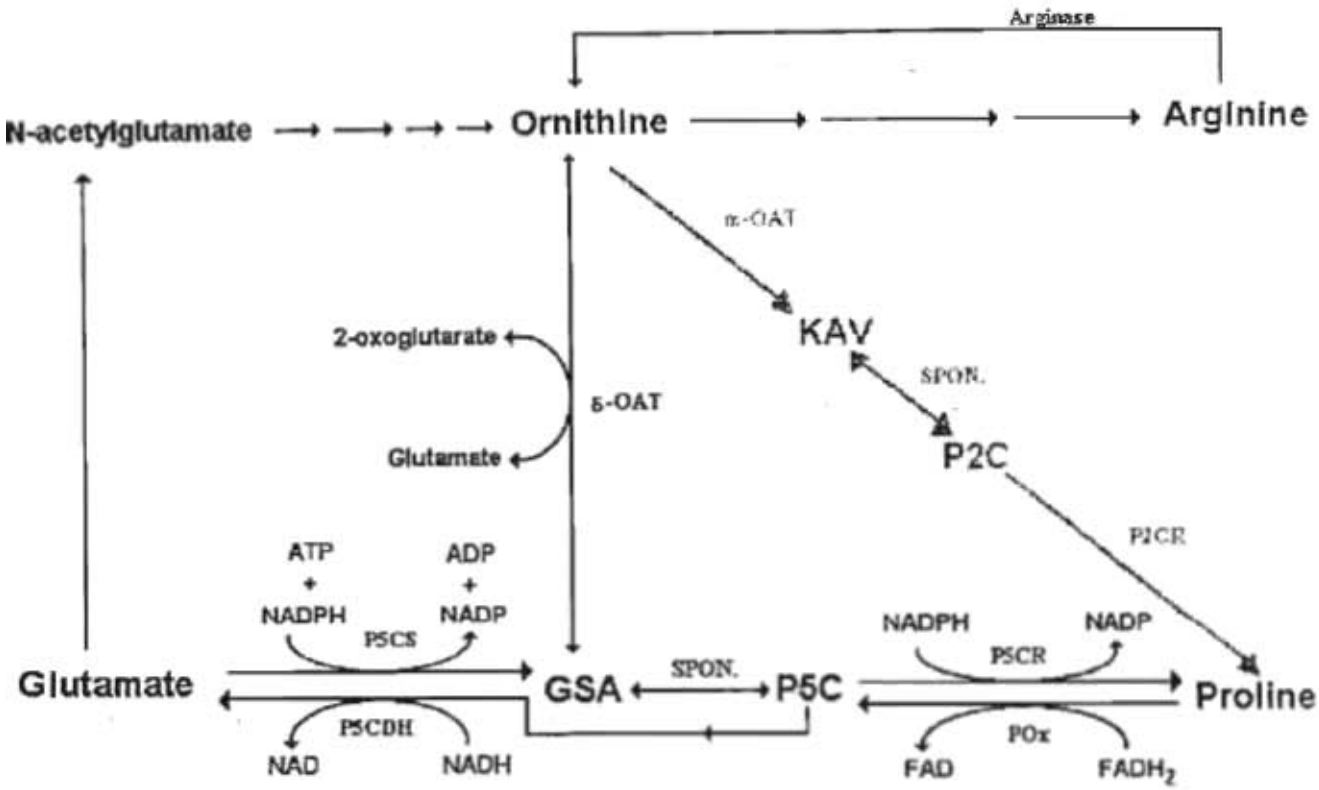


Figure 2.2 : Pathways of proline biosynthesis and degradation in plants. The pathways of proline biosynthesis from glutamate and ornithine converge at the point of formation of GSA, with the subsequent spontaneous cyclisation of GSA to P5C and its reduction to proline being common to both pathways. Abbreviations of intermediates are : GSA, glutamic-γ-semialdehyde; KAV, α-keto-δ-aminovalerate; P2C, Δ^1-pyrroline-2-carboxylate; P5C, Δ^1-pyrroline-5-carboxylate. Abbreviations of enzymes are : α-OAT, ornithine-α-aminotransferase; δ-OAT, ornithine-δ-aminotransferase; P5CDH, P5C dehydrogenase; P2CR, P2C reductase; P5CR, P5C reductase; P5CS, P5C synthase; POx, proline oxidase. The interconversions of GSA and P5C as well as KAV and P2C are spontaneous cyclisations. Conclusive evidence supporting the biosynthesis of proline from ornithine via KAV and P2C (shown in grey) is currently not available.

To date, no data has been presented to support the contention that any of the enzymatic conversions in the biosynthetic pathway from glutamate are reversible. The final step in proline formation and initial step in proline degradation are catalysed by different enzymes. The formation and catabolism of the intermediates GSA and P5C, which are in equilibrium with one another, are also catalysed by separate enzymes. The reaction catalysed by δ-OAT may be reversible.

Adapated from Wagner and Backer (1992) and Delauney *et al.* (1993).

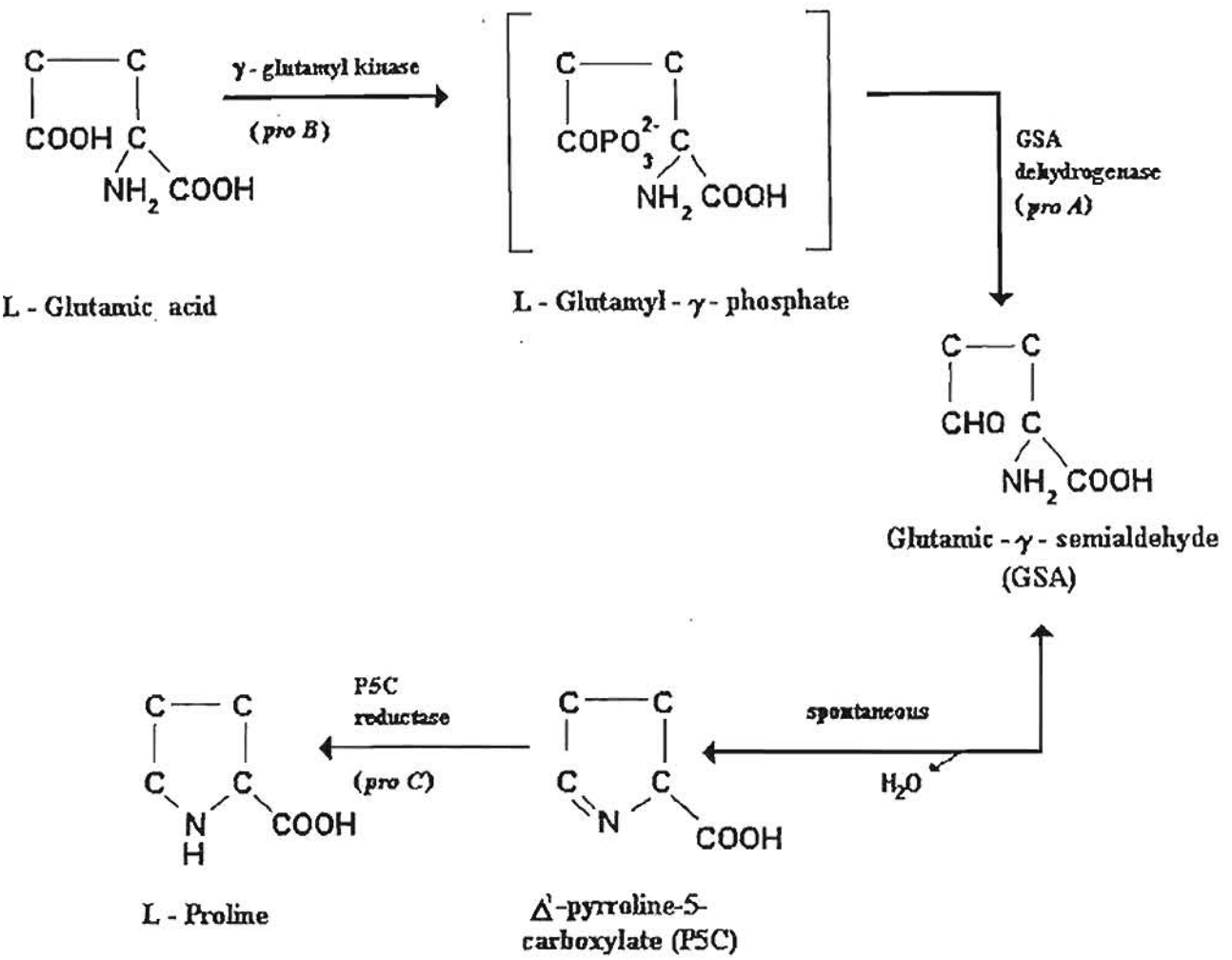


Figure 2.3 : Proline biosynthesis from glutamate in *Escherichia coli*. L-glutamate is activated by γ-glutamyl kinase to form L-glutamyl-γ-phosphate. This unstable intermediate is subsequently converted to glutamic-γ-semialdehyde (GSA) by GSA dehydrogenase. It is believed that the first two enzymes form a complex which ensures that the glutamyl-γ-phosphate remains enzyme bound. Glutamic-γ-semialdehyde spontaneously cyclises to form Δ^1-pyrroline-5-carboxylate (P5C). In the final step, P5C is converted to L-proline by the action of P5C reductase (P5CR). The names assigned to the genes encoding the three proline biosynthetic enzymes are indicated in parentheses.

Adapted from Delauney and Verma (1993).

incorporation of [^{14}C]-glutamate into proline. This supported the proposed route. A further important step in formulating the model for proline biosynthesis in bacteria was the demonstration that feedback regulation of the pathway by proline was at some stage between glutamate and GSA (Strecker, 1957; Baich and Pierson, 1965). Baich (1969) provided evidence that the ATP-dependent enzymatically-catalysed phosphorylation of glutamate was inhibited by proline. This finding supported the existence of a γ-GK specifically involved in P5C synthesis and led to the formulation of the pathway in its present form. Subsequently, *E. coli* mutants resistant to proline inhibition *in vivo* have been shown to contain γ-GK's with markedly reduced sensitivity to proline (Smith *et al.*, 1984; Rushlow *et al.*, 1984; Csonka *et al.*,1988; Dandekar and Uratsu, 1988). This has confirmed that in *E. coli*, regulation of proline biosynthesis is primarily at the level of feedback inhibition of γ-GK.

Besides *E. coli*, genes and enzymes involved in proline biosynthesis from glutamate have also been studied in microorganisms such as *Pseudomonas aeruginosa* (Krishna and Leisinger, 1979; Krishna *et al.*, 1979; Savioz *et al.*, 1990), *Salmonella typhimurium* (Mahan and Csonka, 1983); *Serratia marcescens* (Omori *et al.*, 1991), *Campylobacter jejuni* (Lee *et al.*, 1985; Louie and Chan, 1993), *Methanobrevibacter smithii* (Hamilton and Reeve, 1985), *Desulphovibrio desulphuricans* (Fons *et al.*, 1991), *Clostridium sporogenes* (Costilow and Cooper, 1978), *Treponema pallidum* (Gherardini *et al.*, 1990), *Bacillus subtilis* (Lewis and Wake, 1989; Ahn and Wake, 1991), *Thermus thermophilus* (Hoshino *et al.*, 1994) and *Saccharomyces cerevisiae* (Brandriss, 1979; Matsuzawa and Ishiguro, 1980a, 1980b; Tomenchok and Brandriss, 1987; Brandriss and Falvey, 1992; Li and Brandriss, 1992). Proline biosynthesis from glutamate has also been characterised at the enzyme and genetic levels in mammalian systems (Smith *et al.*, 1980; Yeh *et al.*, 1981; Henslee *et al.*, 1983; Wakabayashi and Jones, 1983; Merrill *et al.*, 1989; Wakabayashi *et al.*, 1991; Dougherty *et al.*, 1992).

In all organisms studied to date, proline biosynthesis from glutamate occurs via the same pathway, originally defined in *E. coli*. This conservation of the pathway has been confirmed by a range of complementation experiments involving diverse organisms often from different biological kingdoms. For example, the *PRO1, PRO2* and *PRO3* genes from *S. cerevisiae*, encoding γ-GK, GSA dehydrogenase and P5CR respectively complement the corresponding *proB*, *proA* and *proC* mutants of *E. coli* and *S. typhimurium* (Tomenchok and Brandriss,

1987). Similarly, the *E. coli proB* gene restores proline prototrophy to the *PRO1* mutant of *S. cerevisiae* (Otser *et al.*, 1988). A 4.8 kb fragment of *C. jejuni* DNA is able to complement both *proA* and *proB* mutants of *E. coli* (Lee *et al.*, 1985). A cDNA from the plant *Vigna aconitifolia* encoding a bifunctional enzyme capable of P5C synthesis complements *proB*, *proA* and *proBA* mutants of *E. coli* (Hu *et al.*, 1992). A human cDNA encoding P5CR can complement the corresponding *PRO3* mutant of *S. cerevisiae* (Dougherty *et al.*, 1992). Complementation of *E. coli proC* mutants has also been reported with genes encoding P5CRs from soybean (Delauney and Verma, 1990), *T. pallidum* (Gherardini *et al.*, 1990) and *A. thaliana* (Verbruggen *et al.*, 1993).

Study of proline biosynthesis in microbial and animal systems and use of comparative enzymology and physiology has contributed greatly to our understanding of proline biosynthesis in plants. However, studies on proline biosynthesis in plants have focused in particular on the role played by the relevant genes and enzymes in osmotic adjustment during hyperosmotic stress (Section 2.1.2). Despite the wealth of information on the enzymology of proline biosynthesis from glutamate gleaned from studies conducted using *E. coli* and other microorganisms, this approach has been less useful in elucidating the osmotic regulation of proline accumulation in plants.

Whereas proline accumulation is a primitive response to osmotic stress, conserved throughout evolution in both eukaryotes and prokaryotes (Measures, 1975), the sources of this accumulation differ. Members of the Enterobacteriaceae accumulate proline as a response to stress by enhanced uptake of exogenous proline and neither synthesis nor catabolism of proline is subject to osmotic control (Csonka, 1989). Osmotic stress has no effect on the rate of proline synthesis or degradation in these organisms, but instead stimulates the activity of the ProP and ProU proline transport systems by inducing transcription of the *proP* and *proU* operons (Csonka and Hanson, 1991). A third and major proline permease, encoded by the *putP* gene is not affected by an increase in the osmolarity of the medium (Csonka, 1989). The PutP protein is required for transport of proline when this metabolite is used as a carbon or nitrogen source (Csonka, 1989). The products of the *proU* and *proP* genes also recognise glycine betaine as a substrate (Cairney *et al.*, 1985a, 1985b). Proline accumulation in *Staphylococcus aureus* at low water potential is also by transport (Anderson and Witter, 1982). In contrast, proline synthesis and degradation appears to be under osmotic control in other Gram-positive bacteria (Csonka, 1989). Whatmore *et*

al. (1990) have shown that in *B. subtilis*, osmotic upshock in a minimal medium increases proline synthesis. However, most eubacteria appear to depend on exogenous proline as an osmoprotectant (Csonka and Hanson, 1991).

To date, the most extensive information obtained concerning proline biosynthesis in plants has been obtained by molecular studies. The control of proline biosynthesis in plants is more complex than in prokaryotic systems since there are biosynthetic routes from both glutamate and ornithine (Delauney and Verma, 1993). Both of these biosynthetic routes are regulated at the levels of enzyme activity and gene expression. Proline metabolism in plants also differs from that in mammalian systems in several important respects, particularly in relation to regulation and subcellular compartmentation.

2.2.1.1 Synthesis of P5C

For many years, understanding of proline biosynthesis lagged far behind that of other amino acid biosynthetic pathways. In particular, despite circumstantial evidence in favour of validity of the first two steps of proline proposed by Vogel and Davis (1952), formal proof of the mechanism of P5C synthesis was elusive. This can largely be ascribed to the extreme lability of γ-glutamyl phosphate to nucleophilic attack and its tendency to cyclise to 5-oxopyrrolidine-2-carboxylate (Strecker, 1957). This hampered efforts to develop specific and sensitive assays for the first two enzymes of the pathway. It was only with the convergence of enzymology and recombinant DNA technology in the early 1980s that conclusive evidence for the pathway emerged in prokaryotic systems. Owing to the comparitive difficulty of enzyme assay in plants added to the low abundance of amino acid biosynthetic enzymes often observed within plant cells (Matthews *et al.*, 1988), conclusive demonstration of a P5C-synthesising reaction in plants was not forthcoming for over 25 years, despite concerted efforts by several research groups.

Recently, the previously unknown nature of P5C synthesis in plants was resolved by the isolation of a cDNA encoding a bi-functional enzyme Δ^1-pyrroline-5-carboxylate synthase (P5CS) in *Vigna aconitifolia* (Hu *et al.*, 1992). Assay of the recombinant enzyme and complementation of an *E. coli proBA* mutant indicated that in *Vigna*, P5CS catalyses the first two steps in proline biosynthesis from glutamate. This represented the first unequivocal

evidence for the existence of a P5C-synthesising reaction in plants. A clone encoding a bifunctional γ-GK/GSA dehydrogenase has also been identified from a tomato cDNA library (Garcia-Rios *et al.*, 1991). *Vigna* P5CS has both γ-GK and GSA dehydrogenase activities. The two enzymatic domains of P5CS correspond to the products of the *proB* and *proA* genes of *E. coli* and contain a leucine zipper in each domain, which may facilitate inter- or intra-molecular interaction of the protein (Hu *et al.*, 1992).

A number of observations concerning the enzymology and genetics of P5C synthesis in bacteria are consistent with the existence of a bifunctional enzyme responsible for P5C synthesis in plants. Consideration of the lability of free γ-glutamyl phosphate led to the notion that in *E. coli*, this compound may exist *in vivo* as an enzyme-bound intermediate (Baich, 1969). Furthermore, Smith *et al.* (1984) demonstrated that to be enzymatically active, *E. coli* γ-GK needs to be complexed with GSA dehydrogenase in an obligatory complex which is extremely labile *in vitro*. The first two enzymes of proline biosynthesis in *E.coli* are believed to function in a complex comprising six subunits of each polypeptide (Deutch *et al.*, 1984). Previous gel filtration studies with crude extracts (Hayzer and Moses, 1978) had suggested the existence of an enzyme complex catalysing the first two reactions in proline biosynthesis and thereby ensuring the direct transfer of the unstable intermediate. Previous attempts to purify γ-GK had led to the dissociation of the two enzymes with a consequent loss of activity of the first enzyme. Purification of γ-GK from *E.coli* (Smith *et al.*, 1984) was successful because a strain containing the *proBA* genes in a multicopy expression vector (Deutch *et al.*, 1984) yielded starting material of high specific activity. Furthermore, the coupled-enzyme assay used throughout the purification, which was based on the NADPH-dependent reduction of γ-glutamyl phosphate by GSA dehydrogenase, satisfied the requirement of γ-GK for the second enzyme in the pathway.

Genetic evidence also exists to justify the existence of a bifunctional enzyme catalysing P5C synthesis in plants. In *E. coli* and *S. marcescens*, the *proB* gene is located directly upstream of the *proA* gene, with their coding sequences located within 15 and 12 bp of each other (Deutch *et al.*, 1984; Omori *et al.*, 1991). In both species, both genes appear to share a common promoter and to be transcribed into a single mRNA encoding both proteins. Similarly, in *S. typhimurium*, the *proB* and *proA* genes form a single operon, with *proB* located proximal to the promoter (Mahan and Csonka, 1983). This arrangement is consistent with the suggestion that the kinase and reductase form a molecular complex ensuring direct

transfer of the intermediate and protecting it from nucleophilic attack (Baich, 1969). Biosynthesis of both enzymes in close proximity is likely to result in relatively high localised concentrations of both enzymes. This is likely to facilitate formation of a physical complex (Deutch *et al.*, 1984). Furthermore, although the genes encoding the two P5C-synthesising enzymes in *C. jejuni* are not arranged in tandem (Louie and Chan, 1993), a 4.8 kb fragment of *C. jejuni* DNA is capable of complementing an *E. coli proBA* mutant (Lee *et al.*, 1985). This suggests that they are in the same vicinity on the chromosome.

While the *proBA* genes comprise an operon at 5.8 min on the *E. coli* chromosome, *proC* represents a distinct transcriptional unit located at 8.9 min on the genetic map (Bachmann, 1990). In the genetic map of *S. typhimurium*, the three *pro* genes are arranged in a similar fashion (Sanderson and Hartman, 1978). The *proB* and *proA* genes of all prokaryotes appear to be closely related to each other in their location on the chromosome. It has even been suggested by Hu *et al.* (1992) that the high level of sequence similarity of the *E. coli* ProB and ProA proteins indicates that they arose by duplication of an ancestral gene. It seems possible that over time, mutation of a stop codon at the 3' end of the gene encoding γ-GK has enabled the fusion of these two genes in plants to form a gene encoding a bifunctional enzyme.

A similar event appears to have occurred in the evolution of another bifunctional plant amino acid biosynthetic gene encoding aspartate kinase-homoserine dehydrogenase (Wilson *et al.*, 1991; Weisemann and Matthews, 1993; Ghislain *et al.*, 1994). Although P5CS activity has been detected in mammalian cells (Smith *et al.*, 1980; Wakabayashi and Jones, 1983; Wakabayashi *et al.*, 1983; Kramer *et al.*, 1985; Wakabayashi *et al.*, 1991), it is currently not known whether this is due to a single bifunctional enzyme or to separate enzymes associated in a complex.

In *S. cerevisiae*, two separate genes (*PRO1* and *PRO2*) encode γ-GK and GSA dehydrogenase (Tomenchok and Brandriss, 1987; Li and Brandriss, 1992). However, it is likely that in yeast these first two enzymes also form a complex to channel the unstable intermediate γ-glutamyl phosphate. Tomenchok and Brandriss (1987) reported that the *PRO1* gene complemented bacterial *proB* strains carrying deletion mutations but not those carrying point mutations. This was interpreted to suggest that yeast γ-GK can complex with the bacterial GSA dehydrogenase if it is not already complexed with a defective bacterial

kinase polypeptide (Tomenchok and Brandriss, 1987).

As in bacterial systems (Smith *et al.*, 1984; Dandekar and Uratsu, 1988; Omori *et al.*, 1991), stimulation of proline biosynthesis in plants has been related to loss of feedback inhibition of proline biosynthesis by the end product of the pathway. Radioisotope experiments (Boggess *et al.*, 1976a) suggest that this loss of feedback inhibition occurs at the level of P5C formation. Accordingly, P5CS from *V. aconitifolia* is sensitive to feedback inhibition by proline (Hu *et al.*, 1992) and is thus the most likely site for the regulation of proline biosynthesis in plants. However, the sensitivity of the recombinant *Vigna* enzyme to feedback inhibition was thirty times less than that of wild-type *E.coli* γ-GK (Hu *et al.*, 1992). The mechanism whereby this feedback inhibition is lost under conditions of stress remains unknown. However, it is of interest to note that a potential phosphorylation site has been identified in P5CS from *V. aconitifolia* (Hu *et al.*, 1992). This may be of some regulatory significance in reducing feedback inhibition of P5C synthesis in plants during stress.

In contrast to the bacterial and plant enzymes, conversion of glutamate to P5C in animals is not inhibited by proline. However, ornithine inhibits mammalian P5CS activity (Smith *et al.*, 1980; Henslee *et al.*, 1983). The possibility of a similar regulation via this alternative precursor of P5C was not investigated by Hu *et al.* (1992). Furthermore, *in vitro* studies with purified *E. coli* γ-GK (Smith *et al.*, 1984) suggest that the activity of the enzyme is modulated not only by proline but also by glutamate and ADP. The sigmoidal saturation kinetics of γ-GK are likely to render it insensitive to small fluctuations in the concentration of glutamate, thereby ensuring a constant rate of proline biosynthesis at varying intracellular concentrations of glutamate (Smith *et al.*, 1984). Competitive inhibition by ADP with respect to ATP is most likely responsible for inhibition of the enzyme under conditions of energy depletion and therefore responsible for fine-tuning proline biosynthesis with the energy charge of the cell (Smith *et al.*, 1984). Partially purified γ-GK from *Pseudomonas aeruginosa* also demonstrates inhibition of catalytic activity by ADP (Krishna and Leisinger, 1979). Future investigations of the allosteric regulation of plant P5CS are likely to be of interest in the overall regulation of proline accumulation under stress conditions.

Although not purified to homogeneity, mammalian P5CS activity has been demonstrated in a mitochondrial membrane fraction (Wakabayashi and Jones, 1983; Wakabayashi *et al.*,

1983). The reaction is dependent on ATP and NADPH. The cofactor NADH cannot replace NADPH (Wakabayashi and Jones, 1983). Similarly, a proline biosynthetic activity dependent on glutamate, ATP, NADPH and Mg^{2+} shown in a cell-free system prepared from blowfly abdomen, was associated exclusively with the mitochondrial fraction (Wadano, 1980). This suggests that the same situation applies in insects. However, the absence of an N-terminal transit peptide in *Vigna* P5CS suggests a cytosolic location for the enzyme in plants (Hu *et al.*, 1992).

Transcription of the *Vigna* P5CS gene is induced in roots subjected to salt stress (Hu *et al.*, 1992). In this respect, regulation of P5C synthesis in plants appears to differ from that in prokaryotes. For example, exposure of *S. marcescens* to high salt concentrations does not affect transcription of the *proBA* operon (Omori *et al.*, 1991). This demonstrates that regulation of P5C synthesis at the transcriptional level is not subject to osmoregulation in this species. Neither synthesis nor catabolism of proline in members of the Enterobacteriaceae appears to be subject to osmotic control (Csonka, 1989).

The possibility that transcription of the *Vigna* gene encoding P5CS may be inactivated by proline was not investigated by Hu *et al.* (1992). However, regulation of plant P5CS activity by proline at the genetic level is unlikely. Expression of *proB* in *E. coli* (Deutch *et al.*, 1984) and *S. marcescens* (Omori *et al.*, 1991) is insensitive to feedback inhibition by proline. Similarly, in *S. cerevisiae*, *PRO1* expression is not repressed by addition of exogenous proline to the growth medium (Li and Brandriss, 1992).

2.2.1.2 Reduction of P5C

Of all the proline biosynthetic enzymes, Δ^1-pyrroline-5-carboxylate reductase (P5CR) has been the most widely studied at the biochemical and genetic levels. This can largely be attributed to the relative ease with which this enzyme may be assayed in comparison with the lack of a reliable assay for the P5C-synthesising reaction in plants (Rayapati *et al.*, 1989). Much of the interest in P5CR in animal systems is associated with its relationship to control of the flux of P5C, which has been identified as an important source of intercellular redox potential and a regulator of cellular metabolism (Phang, 1985; Mixson

and Phang, 1988; Yeh and Phang, 1981, 1988; Merrill *et al.*, 1989). By virtue of its modulation of P5C, the level and allosteric properties of P5CR impinge on several physiologically important processes common to both plant and animal cells. Furthermore, its location at the branch point between the proline synthetic routes from glutamate and ornithine suggests that P5CR may be a control point in proline biosynthesis from these two precursors (Phang, 1985; LaRosa *et al.*, 1991).

Activity of P5CR in higher plants was first detected by Meister *et al.* (1957). To date, plant P5CR has been purified to apparent homogeneity from *Hordeum vulgare* (barley; Krueger *et al.*, 1986), *Nicotiana tabacum* (tobacco; LaRosa *et al.*, 1991) and *Glycine max* (soybean; Chilson *et al.*, 1991; Szoke *et al.*, 1992). Activity of P5CR has also been measured in cotyledons of *Arachis hypogaea* (peanut; Mazelis and Fowden, 1969, 1971), wheat germ (*Triticum aestivum*; Mazelis and Creveling, 1974), cotyledons of *Cucurbita moschata* (pumpkin; Splittstoesser and Splittstoesser, 1973; Rena and Splittstoesser, 1975), cell suspension cultures of *Mesembryanthemum nodiflorum* (Treichel, 1986) and *Solanum tuberosum* (potato; Corcuera *et al.*, 1989) as well as in green leaves and etiolated shoots of *Pisum sativum* (pea; Rayapati *et al.*, 1989).

In addition, the enzyme has been partially purified and characterised from various animal sources (Smith and Greenberg, 1956; Meister *et al.*, 1957; Smith and Greenberg, 1957; Adams and Goldstone, 1960; Greenberg, 1962; Peisach and Strecker, 1962; Strecker, 1971), bacteria (Meister *et al.*, 1957; Adams and Goldstone, 1960; Costilow and Cooper, 1978; Meile and Leisinger, 1982), *Neurospora crassa* (Yura and Vogel, 1959) and *Saccharomyces cerevisiae* (Matsuzawa and Ishiguro, 1980a, 1980b). Human erythrocyte P5CR has been purified to homogeneity (Merrill *et al.*, 1989).

Genes encoding P5CR have been isolated and sequenced from a range of phylogenetically distinct organisms. These include the plants soybean (Delauney and Verma, 1990), pea (Williamson and Slocum, 1992) and *Arabidopsis thaliana* (Verbruggen *et al.*, 1993). Sequences of genes encoding P5CRs from *Homo sapiens* (human; Dougherty *et al.*, 1992), *S. cerevisiae* (Brandriss and Falvey, 1992), the archaebacterium *Methanobrevibacter smithii* (Hamilton and Reeve, 1985) and several eubacteria have also been published. The eubacterial sequences include those from Gram-negative bacteria including *Escherichia coli*

(Deutch *et al.*, 1984), *Pseudomonas aeruginosa* (Savioz *et al.*, 1990) and *Thermus thermophilus* (Hoshino *et al.*, 1994) as well as Gram-positive bacteria. An open reading frame in the *Bacillus subtilis* genome encoding a product with high similarity to the *E. coli proC* gene product has been sequenced by Lewis and Wake (1989). In addition, sequences of genes encoding P5CRs from *Mycobacterium leprae* (author(s) unknown, GenPept Accession No. U00018) and the spirochaete *Treponema pallidum* (F.C. Gherhardini, C.R. Moomaw and P.J. Bassford, Swiss Prot Accession No. P27771) have been entered in international sequence databases.

The enzymes encoded by these genes vary in size from 251 (*M. smithii*; Hamilton and Reeve, 1985) to 319 (*H. sapiens*; Dougherty *et al.*, 1992) amino acids in length. However, the human enzyme appears to be exceptionally large. It contains a C-terminal extension with no homology to other sequences P5CRs. This C-terminal region is possibly of regulatory importance (Dougherty *et al.*, 1992).

In cases where data indicating the size of the native enzyme are available, P5CR appears to be a multimer comprised of identical subunits. From Table 2.3, it is apparent that the size of the P5CR monomer from all sources is approximately 30 000 Da. However, the number of subunits in the holoenzyme is variable between species. Whereas P5CR from human erythrocytes (Merrill *et al.*, 1989) and *E. coli* (Deutch *et al.*, 1982) appears to be a ten-mer or twelve-mer, the enzyme from rat lens (Shiono *et al.*, 1986 cited by Merrill *et al.*, 1989) is an octamer and P5CR from *S. cerevisiae* is a tetramer (Matsuzawa and Ishiguro, 1980a; Brandriss and Falvey, 1992). Reported sizes of native plant P5CRs vary from within the range of 200 000 Da in wheat germ (Mazelis and Creveling, 1974) to 480 000 Da in barley (Krueger *et al.*, 1986). Although native polyacrylamide gel electrophoresis (PAGE) is not accurate for determinations of molecular mass, LaRosa *et al.* (1991) estimated native tobacco P5CR to be between 300 000 and 400 000 Da. These data suggest that most plant P5CRs may comprise at least twelve identical subunits.

However, it is questionable whether this multimeric conformation is essential for enzymatic activity. Szoke *et al.* (1992) purified monomeric soybean nodule P5CR (approximately 29 000 Da) to homogeneity by overexpression of the corresponding cDNA in *E. coli*. Assembly of the multimeric form apparently did not occur in this foreign host; however, activity was still evident. Since translation was initiated from the start codon of the soybean

Table 2.3 : Native and subunit molecular weights of P5CRs characterised to date.

Species	Molecular Weight		Reference
	Native[a]	Subunit[b]	
Bacteria			
Escherichia coli	320 000	ND	Rossi *et al.* (1977a)
	280 000	28 112	Deutch *et al.* (1982)
Pseudomonas aeruginosa	95 000		Krishna *et al.* (1979)
		28 097	Savioz *et al.* (1990)
Thermus thermophilus	50 000 - 190 000	27 819	Hoshino *et al.* (1994)
Mycobacterium leprae	ND	30 240	Author(s) unknown; GenPept Acc. No. U00018
Treponema pallidum	170 000	28 142	Gherardini *et al.*(1990) Swiss Prot Acc. No. P27771
Methanobrevibacter smithii	ND	27 836	Hamilton and Reeve (1985)
Yeasts			
Saccharomyces cerevisiae	125 000		Matsuzawa and Ishiguro (1980a)
		30 120	Brandriss and Falvey (1992)
Plants			
Triticum aestivum L.	approx. 200 000	ND	Mazelis and Creveling (1974)
Hordeum vulgare L.	220 000 - 480 000	30 000 (SDS-PAGE)	Krueger *et al.* (1986)
Mesembryanthemum nodiflorum L.	324 000	ND	Treichel (1986)
Glycine max L.	ND	28 586	Delauney and Verma (1991)
Nicotiana tabacum L.	300 000 - 400 000 (PAGE)	ND	LaRosa *et al.* (1991)
Pisum sativum L.	ND	28 242	Williamson and Slocum (1992)
Arabidopsis thaliana (L.) Heynh.	ND	28 626	Verbruggen *et al.* (1993)
Animals			
Homo sapiens	300 000 - 350 000	30 000 (SDS-PAGE)	Merrill *et al.* (1989)
		33 400	Dougherty *et al.* (1992)

[a] Deduced using gel filtration, unless otherwise specified
[b] Deduced from the nucleotide sequence of the corresponding gene, unless otherwise specified
ND Not determined

cDNA (Delauney and Verma, 1990), the enzyme purified was not a fused protein (Szoke *et al.*, 1992).

Krueger *et al.* (1986) ascribed the variation in the number of subunits participating in the P5CR homopolymer to differences in the buffer environment of the enzyme. Gel filtration in Tris buffer indicated a native molecular weight of 480 000 Da, whereas when phosphate buffer was used, a fairly broad elution profile with a main peak of molecular weight 220 000 Da and additional distinct shoulders on both sides of the main peak was observed (Krueger *et al.*, 1986). Thus, differences in the buffers used in different purification strategies for P5CRs may explain some of the heterogeneity observed in the reported sizes of native enzymes. LaRosa *et al.* (1991) found no differences in the electrophoretic mobilities of native P5CRs from NaCl-adapted and non-adapted tobacco cells. However, as suggested by Krueger *et al.* (1986), the possibility that changes in quaternary structure of P5CRs by aggregation or disaggregation of the subunits may be of regulatory significance cannot be disregarded.

Although the reaction catalysed by P5CR is not normally rate-limiting in proline biosynthesis in plants (LaRosa *et al.*, 1991; Szoke *et al.*, 1992), increased activities of P5CR and enhanced expression of the corresponding gene have been shown in a number of plants in which proline participates in osmotic adjustment. Activity of P5CR increased approximately four fold in response to salt stress in both the halophyte *Mesembryanthemum nodiflorum* (Treichel, 1986) and the halophytic alga *Chlorella autotrophica* (Laliberte and Hellebust, 1989b). A four-fold increase in activity of P5CR from the epidermal cells of water-stressed barley leaves has also been reported (Argandona and Pahlich, 1991). Both KCl and $MgCl_2$ increased P5CR activity in pea chloroplasts by at least two fold (Rayapati *et al.*, 1989). Furthermore, levels of P5CR transcript increased six-fold in the roots of soybean seedlings (Delauney and Verma, 1990) and approximately five-fold in the roots of pea seedlings (Williamson and Slocum, 1992) in response to short-term salinisation. Verbruggen *et al.* (1993) have reported a correspondingly high induction of P5CR transcription in salt-stressed *Arabidopsis* seedlings. This transcriptional regulation of P5CR in response to osmotic stress is consistent with the induction of *Vigna* P5CS gene transcription by salt stress (Hu *et al.*, 1992).

These findings suggest that even if P5CR does not catalyse the rate-limiting step in proline biosynthesis, transcription of the corresponding gene is an osmotically sensitive process. It is tempting to speculate that stress-inducible transcription of the proline biosynthetic genes and/or translation of the corresponding transcripts indicates that proline accumulation is critical to survival of the stress. However, it may be that the process is simply sensitive to the stress rather than being part of a tolerance mechanism.

Amino acid biosynthetic pathways are generally compartmentalised in eukaryotes. Most amino acid biosynthetic genes characterised to date encode enzymes destined for the chloroplast (Coruzzi, 1991). Accordingly, P5CR activity has been detected in the chloroplast fraction of tobacco (Noguchi *et al.*, 1966) and pea (Rayapati *et al.*, 1989) leaves. The latter workers reported that P5CR in chloroplasts can account for all of the pea leaf P5CR activity (Rayapati *et al.*. 1989). Rayapati *et al.* (1989) also detected P5CR activity within the etioplasts of etiolated seedlings. On the basis of its kinetic parameters, this enzyme appeared to be different to P5CR found in chloroplasts (Rayapati *et al.*, 1989). However, in soybean root nodules, P5CR is in the cytosol and not in plastids (Kohl *et al.*, 1988).

Interestingly, the curve showing changes in activity of P5CR from mung bean hypocotyls with increasing pH has a shoulder between pH 6.0 and 6.4 that corresponds with the optimal P5CR activity from barley shoots (Elthon and Stewart, 1984). In contrast, the optimal activity of mung bean P5CR at pH 8.0 corresponds with a shoulder of activity in the barley enzyme (Elthon and Stewart, 1984). It has been speculated by these workers that this may reflect the existence of two P5CR isoforms in plant tissues. The same phenomenon was noted in the pH dependence of P5CRs from pumpkin cotyledons (Rena and Splittstoesser, 1975) and pea plastids (Rayapati *et al.*, 1989). However, although P5CR from cultured tobacco cells also showed a smooth, broad response to changes in pH characteristic of plant P5CRs (Noguchi *et al.*, 1966; Splittstoesser and Splittstoesser, 1973; Miler and Stewart, 1976; Laliberte and Hellebust, 1989b; Szoke *et al.*, 1992), no major shoulders or dual pH optima were apparent (LaRosa *et al.*, 1991). Similarly, Chilson *et al.* (1991) failed to observe a biphasic profile for the activity of P5CR from soybean nodules with changes in pH. This study used a P5CR purified over 2 300 fold (Chilson *et al.*, 1991). A single pH optimum between pH 7.0 and 7.6 was observed for P5CRs from *Cucurbita maxima* and *Cucurbita moschata* cotyledons (Splittstoesser and Splittstoesser, 1973). The P5CRs from five halophytes studied by Treichel (1986) all exhibited a single pH optimum of

approximately pH 7. Furthermore, although P5CR activities from chloroplasts and etioplasts each had two separate pH optima, native isoelectric focusing revealed only a single band of P5CR activity with a pI of 7.8 (Rayapati *et al.*, 1989). It therefore remains uncertain whether multiple pH optima of certain plant P5CRs reflect the existence of isozymes or not.

Nevertheless, convincing evidence for the existence of P5CR isoenzymes has been presented in soybean, where 15% of P5CR activity in the leaves was reported to be associated with the plastidic fraction (Szoke *et al.*, 1992). Western blot assays of proteins from cytosol and plastid fractions of soybean leaves indicated that an antibody specific for soybean P5CR reacted with a slightly smaller peptide in the chloroplast than in the cytosol (Szoke *et al.*, 1992). These workers suggested that this smaller peptide may be a processed P5CR devoid of a transit peptide. However, expression of a soybean gene encoding P5CR in transgenic tobacco led to the accumulation of the enzyme exclusively in the cytoplasm. This demonstrated the absence of a transit peptide capable of targeting the enzyme to the plastid. Chilson *et al.* (1991) have also provided evidence for the existence of isozymes of P5CR in soybean, although these workers were unable to prove conclusively the apparently different polypeptides did not represent different degrees of self association of the same polypeptide, its interaction with other cellular constituents or post-translational modifications.

Genomic Southern analysis has suggested the presence of multiple copies of P5CR-encoding genes in soybean (Delauney and Verma, 1990), pea (Williamson and Slocum, 1992) and possibly also in *Arabidopsis* (Verbruggen *et al.*, 1993). However, throughout the purification of tobacco P5CR, no evidence of multiple peaks of activity was observed in chromatographic profiles (LaRosa *et al.*, 1991).

The absence of a convincing transit peptide in the primary structure of the peptides corresponding to any of the plant P5CR clones isolated to date (Delauney and Verma, 1990, Williamson and Slocum, 1992; Verbruggen *et al.*, 1993) suggests that they all encode cytosolic enzymes. The apparent presence of multiple copies of P5CR genes in legumes may be due to cross-hybridisation or the presence of either pseudogenes or identical functional copies of the gene in these species. In *Arabidopsis*, multiple copies are only apparent under hybridisation conditions of medium to low stringency (Verbruggen *et al.*, 1993).

Therefore, the genetic basis for the existence of chloroplastic P5CR or the pH dependence of P5CR activity from certain plant species remains uncertain. Mammalian P5CR activity is located primarily in the cytosol (Meister *et al.*, 1957; Peisach and Strecker, 1962). Soybean nodule (Kohl *et al.*, 1988) and yeast (Brandriss and Falvey, 1992) P5CRs have also been reported to be cytosolic proteins. One possibility is that chloroplastic P5CR may be encoded by a chloroplastic gene and that the enzyme is not translocated from the cytoplasm to the chloroplast in a manner similar to the translocation of other nuclearly-encoded chloroplastic enzymes. Preliminary, yet inconclusive evidence in favour of a P5CR gene in the soybean plastome was reported by Delauney and Verma (1990).

Similar controversy surrounds the existence of multiple genes encoding P5CR in animal systems. At least two forms of P5CR may exist in humans. They have been distinguished primarily on the basis of their relative affinities for coenzymes and sensitivities to inhibition by proline and NADP (Yeh and Phang, 1981). One form, typified by that found in proliferating fibroblasts (Yeh *et al.*, 1981), seems to be involved in the generation of proline for protein synthesis. The erythrocyte enzyme, an example, of the second type, appears to have a primary role in transferring intercellular redox potential (Yeh and Phang, 1981). However, Dougherty *et al.* (1992) have suggested that human P5CR is encoded by a single copy gene. These workers suggested that differential splicing of the P5CR mRNA or differential use of polyadenylation sites might account for the presence of at least two bands when poly(A) mRNA is probed with the human P5CR cDNA (Dougherty *et al.*, 1992). Allosteric regulation or post-translational modifications of the same gene product in different tissues also remain possible explanations for the different activities (Dougherty *et al.*, 1992).

Kinetic studies have indicated that all P5CRs isolated to date exhibit Michaelis-Menten kinetics. This has permitted the determination of the kinetic parameters of enzymes from different sources (Table 2.4).

Enzymatic studies have indicated that most P5CRs characterised to date can use either NADH or NADPH to support the reduction of P5C (Noguchi *et al.*, 1966; Mazelis and Fowden, 1971; Mazelis and Creveling, 1974; Rena and Splittstoesser, 1975; Miler and Stewart, 1976; Rossi *et al.*, 1977a; Krishna *et al.*, 1979; Matsuzawa and Ishiguro, 1980a; Yeh *et al.*, 1981; Deutch *et al.*, 1982; Krueger *et al.*, 1986; Treichel, 1986; Hellebust and Larochelle, 1988; Kohl *et al.*, 1988; Merrill *et al.*, 1989; Rayapati *et al.*, 1989; LaRosa

Table 2.4 : Affinities of P5CRs for substrate and pyridine nucleotide cofactors.

Species	Ligand K_m		Reference
Bacteria			
Escherichia coli	P5C	0.15 mM (0.09 - 0.14 mM NADPH)	Rossi *et al.* (1977)
	P5C	0.14 mM (0.15 - 0.20 mM NADH)	
	NADPH	0.03 mM (0.3 - 0.8 mM P5C)	
	NADH	0.23 mM (0.3 - 0.8 mM P5C)	
Pseudomonas aeruginosa	P5C	0.09 mM (0.2 mM NADH)	Krishna *et al.* (1979)
	P5C	0.12 mM (0.2 mM NADPH)	
	NADPH	0.20 mM (0.5 mM P5C)	
	NADH	0.05 mM (0.5 mM P5C)	
Desulfovibrio desulfuricans Norway	P5C	0.1 - 0.2 mM	Fons *et al.* (1991)
	NADH	0.3 mM	
Yeasts			
Saccharomyces cerevisiae	P5C	0.08 mM (0.11 mM NADH)	Matsuzawa and Ishiguro (1980a)
	NADPH	0.0056 mM (0.16 mM P5C)	
	NADH	0.0048 mM (0.16 mM P5C)	
Protozoans			
Acanthamoeba castellanii	P5C	0.012 mM (0.4 mM NADH)	Hellebust and Larochelle (1988)
	P5C	0.320 mM (0.4 mM NADPH)	
	NADPH	0.100 mM	
	NADH	0.015 mM	
Plants			
Nicotiana tabacum L.	P5C	0.15 - 0.18 mM (0.41 mM NADPH)	LaRosa *et al.* (1991)
	P5C	0.60 mM (0.51 mM NADH)	
	NADPH	0.03 mM (0.7 mM P5C)	
	NADH	0.51 mM (0.7 mM P5C)	
Glycine max L.	P5C	0.12 mM (0.68 mM NADPH)	Kohl *et al.* (1988)
	P5C	0.20 mM (0.68 mM NADH)	
(root nodules)	NADPH	0.06 mM (0.5 mM P5C)	
	NADH	1.55 mM (0.5 mM P5C)	
(leaves)	P5C	0.212 mM (0.24 mM NADPH)	Szoke *et al.* (1992)
	P5C	0.179 mM (0.24 mM NADH)	
	NADPH	0.031 mM (0.48 mM P5C)	
	NADH	0.385 mM (0.48 mM P5C)	
Pisum sativum L. (chloroplasts)	NADPH	0.12 mM (2 mM P5C)	Rayapati *et al.* (1989)
	NADH	0.19 mM (2 mM P5C)	
(etiolated seedlings)	NADPH	0.10 mM (2 mM P5C)	
	NADH	0.43 mM (2 mM P5C)	
Animals			
Homo sapiens (erythrocytes)	P5C	0.23 mM (1.0 mM NADPH)	Merrill *et al.* (1989)
	P5C	1.49 mM (5.0 mM NADH)	
	NADPH	0.04 mM (0.7 mM P5C)	
	NADH	0.64 mM (4.0 mM P5C)	
(fibroblasts)	P5C	0.20 mM (0.27 mM NADPH)	Yeh *et al.* (1981)
	P5C	0.41 mM (0.36 mM NADH)	
	NADPH	0.09 mM (0.04 mM P5C)	
	NADH	0.12 mM (0.04 mM P5C)	

et al., 1991; Szoke *et al.*, 1992).

Although soybean P5CR *in vitro* uses both NADPH and NADH as cofactors, its K_m value for NADH is approximately twelve-fold higher than for NADPH (Szoke *et al.*, 1992; Table 2.4). Kohl *et al.* (1988) also found the soybean nodule enzyme to possess a considerably higher K_m for NADH than NADPH, although these workers estimated it to be approximately 25 times larger (Table 2.4). Although Miler and Stewart (1976) also reported a lower K_m of soybean P5CR for NADPH than NADH, activity was higher with NADH.

The P5CRs from tobacco chloroplasts (Noguchi *et al.*, 1966), peanut cotyledons (Mazelis and Fowden, 1971), barley leaves (Krueger *et al.*, 1986) and pea chloroplasts (Rayapati *et al.*, 1989) are also more active with NADPH than NADH. Turnover rates of P5CR from *M. nodiflorum* are four to five fold higher with NADPH than NADH (Treichel, 1986).

In contrast, although P5CR in etioplasts of etiolated pea shoots has a lower K_m for NADPH than NADH (Table 2.4), its V_{max}/K_m ratios with the two cofactors indicates that it is more active with NADH than NADPH (Rayapati *et al.*, 1989). The P5CR from wheat germ was reported to use NADPH at only 25% of the rate of use of NADH (Mazelis and Creveling, 1974). Rena and Splittstoesser (1975) reported that P5CR from pumpkin cotyledons had a 2.5 fold greater activity with NADH than NADPH. Earlier workers (Splittstoesser and Splittstoesser, 1973) reported the activity of pumpkin P5CR to be 4.5 fold greater in the presence of NADH than NADPH.

A ten to twenty fold higher affinity for NADPH than for NADH has also been observed for purified human erythrocyte P5CR (Merrill *et al.*, 1989; Table 2.4). These workers also reported a five to ten fold higher affinity for P5C with NADPH as a cofactor than with NADH. In contrast, no significant difference in K_m was detected for P5C in the presence of NADPH or NADH using soybean P5CR (Kohl *et al.*, 1988; Szoke *et al.*, 1992). Nevertheless, LaRosa *et al.* (1991) reported that the K_m of tobacco P5CR for P5C is lower in the presence of NADPH than NADH (Table 2.4). Like most of the other plant P5CRs characterised, tobacco P5CR also displays a preference for NADPH over NADH as a reductant (LaRosa *et al.*, 1991; Table 2.4).

This cofactor preference is not reflected in all P5CRs studied. For example, P5CR from *S. cerevisiae* does not appear to display any preference for either NADH or NADPH (Matsuzawa and Ishiguro, 1980a; Table 2.4). Affinity of P5CR from the soil amoeba *Acanthamoeba castellanii* (Hellebust and Larochelle, 1988) for NADPH is almost an order of magnitude lower than for NADH and maximum activity with NADH is eight times higher than with NADPH (Table 2.4). The P5CR from *P. aeruginosa* also appears to preferentially use NADH over NADPH (Krishna *et al.*, 1979; Table 2.4). A highly purified form of *Clostridial* P5CR specifically used NADH (Costilow and Cooper, 1978). Similarly, an unfractionated system from blowfly flight muscle (Balboni, 1978) catalysed P5C reduction to proline only in the presence of NADH.

Given that NADPH is metabolically more expensive than NADH, the definite cofactor preference exhibited by plant P5CRs is likely to be of some significance. It supports the hypothesis that P5CR activity is involved in the regulation of cellular redox potential by affecting the level of reduction of the NADPH pool and is therefore important in metabolic regulation (Section 2.1.2.4).

On the basis of the kinetic data presented in Table 2.4, there appear to be two classes of P5CR, which may be classified on the basis of their preference for either phosphorylated or non-phosphorylated pyridine nucleotide cofactors. It is interesting to note that the enzymes with higher affinity for NADH generally have correspondingly higher affinities for P5C in comparison with those using NADPH (Table 2.4).

In the *E. coli* P5CR, proline and NADP, the end-products of the reaction, act as competitive inhibitors of the enzyme (Rossi *et al.*, 1977a). The cofactor NAD does not inhibit *E. coli* P5CR (Rossi *et al.*, 1977a). The P5CR from *P. aeruginosa* is also inhibited by proline, but not by high concentrations of NAD or NADP (Krishna *et al.*, 1979). However, the high levels of proline required for inhibition of P5CRs from *E. coli* and *P. aeruginosa* make it arguable as to whether this is of any physiological significance (Krishna *et al.*, 1979). Nevertheless, in mammalian cell culture systems, it has been suggested that the high sensitivity of P5CR to proline is an effective control of proline synthesis (Valle *et al.*, 1973; Valle *et al.*, 1975). Inhibition of proline synthesis from glutamate via inhibition of P5CR in mammalian cultured cells has also been reported by Eagle *et al.* (1965). Rena and Splittstoesser (1975) reported 25% inhibition of P5CR from pumpkin cotyledons by 20 mM

proline. However, soybean P5CR is only inhibited at proline concentrations five times in excess of this (Miler and Stewart, 1976). Slight inhibition of P5CR from pumpkin cotyledons by 50 mM proline has also been reported by Splittstoesser and Splittstoesser (1973).

Soybean P5CR activity is inhibited by NADP, but not by proline, NAD or ATP (Kohl *et al.*, 1990; Szoke *et al.*, 1992). Both NADH- and NADPH-dependent activities of soybean P5CR are inhibited by NADP (Miler and Stewart, 1976). Inhibition of P5CR by NADP has also been reported for the enzyme from pumpkin cotyledons (Splittstoesser and Splittstoesser, 1973; Rena and Splittstoesser, 1975). The same situation applies to P5CR isolated from human erythrocytes (Merrill *et al.*, 1989). This, together with the cofactor preference exhibited by P5CRs directly supports the hypothesis that P5CR activity is involved in regulation of cellular redox potential and is therefore important in metabolic regulation (Section 2.1.2.4).

There is indirect evidence of at least two forms of P5CR in animal tissues. They are distinguished primarily on the basis of their relative affinities for coenzymes and their sensitivities to inhibition by proline and NADP (Yeh and Phang, 1981). One form, typified by that found in cultured fibroblasts (Yeh *et al.*, 1981) and a lymphoblastoid cell line (Lorans and Phang, 1981), seems to be involved in the generation of proline for protein synthesis. This form has a similar affinity for either cofactor and an affinity for P5C that is not dependent on the choice of cofactor. This form is inhibited by proline but not by NADP. The primary role of a second form, identified in human erythrocytes (Merrill *et al.*, 1989), bovine retina (Matsuzawa, 1982 cited by Merrill *et al.*, 1989) and rat lens (Shiono *et al.*, 1986 cited by Merrill *et al.*, 1989) has a twenty to sixty fold lower K_m for NADPH *versus* NADH and has a five to twelve fold higher affinity for P5C with NADPH as cofactor (Merrill *et al.*, 1989). Purified erythrocyte P5CR uses NADPH exclusively when both pyridine nucleotides are available at physiologic concentrations (Merrill *et al.*, 1989). This second form, which is inhibited by NADP, but not by proline, appears to be involved in the transfer of intercellular redox potential by the production of NADP which is necessary for activity of the oxidative pentose phosphate pathway.

Comparable evidence for P5CR isozymes in plant tissues is not as strong. Plants require P5CR for proline synthesis needed for basic housekeeping functions such as protein and cell

wall biosynthesis. It is appealing to speculate that a possible means whereby these requirements may be regulated independently of stress-induced proline biosynthesis may be via alternative forms of P5CR. Nevertheless, a proline sensitive P5CR analogous to the enzyme found in human fibroblasts (Yeh *et al.*, 1981) has not been described in plant tissues. Partially-purified P5CR from soybean exhibited partial inhibition by proline only at very high concentrations (Miler and Stewart, 1976). Subsequent workers (Kohl *et al.*, 19888; Szoke *et al.*, 1992) have failed to confirm any inhibitory effect of proline on soybean P5CR. Splittstoesser and Splittstoesser (1973) reported only very weak inhibition of pumpkin P5CR by proline. However, it is interesting to note that activity of P5CR in pumpkin in the presence of both NADH and NADPH was equal to the activities in the presence of either NADH and NADPH alone (Rena and Splittstoesser, 1975). This was interpreted to suggest the presence of two isoforms of P5CR; one using NADH and the other NADPH (Rena and Splittstoesser, 1975). Nevertheless in an extensive study of P5CRs from both salt-adapted and non-adapted tobacco cells, LaRosa *et al.* (1991) failed to demonstrate any differences in the enzyme from both sources in their affinities for cofactors or P5C, their pH profiles, chromatographic behaviour during purification or electrophoretic mobility of the native enzymes. In contrast however, Treichel (1986) reported that adaptation of *M. nodiflorum* cells to 200 mM and 400 mM NaCl resulted in a drop in the K_m of P5CR for P5C.

The osmoregulation of P5CR expression in plants (Delauney and Verma, 1990; Williamson and Slocum, 1992; Verbruggen *et al.*, 1993) contrasts with a lack of transcriptional regulation of *Arabidopsis* P5CR expression by proline (Verbruggen *et al.*, 1993). Exogenous proline also has no effect on expression of the *PRO3* gene from *S. cerevisiae* (Brandriss and Falvey, 1992). In *Salmonella typhimurium*, the *proC* gene, unlike most other amino acid biosynthetic genes, is constitutively expressed (Brady and Csonka, 1988). Proline starvation or excess has little effect on β-galactosidase synthesis in strains of *S. typhimurium* carrying *proC-lacZ* operon fusions (Brady and Csonka, 1988).

However, Deutch *et al.* (1982) failed to correlate gene dosage with P5CR activity when *proC* was cloned into multicopy expression plasmids and introduced into *E. coli*. This lack of a gene dosage effect appears to be specific for *proC*, since it was not observed with either the *proB* or *proA* genes in *E. coli* (Smith *et al.*, 1984). Deutch *et al.* (1984) suggested that the lack of a strong gene dosage effect for the *proC* gene in *E. coli* might be the result of either autogenous regulation of *proC* by proline or a requirement for positive activation

of *proC*.

This argument in favour of possible transcriptional regulation of P5CR is bolstered by the observation of Rossi *et al.* (1977b) that *argD*⁻ mutants of *E. coli* (deficient in acetylornithine-δ-aminotransferase) possess elevated levels (20-60% higher) of P5CR activity in crude extracts. However, the nature of this mutation has not yet been elucidated.

Despite the lack of regulation of exogenous proline on *PRO3* expression in *S. cerevisiae* (Brandriss and Falvey, 1992), comparison of the yeast *PRO1* and *PRO3* promoters revealed several highly homologous regions (Li and Brandriss, 1992). It was speculated by these workers that these homologous sequences may be binding sites for a regulatory protein capable of coordinating expression of the yeast proline biosynthetic genes (Li and Brandriss, 1992). Many amino acid biosynthetic genes in *S. cerevisiae* are regulated by a global system known as the general control of amino acid biosynthesis (Hinnebusch, 1988). Starvation for a single amino acid in the pathways regulated by this system causes an increase in RNA and enzyme production of all the co-regulated members of the group.

In this context it is interesting to note that the 5' untranslated region of the *Arabidopsis* P5CR gene contains a GCN4-analogous binding site motif (Verbruggen *et al.*, 1993). The transcriptional factor GCN4 is responsible for regulating expression of genes involved in the general control of amino acid biosynthesis in yeast (Arndt and Fink, 1986). Furthermore, molecular analysis of the aspartate kinase-homoserine dehydrogenase gene from *Arabidopsis* recently presented by Ghislain *et al.* (1994) also revealed the presence of a sequence similar to the GCN4 binding site in the vicinity of the TATA box. Although it is currently not known whether a system of general control of amino acid biosynthesis exists in plants as in yeast, it is tempting to speculate that these regulatory elements may be involved in controlling carbon flux into amino acid biosynthesis in plants.

In *S. cerevisiae*, the *PRO1* and *PRO2* genes are regulated by the general amino acid control system (Li and Brandriss, 1992). In contrast, despite the presence of a sequence identical to the core GCN4 binding site, transcription of the *PRO3* gene does not appear to be under the control of the GCN4 protein (Brandriss and Falvey, 1992). However, the possibility that in plants transcription of the gene encoding P5CR may also be regulated by the cellular status of other amino acids cannot be disregarded. The *Arabidopsis* P5CR sequence

presented by Verbruggen *et al.* (1993) is the only genomic plant sequence currently available. Isolation of genomic clones encoding P5CRs from other plant species may confirm this intriguing possibility.

2.2.2 Synthesis from ornithine

In eukaryotes, the control of proline biosynthesis is more complex than in *Escherichia coli*, since proline may also be synthesised from the non-protein amino acid ornithine (Adams and Frank, 1980). Ornithine may either be carbamylated to citrulline for participation in the urea cycle or transaminated to glutamic-γ-semialdehyde (GSA) and its tautomer Δ^1-pyrroline-5-carboxylate (P5C) for subsequent conversion to proline or glutamate (Jones, 1985).

Ornithine δ-aminotransferase (OAT; L-ornithine:2-oxo-acid aminotransferase, EC 2.6.1.13) catalyses the condensation of ornithine with 2-oxoglutarate to form glutamate and GSA. The GSA is in chemical equilibrium with its enamine P5C, which may be reduced to proline or oxidised to glutamate (Strecker, 1965).

Activity of plant OAT has been detected in mung bean (Bone, 1959), wheat (Kleczhowski and Kretovich, 1960), sunflower (Smith, 1962), peanut (Mazelis and Fowden, 1969), pumpkin (Splittstoesser and Fowden, 1973), squash (Lu and Mazelis, 1975) and pea (Kleczhowski and Kretovich, 1960; Taylor and Stewart, 1981).

A cDNA encoding OAT has recently been isolated from *Vigna aconitifolia* (Delauney *et al.*, 1993). This cDNA was isolated by complementation of an *E. coli proBA* proline auxotroph grown in the presence of ornithine. The gene product was capable of transaminating ornithine to GSA, thereby by-passing the block in GSA synthesis from glutamate in the *proBA* mutant and thus enabling the auxotroph to growth prototrophically on ornithine (Delauney *et al.*, 1993).

On the basis of enzymatic studies conducted in several species (Bone, 1959; Smith, 1962; Taylor and Stewart, 1981), and the presence of a mitochondrial transit peptide in the *Vigna*

OAT cDNA (Delauney *et al.*, 1993), it is likely that OAT is located in mitochondria in plants. Mammalian OAT is also a mitochondrial matrix enzyme which is synthesised with an N-terminal transit peptide that is cleaved during entry into the mitochondrion (Inana *et al.*, 1986). In contrast, OAT from *Saccharomyces cerevisiae* is cytosolic (Degols, 1987).

Examination of the properties of *Vigna* OAT expressed in *E. coli* indicated that *in vitro* the enzyme catalysed an irreversible conversion of ornithine to GSA and was not inhibited by proline (Delauney *et al.*, 1993). Reversibility would suggest that conversion of P5C to ornithine could supply intermediates for citrulline synthesis. Several workers using animal systems have pointed out that P5C is one of only a few metabolites of both the urea and TCA cycles (Hagedorn and Phang, 1983; Wakabayashi and Jones, 1983; Phang, 1985). Biosynthesis and degradation of proline, ornithine and glutamate may occur by single-step interconversions of P5C. It has therefore been proposed that regulation of P5C metabolism and the relative sizes of the P5C pools in the cytosol and mitochondria may be important in controlling relative flux through these pathways and, in addition, subcellular and tissue compartmentalisation of P5C may prevent futile cycling of these amino acids. However, the irreversibility of recombinant *Vigna* OAT, which contrasts with the freely reversible reaction catalysed by OAT in animal systems (Jones, 1985), suggests that an analogous situation may not apply in plants. Therefore, evidence available to date does not support the conversion of P5C to ornithine in plants.

However, equilibrium of the reaction catalysed by mammalian OAT is towards GSA formation (Strecker, 1965). Jones (1985) has argued that although many workers have reported mammalian OAT activity to proceed solely towards GSA, this is because the chemical formation and stability of GSA and its tautomer P5C pull the reaction in this direction *in vitro*. *In vivo*, the concentration of the substrates 2-oxoglutarate and glutamate as well as ornithine and P5C and the strength of their binding to the enzyme may be more important in determining the directionality of the reaction (Jones, 1985). Therefore, the possibility that P5C might regulate flux through the TCA and urea cycles in plants cannot be completely eliminated. Metabolic labelling studies may resolve this issue.

On the basis of examining the levels of P5CS and OAT mRNA transcripts in seedlings of *V. aconitifolia*, Delauney *et al.* (1993) concluded that both pathways contribute to proline synthesis under normal physiological conditions. In plants subjected to salinity stress, OAT

mRNA levels were markedly depressed concomitant with the elevation of P5CS transcript levels. Similarly, in nitrogen-starved plants, reduced levels of OAT mRNA accompanied increased amounts of P5CS transcript (Delauney *et al.*, 1993). This was interpreted to imply that synthesis from glutamate predominates under stress conditions such as high salinity and nutrient limitation. In contrast, synthesis from ornithine assumes prominence under high nitrogen input (Delauney *et al.*, 1993). This model is in keeping with the generally accepted tenet that glutamate is the primary precursor of proline accumulated in response to the imposition of osmotic stress (Boggess and Stewart, 1976). However, it introduces the complication that the relative importance of the proline biosynthetic pathway from glutamate during stress may be affected by the nitrogen status of the plant. The relative contributions of the two pathways in different tissues and during different developmental stages remain to be defined.

In contrast to the findings of Boggess and Stewart (1976) and Delauney *et al.* (1993), Wrench *et al.* (1977) found that in osmotically stressed *Helianthus tuberosus* tubers, arginine the immediate precursor of ornithine, is quantitatively more important than glutamate in proline biosynthesis. A similar situation applies in wilted bean leaves (Stewart and Boggess, 1977). Furthermore, proline arises from both glutamate and arginine in water-stressed *Cyclotella cryptica* cells (Liu and Hellebust, 1976a, 1976b). Clearly, closer examination of the relative contributions of the pathways of proline synthesis from glutamate and ornithine in several species during stress is warranted.

The complexity of proline biosynthesis in plants is increased even further by the possibility that two pathways may exist for the synthesis of proline from ornithine. Radiolabelling experiments conducted by Mestichelli *et al.* (1979) suggest the existence of an alternative pathway involving the intermediates α-keto-δ-aminovalerate and its cyclic tautomer Δ^1-pyrroline-2-carboxylate (P2C). In the conversion of radiolabelled ornithine to proline, the δ-hydrogen atoms were maintained, while the α-hydrogen atoms were lost. This was interpreted to indicate a route via α-keto-δ-aminovaleric acid in addition to the route via GSA. It would implicate the existence of two additional enzymes involved in proline biosynthesis, namely ornithine-α-aminotransferase (EC unassigned) and Δ^1-pyrroline-2-carboxylate reductase (P2CR; EC 1.5.1.1). This proposal has subsequently been challenged by Adams and Frank (1980) who asserted that isotope exchange reactions prior to conversion of ornithine to proline as well as kinetic isotope effects may have led to

misinterpretation of the data.

Although Delauney *et al.* (1993) did not recover ornithine-α-aminotransferase and P2CR cDNAs in their search for cDNAs encoding enzymes of proline synthesis in higher plants, the complementation strategy used by these workers was not suitable for cloning these genes. Cloning of ornithine-δ-aminotransferase was possible because the *proBA* deficient *E. coli* strain employed harboured a functional *proC* gene encoding P5CR. Cloning of an ornithine-α-aminotransferase by the same approach would require a *proBA* deficient strain harbouring a gene encoding a functional P2CR. Since *E. coli* does not possess P2CR activity (Leisinger, 1987), a complementation strategy is not suitable for screening for the gene encoding ornithine-α-aminotransferase. Likewise, a P2CR cDNA would not be expected to complement *proC* mutants of *E. coli*. Therefore, although molecular evidence is currently lacking, the possibility that a parallel pathway of proline synthesis from ornithine exists, via P2C, cannot be yet be completely excluded. If correct, it would introduce the possibility of a further level of regulation of proline biosynthesis via control of ornithine-α-aminotransferase or P2CR at either the level of enzyme activity or gene expression.

Although no evidence for the pathway of proline synthesis from ornithine via P2C was found in rat intestinal mucosa (Wakabayashi and Jones, 1983), P2CR has been reported to have been partially purified from rat kidney (Meister, 1962). Other early studies of proline metabolism also implicated P2C as an intermediate in proline oxidation (Blanchard *et al.*, 1944; Ling and Hedrick, 1964). Nevertheless, subsequent workers have failed to validate these findings. Therefore, until more convincing evidence in favour of P2C as an intermediate in proline metabolism is presented, the reaction catalysed by P5CR remains the only committed step in proline biosynthesis. Irrespective of whether or not proline may also arise from ornithine via a P2C intermediate, P5CR is likely to be of critical importance in proline biosynthesis. Its location at the branch point between P5C synthesised from both glutamate and ornithine suggests a possible regulatory role for P5CR in controlling flux from the two routes.

In mammalian tissue, ornithine inhibits P5CS activity (Henslee *et al.*, 1983). Although this possibility was not investigated by Hu *et al.* (1992), proline had no effect on activity of recombinant *Vigna* OAT activity (Delauney *et al.*, 1993). This contrasts with an earlier report (Splittstoesser and Fowden, 1973) that OAT activity in pumpkin seed is inhibited by

proline. However, Lu and Mazelis (1975) also reported no inhibition of *Cucurbita pepo* OAT activity by proline, but inhibition by valine, leucine and isoleucine. Activity of the recombinant *Vigna* OAT was also inhibited by valine (Delauney *et al.*, 1993).

Although *E. coli* synthesises proline exclusively from glutamate (Leisinger, 1987), many Gram-positive bacteria also synthesise proline from ornithine. For example, *Clostridium botulinium* and *C. sporogenes* use ornithine cyclase (E.C. 4.3.1.12) (Muth and Costilow, 1974). Others, such as *Bacillus subtilis* and *Pseudomonas aeruginosa*, rely on acetylornithine-δ-transaminase (E.C. 2.6.1.11) for biosynthesis of proline from ornithine (Albrecht and Vogel, 1964). Members of the genus *Agrobacterium* convert ornithine into proline using ornithine cyclodeaminase (Schindler *et al.*, 1989). However, to date none of these activities have been detected in higher plants.

The diversity of metabolic routes from ornithine to proline among these microbial systems suggests that it may be more difficult to extrapolate directly from these systems to plants than was originally believed. Although proline synthesis from ornithine does not appear to be important under conditions of stress, much remains to be learnt about the contribution of ornithine to the proline pool and the regulation of flux through the pathways from both precursors. To date, molecular studies suggest that the choice of pathway followed depends on the nitrogen status of the plant. High nitrogen input induces OAT gene expression possibly via an accumulation of ornithine or arginine (Delauney *et al.*, 1993). A similar induction of OAT gene expression by exogenous arginine occurs in yeast (Brandriss and Magasanik, 1980). However, these data in plants still need to be corroborated by assaying enzyme activity and measuring the metabolic flux through the respective pathways *in vivo* in order to establish the role of the metabolites under different physiological conditions.

2.2.3 Proline degradation

Besides enhanced synthesis, proline accumulation is also be controlled at the level of proline degradation. Nevertheless, although inhibition of proline oxidation undoubtably contributes to proline accumulation during stress, it is not sufficient in itself to account for the observed rates of accumulation (Elthon and Stewart, 1984).

In plants, the oxidation of proline to glutamate occurs within mitochondria (Boggess *et al.*, 1978; Huang and Cavalieri, 1979; Elthon and Stewart, 1981; Sells and Koepe, 1981; Elthon and Stewart, 1982). This is also the case in other eukaryotes studied, including *Saccharomyces cerevisiae* (Krzywicki and Brandriss, 1984; Brandriss and Krzywicki, 1986; Wang and Brandriss, 1987) and animals such as rat (Brunner and Neupert, 1969; Small and Jones, 1990) and *Drosophila melanogaster* (Hayward *et al.*, 1993).

Proline oxidation is catalysed by the sequential action of two enzymes, proline oxidase (L-proline : O_2 oxidoreductase; EC 1.4.3), and P5C dehydrogenase (Pyrroline-5-carboxylate : $NAD(P)^+$ oxidoreductase; EC 1.5.1.12). Proline oxidase converts proline to P5C. The oxidation of P5C to glutamate is catalysed by P5C dehydrogenase (Figure 2.2). In plants, both proline oxidase (Boggess *et al.*, 1978) and P5C dehydrogenase (Stewart and Lai, 1974; Boggess *et al.*, 1975) are bound to the matrix side of the inner mitochondrial membrane (Elthon and Stewart, 1981, 1982). In prokaryotes, such as *Escherichia coli* (Scarpulla and Soffer, 1978) and *Salmonella typhimurium* (Menzel and Roth, 1981), proline oxidase is associated with the plasma membrane. In *E. coli* (Frank and Ranhand, 1964) and *S. typhimurium* (Ratzkin and Roth, 1978), P5C dehydrogenase is believed to be physically associated with proline oxidase in a protein complex.

As with the enzymes from *E. coli* (Frank and Ranhand, 1964), the yeast *Hansenulla subpelliculosa* (Ling and Hedrick, 1964) and mammals (Johnson and Strecker, 1962; Strecker, 1971) plant proline oxidase is an oxygen-dependent flavoprotein (Huang and Cavalieri, 1979; Elthon and Stewart, 1982). Although NAD is the preferred electron acceptor for P5C dehydrogenase, NADP may also be used, although its use yields lower rates of activity (Stewart and Lai, 1974).

Proline oxidation is believed to be responsible for transferring electrons into the first portion of the electron transport chain (Elthon and Stewart, 1981, 1982). The oxidation of both proline and P5C in maize mitochondria is inhibited by rotenone (Elthon and Stewart, 1981). This indicates that electrons from these substrates enter the respiratory chain prior to at least one of the rotenone sensitive iron-sulphur proteins (Elthon and Stewart, 1981). Similarly, respiratory deficient yeast strains contain inactive proline oxidase because of a non-functional electron transport chain (Wang and Brandriss, 1987). However, *in vitro* activity can be detected if artificial electron acceptors are included in the assay (Wang and

Brandriss, 1987). Two P5C dehydrogenases have been found in plant mitochondria. One oxidises P5C derived from proline and the other oxidises P5C from ornithine (Elthon and Stewart, 1982). The latter enzyme forms a complex with mitochondrial OAT (Elthon and Stewart, 1982).

Activity of plant proline oxidase is reduced during salt stress (Stewart *et al.*, 1977; Rayapati and Stewart, 1991). This is most probably attributable at least in part to an alteration in the permeability of mitochondrial membranes concomitant with water stress and salinisation (Miller *et al.*, 1971; Nir *et al.*, 1971; Jolivet *et al.*, 1990). However, the oxidation of proline is more sensitive to water stress than is the oxidation of other mitochondrial substrates (Sells and Koepe, 1981). Results presented by these workers suggest that the decreased oxidation of proline observed in mitochondria from water-stressed maize seedlings is the result of an alteration in membrane integrity that uniquely affects the enzymes involved in proline degradation (Sells and Koepe, 1981).

A role for the enzymes catalysing proline degradation in recovery from stress has already been discussed in Section 2.1.2.4. Proline oxidation is believed to provide energy and carbon skeletons for a number of processes. It is likely that proline oxidation replenishes TCA cycle intermediates to keep them at the appropriate levels needed for the high energy demands upon relief from stress.

To date, genes encoding proline oxidase or P5C dehydrogenase have not been isolated from plants. Nevertheless, isolation of the *PUT1* and *slgA* genes encoding proline oxidase in *S. cerevisiae* (Wang and Brandriss, 1987) and *D. melanogaster* (Hayward *et al.*, 1993) respectively, has been reported. In both organisms, proline oxidase is a nuclearly encoded, cytoplasmically synthesised and mitochondrially imported enzyme. The *PUT2* gene encoding P5C dehydrogenase in *S. cerevisiae* has also been cloned and characterised (Brandriss, 1983; Krzywicki and Brandriss, 1984). In yeast, expression of P5CR dehydrogenase is regulated by proline at the transcriptional level. Both *PUT1* and *PUT2* are regulated by a control element encoded by the *PUT3* gene (Brandriss and Magasanik, 1979a, 1979b; Brandriss, 1987). Mutations in the *PUT3* gene result in constitutive expression of both proline oxidase and P5C dehydrogenase (Brandriss and Magasanik, 1979b; Brandriss, 1987). The existence in plants of an analogous regulatory factor resembling the *PUT3* gene product has not been demonstrated to date. However, in this context it is interesting to note that the rapid decline

in proline oxidation by mitochondria isolated from stressed maize seedlings after only very modest water stress does not occur when osmotic stress is imposed on isolated mitochondria from non-stressed seedlings (Sells and Koepe, 1981). This suggests that such a regulatory element in plants may account for the differential regulation of proline oxidation under stressed and non-stressed conditions.

Mutants of *Arabidopsis thaliana* resistant to the proline analogue azetidine-2-carboxylic acid have been isolated (Verbruggen and Jacobs, 1987; Khan and Lehle, 1991). It is possible that these may be deficient in proline oxidase activity. Alternatively, they may be deficient in a mitochondrial proline transporter needed for transport of cytosolic proline into mitochondria. In light of the fact that proline uptake into the mitochondrial matrix is both stereospecific and reversibly sensitive to sulfhydryl reagents (Cavalieri and Huang, 1980) in addition to being sensitive to uncouplers (Elthon and Stewart, 1981), the uptake of proline into mitochondria is likely to be a carrier-mediated, energy-dependent process. Proline oxidation in such mutants may therefore be blocked either at the transport or the dehydrogenase level. Only one of the three *Arabidopsis* mutants resistant to azetidine-2-carboxylic acid possessed significantly elevated levels of proline (Verbruggen and Jacobs, 1987). Mutants resistant to toxic proline analogues are likely to be useful tools in future investigations into the importance of a decline in proline oxidation in plants experiencing stress. Furthermore, they may indicate whether transport of cytoplasmic proline into mitochondria is specifically affected by relatively short periods of water deprivation. They are also likely to be of value in the isolation of plant genes involved in proline oxidation.

Inactivation of proline oxidase is likely to be an important target in future attempts to overproduce proline in plants. Bloom *et al.* (1983) reported that an *E. coli* strain unable to degrade proline owing to a chromosomal deletion in the gene encoding proline oxidase was capable of yielding 27 g proline per litre, with 40% conversion of glucose to proline. However, this strain also contained a vector carrying a *proBA* operon encoding a feedback resistant γ-GK and a *proC* gene. It is therefore not known at what level the proline overproduction was facilitated.

In *Saccharomyces cerevisiae*, proline induces both proline oxidase and P5C dehydrogenase (Brandriss and Magasanik, 1979a, 1979b). The genes encoding both proline oxidase (Wang and Brandriss, 1987) and P5C dehydrogenase (Brandriss, 1983) are regulated by proline at

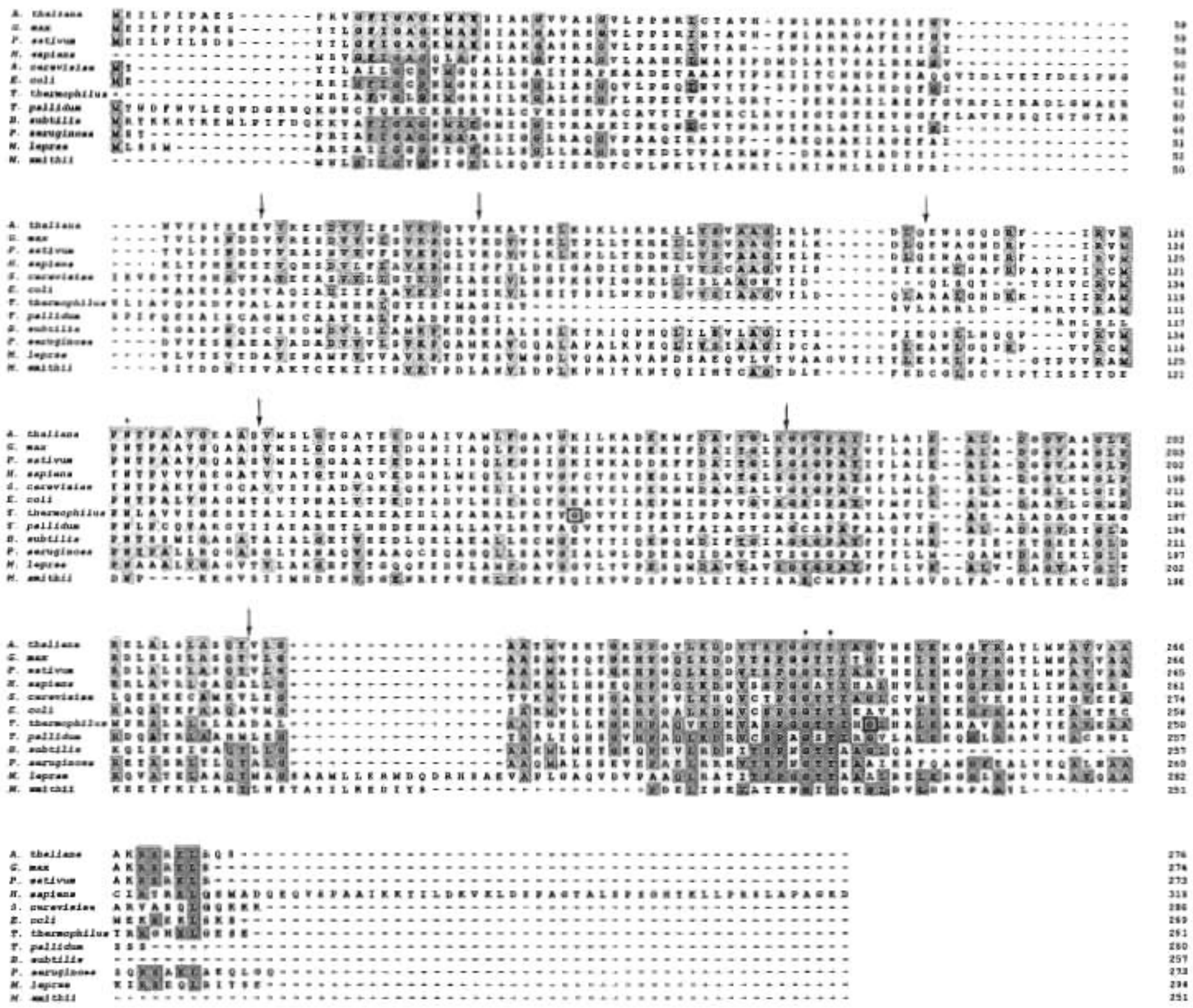


Figure 4.13 : Alignment of twelve P5CR sequences. Alignment was based on the blocks selected in Figure 4.12 using the computer program MACAW (Schuler *et al.*, 1991). Amino acid residues conserved in at least six of the twelve sequences aligned are shaded. Two glycine residues known to be critical for function of P5CR from *T. thermophilus* (Hoshino *et al.*, 1994) are boxed. Positions of introns in the corresponding *Arabidopsis* gene encoding P5CR (Verbruggen *et al.*, 1993) are indicated by arrows. Three residues conserved in all known P5CR sequences are indicated by asterisks above the aligned sequences.

2.3 *Arabidopsis* as a model system in plant molecular studies

Traditionally, plant physiologists and geneticists have studied a wide variety of plant species. These have been selected mainly on the basis of their suitability to specific types of research or because of the economic importance of a particular species concerned (Koorneef, 1991). In particular, tomato (*Lycopersicon esculentum*) has been one of the leading models for genetic studies in dicotyledonous plants (Hille *et al.*, 1989). Rice (*Oryza sativa*) and maize (*Zea mays*) are commonly-studied representative monocotyledonous species. Tobacco (*Nicotiana tabacum*) has also been the subject of numerous studies, owing mainly to its amenability to cell and tissue culture. Despite their agricultural value, representative legumes such as pea (*Pisum sativum*) and soybean (*Glycine max*) have received less attention, largely because of their problematic transformability. However, a feature common to all of these systems is a high degree of genetic complexity (Table 2.5). This causes difficulty in the application of genetic techniques to these plants. The hexaploid nature of wheat (*Triticum aestivum*) complicates mutational analysis (Koorneef, 1991).

Currently a major focus in biology is the accumulation of genetic knowledge at the molecular level. The stated goals of the U.S. Genome Project include the production of fifty megabases of DNA sequence data per year by 1998 and the identification and correlation of genes in humans and model organisms (Collins and Galas, 1993). During the past decade, the challenge of obtaining an integrated view of the molecular basis of plant growth and development has resulted in a reduction in the number of plant species studied. The power of focusing on a single species as a model system of a group of organisms has been shown by the multitude of scientists working in other fields with organisms such as *Escherichia coli*, *Drosophila melanogaster*, *Caenorhabditis elegans*, mouse and man. Increasingly, the small crucifer *Arabidopsis thaliana* (L.) Heynh. (Figure 2.4) is being established as the analogous model system in contemporary plant biology. *Arabidopsis thaliana* is a winter annual endemic to the moderate temperature zones of the world (Rédei, 1992).

A large number of projects currently underway throughout the world are involved in a concerted effort to characterise the *Arabidopsis* genome and patterns of its expression. The most common approach being adopted in the resolution of the latter question is nucleotide sequencing of the termini of randomly selected cDNA clones isolated from different plant

Table 2.5 : Genomic complexity of representative plant species compared with a virus, a chloroplast and human. (Adapted from Grierson and Covey, 1988).

Species	DNA content (pg per genome)	Ploidy	Number of bp in genome
Cauliflower mosaic virus	8.4×10^{-4}		8 024
Pea chloroplast	1.3×10^{-4}		1.24×10^{5}
Escherichia coli	4.0×10^{-3}		3.90×10^{6}
Arabidopsis thaliana (thale cress)	0.1	$2x = 10$	0.93×10^{8}
Glycine max (soybean)	0.9	$4x = 40$	8.69×10^{8}
Hordeum vulgare (barley)	5.6	$2x = 14$	5.40×10^{9}
Lilium davidii	43.2	$2x = 24$	4.17×10^{10}
Lycopersicon esculentum (tomato)	0.75	$2x = 24$	7.24×10^{8}
Nicotiana tabacum (tobacco)	3.9	$4x = 48$	3.70×10^{9}
Oryza sativa (rice)	1.0	$2x = 24$	9.65×10^{8}
Pisum sativum (pea)	4.9	$2x = 14$	4.73×10^{9}
Solanum tuberosum (potato)	2.1	$4x = 48$	2.03×10^{9}
Spinacia oleracea (spinach)	1.0	$2x = 12$	9.65×10^{8}
Triticum aestivum (hexaploid wheat)	17.3	$6x = 42$	1.67×10^{10}
Zea mays (maize)	3.9	$2x = 20$	3.76×10^{9}
Human	6.0	$2x = 46$	5.79×10^{9}

Figure 2.4 : ***Arabidopsis thaliana*** **(L.) Heynh.**

organs and during different stages of development or different physiological states. These partially resolved nucleotide sequences, commonly referred to as expressed sequence tags (ESTs) are used in homology searches of one or more of the nucleotide databases in an attempt to establish their identity. They may subsequent be used to define sequence tagged sites (STS's) in transcript maps of the *Arabidopsis* genome or for a comprehensive survey of genes during different stages of the life cycle. To date, the ends of more than 1152 randomly selected cDNA clones from *Arabidopsis* have been sequenced and a large proportion of these cDNAs have been identified (Höfte *et al.*, 1993). These are distributed over a broad spectrum of metabolic pathways and constitute a comprehensive survey of expressed genes. Similar efforts to characterise expression using ESTs have been reported in human (Adams *et al.*, 1991, 1992, 1993; Okubo *et al.*, 1992), *Caenorhabditis elegans* (McCombie *et al.*, 1992; Waterston *et al.*, 1992), mouse (Höög, 1991), maize (Keith *et al.*, 1993) and *Brassica napus* (Park *et al.*, 1993).

A number of factors qualify the suitability of *Arabidopsis* as a tool in plant genetics. Its nuclear genome (n=5) is the smallest known among higher plants, containing about 0.7-1 x 10^8 bp (Leutweiler *et al.*, 1984). As shown in Table 2.5, all other flowering plants for which similar measurements have been made show haploid genome sizes from several-fold to almost five hundred-fold higher than the *Arabidopsis* measurement. Furthermore, owing to the near-absence of interspersed repetitive DNA, genomic redundancy is very low (Meyerowitz and Pruitt, 1985). What selective advantage, if any, is served by such an unusual genome size and organisation for a flowering plant is uncertain. However, the practical advantage of working with minimal amounts of genomic DNA and repetitive sequences in molecular biological experiments, such as gene cloning and chromosome walking, is obvious.

A number of other attributes operate in favour of *Arabidopsis* as an experimental system. These include high seed yield, easy cultivation in limited space, rapid development, easy crossing with full fertility of the hybrids and the availability of several efficient transformation procedures (Feldman and Marks, 1987; Valvekens *et al.*, 1988). Because of its small size, short life cycle and high fertility, *Arabidopsis* is a choice organism for the isolation of both spontaneous and induced mutations. A wide range of metabolic, developmental, hormonal and disease-resistant mutants of *Arabidopsis* have been isolated (Rédei and Koncz, 1992; Anderson, 1994). Mutants remain one of the major tools enabling

the cloning of certain plant genes. This is possible using molecular techniques such as chromosome walking, tagging, deletion cloning or differential screening of mutant and wild-type mRNAs.

Differential screening of specific mRNAs and the use of heterologous probes, often from organisms outside of the plant kingdom, are widely used for plant gene isolation. This applies particularly to those involved in primary metabolism. However, this approach is not tenable for the isolation of all plant genes, especially those whose products are unique to plants. An important advantage of obtaining a mutant is that one can directly observe what effect the gene has on plant growth by comparing the mutant and wild-type phenotypes.

Several different ecotypes of *Arabidopsis* are currently available (Anderson, 1994). The two most commonly used in molecular studies are the ecotypes Landsberg *erecta* and Columbia. A number of markers, loci and traits which differ between these two ecotype have been identified. For example, the *erecta* mutation carried by Landsberg *erecta* confers a short and erect stature on this plant (Rédei, 1992). Columbia has the wild-type gene and is therefore taller and less compact. The two ecotypes differ in their susceptibility to the fungal pathogen *Peronospora parasitica*. Columbia shows the sensitive phenotype, whereas Landsberg *erecta* displays resistance to the downy mildew (Parker *et al.*, 1993). Recently, Aufsatz and Grimm (1994) demonstrated that a pathogen-inducible gene from *Arabidopsis* is expressed in an ecotype-specific manner. Both ecotypes display differences in their drought-adaptive strategies, with Columbia being more drought tolerant than Landsberg *erecta* (Vartanian *et al.*, 1994).

To date, less attention has been devoted to stress responses of *Arabidopsis* than to deciphering the structure of the genome and characterisation of basic housekeeping genes of this plant. In the past, a common avenue towards understanding responses of plants to environmental stresses has been to compare traits of plants from extreme environmental regimes with those endemic to more moderate habitats. However, it is unlikely that many of the physiological and morphological adaptations associated with survival under conditions of extreme stress are applicable to crop plants grown in fairly mesic environments. In this respect, *Arabidopsis* seems well suited as a model system for stress responses in agriculture. It may be argued that the short growing season of *Arabidopsis* functions as an attribute of drought escape. Unfortunately, to date the adaptive behaviour of this species to

hyperosmotic stress has not been described in any significant detail. Further work in this respect will be necessary to establish exactly how effectively *Arabidopsis* serves as a model species in drought tolerance research.

However, it is likely that many principles of plant development and growth are common to all higher plants. Proline accumulation is a good example of a response that appears to have been highly conserved throughout the plant kingdom (Table 2.1). To date, a range of osmotic stresses have been shown to increase proline levels up to twenty-fold in *Arabidopsis* (Chiang and Dandekar, 1991). Verbruggen *et al.* (1993) reported that following salinisation, proline may account for up to 20% of the free amino acid pool in *Arabidopsis* plantlets. This indicates that this species is a typical proline accumulator. The wealth of genes and mutants that continue to be isolated from *Arabidopsis* make this system particularly attractive to elucidate the relationships between proline biosynthesis and the rest of intermediary metabolism.

Obviously, there are considerations which caution against elimination of other higher plant model systems. Acquisition of specific experimental data focusing on the particular crop in question will always be important. Furthermore, a disadvantage for certain types of research is the small size of *Arabidopsis*, particularly of its seeds and seedlings. For seed development studies, maize is a more attractive alternative (Sheridan, 1988). Studies on floral development and pigmentation are more suited to use of *Petunia* or *Antirrhinum* (Koorneef, 1991). However, having at least one reference plant species which has been extensively characterised is likely to be extremely useful. The ease of detecting a particular gene from one species in another using molecular probes should enable efficient switching between species. In conclusion therefore, although *Arabidopsis* is not an economically important organism, its anthropocentric value rests on the depth of scientific information that can be gathered by its use as a model plant system.

2.4 Fundamentals of plant gene structure and regulation

Over the past twenty years, the striking advances in nucleic acid analysis in general and in DNA cloning and sequencing in particular have made available a great deal of data on the primary structure of several plant genes. This has indicated that many features of plant gene structure are identical to those found in other living organisms (Lefebvre, 1990). Many concepts in plant gene structure and regulation derive from animal, yeast and bacterial systems. This interaction between plant and non-plant research has been and continues to be extremely rewarding. Plant genes use the same genetic code, are split by introns and use regulatory mechanisms that are similar in principle to those characterised in yeast and animal systems.

In basic terms, all genes comprise a coding region surrounded on both sides by non-translated regions. The coding region or open reading frame (ORF) specifies the amino acid sequence of the gene product. Regions preceding the ORF on the 5' end are responsible for regulation of transcription from the DNA template (Lefebvre, 1990). However, eukaryotic genes differ in several important respects from their prokaryotic counterparts. In particular, they are not arranged in operons, and their coding regions are often interrupted by introns. Therefore, primary mRNA transcripts often require post-transcriptional splicing in order to form mature functional mRNAs (Lefebvre, 1990). Although plant genes resemble animal genes more closely than bacterial genes (Lefebvre, 1990), a number of subtle differences distinguish the two. This brief review will deal primarily with aspects of plant gene structure and potential regulatory mechanisms which differ from those found in animal and bacterial systems.

2.4.1 Upstream regulatory regions

Our current understanding of eukaryotic gene regulation implicates the involvement of a number of factors and non-transcribed DNA sequences which may act in two ways to regulate the expression of the gene. Gene activation apparently results from an interplay between *trans*-acting protein factors and *cis*-acting DNA elements (Lefebvre, 1990). *Cis*-acting DNA elements act directly by being associated with the gene at its 5' end. Activity

of *cis*-acting elements is dependent on the binding of specific transcriptional factors. These *trans*-acting factors that bind to *cis*-acting regulatory elements are encoded by genes that are often located on a different chromosome from the target gene (Lefebvre, 1990). Diffusible *trans*-acting factors may act on one or a number of target sites simultaneously (Lefebvre, 1990).

Cis-acting elements have been operationally classified as promoters (usually found within 100 bp upstream of the transcription start site) and enhancers. Eukaryotic promoters consist of two major elements. The Hogness or TATA box or a functionally related sequence is usually located approximately 30 bp upstream of the transcriptional start position (Kuhlemeier, 1992). This site directs RNA polymerase II to the correct transcriptional start site. Centred around 75 bp upstream of the transcription initiation site are one or more elements responsible for regulating the frequency of initiation (Lefebvre, 1990). This second class of *cis*-acting DNA elements includes binding sites for proteins that interact directly with the RNA polymerase complex. These *cis*-acting elements can function at variable distances from the TATA box (Kuhlemeier, 1992).

In animal genes, the CAAT box is a common element found at this position. However, CAAT boxes are often absent from plant genes (Lefebvre, 1990). Certain of the genes encoding the small subunit of ribulose-1,5-bisphosphate carboxylase have no apparent requirement for CAAT or functionally equivalent sequences (Morelli *et al.*, 1985; Kuhlemeier *et al.*, 1987). However, cereal genes have a consensus sequence, the CATC box, 90 bp upstream of the transcription initiation site. This element may substitute for the CAAT box commonly found in animal genes (Kreis *et al.*, 1986). Often the *cis*-acting elements upstream of the TATA box are regulatory. They frequently only enhance or repress transcription under specific cellular or environmental conditions. A classical example is the heat-shock element, which when fused upstream of the TATA box of a reporter gene, increases transcription only at high temperatures (Pelham and Bienz, 1982). Protein footprint studies have resulted in the identification of proteins that bind to the heat-shock element and act as transcription factors. Such heat-shock transcription factors have been cloned from plants (Vierling, 1991).

Ironically, comparatively little is known about plant TATA boxes or genes without TATA boxes (Kuhlemeier, 1992). This contrasts with the multitude of reports that continue to be

made which describe other upstream regulatory elements. Although the TATA box is the binding site for RNA polymerase and its associated factors, other *cis*-acting regulatory elements can bind a variety of DNA-binding proteins. Some of the upstream binding proteins are probably general transcription factors which are present in all or at least most cell types and active under most if not all conditions (Kuhlemeier, 1992). Fairly "general" transcription factors appear to be responsible for the initiation of a basal level of transcription at promoter sites. Other *trans*-acting factors superimpose more specific changes in the expression of particular genes in response to certain environmental or cellular stimuli and account for the finely-tuned regulation of individual genes.

Enhancer elements are another feature of eukaryotic genes which may be involved in the regulation of transcription. Enhancer-binding factors impose a further level of regulation on the basic transcriptional apparatus, thereby modifying the rate of initiation and efficiency of transcription (Lefebvre, 1990). Enhancers are 100-200 bp in length and can act in either orientation and at a considerable distance (up to 30 kb upstream or downstream) from the promoter of the target gene (Lefebvre, 1990). Since they are found on the same chromosome as the target gene, they are *cis*-acting elements. However, they may be 5' or 3' to the transcriptional start and usually act by stimulating the nearest promoter (Lefebvre, 1990).

One of the best studied plant promoters is the cauliflower mosaic virus (CaMV) 35S promoter. This very strong viral promoter produces the 35S genomic RNA of the cauliflower mosaic virus. Early experiments revealed that fusion of approximately 1000 bp of CaMV 35S promoter DNA to various reporter genes resulted in high levels of expression of these genes (Kuhlemeier, 1992). Levels of expression were not affected by various endogenous and environmental cues such as hormones, heat shock and light (Odell *et al.*, 1985; Kay *et al.*, 1987; Ow *et al.*, 1987). Furthermore, fusion of the CaMV 35S promoter to reporter genes results in apparently constitutive expression of the gene in all plant tissues (Kuhlemeier, 1992).

A central problem in research on plant gene regulation is the mechanism of cell- or tissue-specific gene expression (Benfey and Chua, 1989, 1990). Any higher plant comprises a range of cell-types. A division of labour among the various plant tissues ensures that each cell type performs different physiological functions. In any single cell type, only a fraction of the full complement of genes are expressed at any one time. Since all cells contain an

identical complement of DNA, a mechanism must exist which ensures the specific and ordered expression of only those genes of importance in a particular tissue. Maintenance and slight alterations of this differential gene expression are probably the primary determinants of the various cellular phenotypes observed in multicellular organisms.

Current evidence suggests that common sets of cell- or tissue-specific *trans*-acting factors provide the basis for tissue-specific expression. Investigation of the transcriptional regulation of the prototypical plant promoter CaMV 35S has revealed that it is modular and comprised of several *cis* elements, each of which confers tissue-specific expression (Benfey *et al.*, 1989; 1990a). The apparent constitutive expression of the CaMV 35S promoter results from the combinatorial and synergistic interaction of various *cis* elements and not from a single *cis*-acting element that confers constitutive expression (Benfey *et al.*, 1990b). This modular constitution of *cis* elements has been proposed to apply to all plant promoters, regardless of their regulation. It is also likely to serve as a paradigm for elucidating the temporal regulation of gene expression during development and the response of gene activity to extracellular factors, such as environmental extremes.

However, studies on the regulation of plant gene expression are still in their infancy. Transcriptional regulation is likely to be determined not only by the intrinsic properties of the transcription factor and its cognate binding site, but also by complex interactions of multiple factors and multiple binding sites (Kuhlemeier, 1992). A single factor may have different affinities for different binding sites and/or may bind cooperatively. Factors may compete for single or overlapping binding sites, which may result in changed interactions with the RNA polymerase complex (Kuhlemeier, 1992). Recently, Neuhaus *et al.* (1994) provided evidence that the high level of expression of *as-1* element of the CaMV 35S promoter in roots and reduced expression in leaves or cotyledons may be attributed to inability of *as-1* to compete for a limiting amount of its cognate transcription factor(s) present in leaves or cotyledons. Regulation of gene expression may therefore be either at the level of the gene itself or at the level of production of the cognate transcription factor(s). Furthermore, many *trans*-acting factors may not bind to DNA elements directly, but rather interact with the transcriptional apparatus via so-called bridging proteins which bear contact sites for RNA-polymerase and upstream DNA-binding elements (Lewin, 1990).

A growing number of putative plant transcription factors continue to be characterised. Many are structurally related. Most fall into the class of the so-called bZip proteins, which contain a leucine zipper dimerisation motif and a basic DNA-binding domain (Kuhlemeier, 1992). A high degree of similarity in the basic DNA-binding domain has been observed in factors isolated by virtue of their affinity to *cis*-acting elements from genes regulated by cues as different as light and abscisic acid (Kuhlemeier, 1992). Detailed knowledge of how *in vitro* binding specificity occurs is accumulating. However, a current challenge remains the elucidation of how these factors bring about the very precise but diverse regulation of the target genes which they control.

2.4.2 The open reading frame

The open reading frame (ORF) of a gene specifies the amino acid sequence of the gene product. It comprises a number of codons. A codon is a triplet base sequence specifying a particular amino acid. The triplet base code of the nucleic acid sequence (genetic code) in the form of codons governs the sequential joining of amino acids to make the gene product.

2.4.2.1 Codon usage

Since there are 64 possible codons and only 20 amino acids, most amino acids within a gene product are encoded for by several codons. Only tryptophan and methionine are encoded for by single codons. The remaining 18 amino acids are encoded by two to six synonymous codons. The genetic code is therefore highly redundant. However, this redundancy is not random. With the exception of serine, leucine and arginine, all synonymous codons differ only in the third base of the triplet. Pairs of codons of the general form XYC and XYU (where X and Y represent any base) always code for the same amino acid.

The choice among synonymous codons in both prokaryotic and eukaryotic genes is distinctly non-random (Bennetzen and Hall, 1982; Ikemura, 1982; Sharp *et al.*, 1988; Murray *et al.*, 1989; Campbell and Gowri, 1990). Most, if not all, mRNA-coding genes show a bias in the choice of which of several degenerate triplets are used to code for a

particular amino acid. Codons usage data, if analysed using the appropriate statistics, may yield a wide range of information about an ORF. Depending on the species from which the DNA sequence derives, it may be possible to predict whether an ORF is indeed likely to be a gene (States and Gish, 1994). Furthermore, on the basis of the codon usage pattern of an ORF, it may be possible to speculate about the normal level of expression of the gene, its recent evolutionary history or even its chromosomal location (Lloyd and Sharp, 1992).

A useful measure of general codon usage bias is the effective number of codons used in a gene, N_c (Wright, 1990). The value of N_c can range from 20, in the case of extreme bias where one codon is used exclusively for each amino acid, to 61 when the use of all alternative codons is equally likely for a single amino acid (Wright, 1990). The parameter N_c therefore represents a simple measure of how far the codon usage of a gene departs from equal usage of synonymous codons.

Synonymous codon usage patterns vary at two levels (Ikemura, 1985; Sharp *et al.*, 1988). Firstly, codon usage differs among species. However, fairly closely related species such as *Escherichia coli* and *Salmonella typhimurium*, or *Homo sapiens* and *Bos taurus* are generally quite similar (Sharp *et al.*, 1988). In a study of 207 plant genes, Murray *et al.* (1989) demonstrated that in the plant kingdom, there is also a clear distinction between monocotyledonous plants and dicotyledonous plants in the pattern of relative use of synonymous codons. Codon usage in the two taxonomic groups differs primarily in the use of G or C in the degenerate third base (Murray *et al.*, 1989). The nuclear and chloroplast genomes of plants also have different coding strategies (Murray *et al.*, 1989). Secondly, codon usage varies, sometimes considerably, within genes from the same genome (Bennetzen and Hall, 1982; Murray *et al.*, 1989).

Using a multivariate method to analyse codon usage in many mRNAs, Grantham *et al.* (1980, 1981) originally demonstrated clear similarities of synonymous codon usage between different protein genes of the same or taxonomically related genomes and dissimilarities between genes of taxonomically distinct genomes. It is now well established that the non-random choice among synonymous codons can mostly be attributed to the availability of isoaccepting tRNAs within a cell. In both *E. coli* (Bennetzen and Hall, 1982) and *Saccharomyces cerevisiae* (Bennetzen and Hall, 1982; Ikemura, 1982), preferred codons have been shown to be highly homologous to the major isoaccepting tRNA species.

There is considerable variation in the extent to which any given gene uses only preferred codons or instead draws indiscriminately from the entire set of 61 sense codons. It has been shown in *E. coli* and *S. cerevisiae* that genes which are strongly expressed are more biased in their codon usage than genes with a lower level of expression (Bennetzen and Hall, 1982). There is also a strong positive correlation between the abundance of an isoaccepting tRNA species and the occurrence of the corresponding synonymous codon (Bennetzen and Hall, 1982). This suggests that codon usage frequency is a means of modifying translation in these unicellular organisms. A similar situation appears to apply in plants. Investigation of the codon preference of genes encoding highly abundant proteins such as chlorophyll a/b binding protein and the small subunit of ribulose-1,5-bisphosphate carboxylase revealed that these highly expressed genes are more restricted in their codon usage than plant genes in general (Murray *et al.*, 1989). Therefore, natural selection appears to have shaped the pattern of codon usage bias in genes.

Two explanations for the physiological basis of biased codon usage have been proposed by Bennetzen and Hall (1982). The most obvious explanation involves translation rate of mRNA for abundant proteins. Since these proteins are required at high intracellular levels, it could be presumed that there is a selective pressure to translate their mRNAs rapidly and repetitively. Since the concentration of charged cognate tRNA governs the time required to add an amino acid opposite each codon, rapid translation is favoured by the use of triplets for abundant tRNAs. Therefore, in genes encoding proteins required in large amounts, substitution of triplets for high abundance tRNAs by triplets for low abundance isoacceptors might be deleterious. Continued selection for a high output of the particular protein will act to retain a set of preferred codons within the gene (Bennetzen and Hall, 1982). Conversely, speed of translation has little selective value for low abundance proteins or those which do not show a rapid increase in levels over a short period. As a result, mutations to triplets decoded by low abundance tRNAs would not be strongly selected against. Consequently patterns of codon usage in lowly expressed genes are generally more random (Bennetzen and Hall, 1982). However, it is important to remember that codon usage bias is but one of many options available to ensure high output of gene product. Mutations which affect transcription rate or mRNA stability, or rearrangements to augment the number of gene copies are other possible ways of ensuring a high output of the gene product.

Bennetzen and Hall (1982) also proposed another physiological basis for biased codon usage. This involves the deleterious situation which would arise should codons corresponding to rare tRNA isoacceptors be used in mRNAs of high abundance. Use of a rare isoaccepting tRNA species in coding for abundant proteins would draw heavily on the pool of that species. Undoubtably, as a large fraction of the available tRNA molecules would be discharged and the pool depleted, all ribosomes translating mRNAs containing the relevant codons would stall. This block in translation might increase the risk of early termination and/or translational error. In contrast, occasional use of a low abundance aminoacylated tRNA by several genes making proteins of low abundance would not deplete the pool of the rare isoaccepting tRNA substantially. Consequently, strong selection pressure against use of the codon in question would not operate in these genes (Bennetzen and Hall, 1982). Overall, from these two explanations it was deduced that deviation from the observed pattern of codon bias would result in lowered fitness for the organism. Codon usage can therefore be explained by a balance between the forces of mutational bias and translational selection.

In a multicellular organism that has undergone considerable differentiation, such as a higher plant, changes in tRNA profiles have been reported to occur in a tissue-specific manner (Bennetzen and Hall, 1982). This introduces the possibility of codon usage being used as on means of regulating tissue-specific gene expression. If the distribution of isoaccepting tRNAs within the cells of a tissue matches the codon usage bias of abundant cell-specific mRNAs, then the output of major proteins in that tissue may be maximised. Alternatively, codon usage bias may ensure that minor mRNAs may be maintained at a low level within that tissue. Campbell and Gowri (1990) have provided convincing evidence that tissue-specific expression of genes is related to codon usage in monocotyledonous plants. However, tissue-specific expression of the dicotyledonous genes analysed did not appear to be related to differences in codon usage (Campbell and Gowri, 1990).

2.4.2.2 Intron splicing

The open reading frames of most plant genes are split into coding exons and noncoding introns (Lefebvre, 1990). Exceptions include the genes encoding zein storage proteins in maize (Heidecker and Messing, 1986), the chlorophyll *a/b* binding protein of wheat

(Kuhlemeier *et al.*, 1987) and the *Arabidopsis* disease resistance gene *RPS2* (Mindrinos *et al.*, 1994).

Introns may have resulted from genetic recombination, thereby bringing protein-domains originally encoded by separate genes. The presence of introns may ensure that coding sequences are kept intact during genetic crossing over events (Lefebvre, 1990). Alternatively, over time introns may also have increased the likelihood of DNA crossing-over events, thereby enabling evolution of multi-domain proteins. In certain plant genes, such as the gene encoding leghaemoglobin (Bogusz *et al.*, 1988), introns are inserted between DNA sequences that encode separate domains of the protein. The soybean leghaemoglobin gene has one more intron than the animal haemoglobins (Jensen *et al.*, 1981), although the first and third introns appear to be precisely homologous in position to all other known functional haemoglobin genes. This suggests that homologous intron/exon splice junctions have been maintained in the globin genes throughout evolution. However, conservation of intron splice sites in genes is not always evident. For example, there is no obvious relationship between the distribution of introns within plant actin genes and the functional domains of the corresponding gene products (Grierson and Covey, 1988). A comparatively recent discovery is that introns may affect plant gene expression. The first intron of the maize gene encoding alcohol dehydrogenase is required for high levels of transcription of the gene (Leuhrsen and Walbot, 1991).

Precise excision of intron sequences from the primary gene transcript (pre-mRNA) and concomitant ligation of the exons is accomplished through the nuclear multistep process of pre-mRNA splicing (Padgett *et al.*, 1986; Maniatis and Reed, 1987; Sharp, 1987). Many introns in genes from lower eukaryotes have the capacity for self-splicing in the absence of proteins or an energy supply (Zaug *et al.*, 1983). However, introns of most plant genes required specific small nuclear ribonucleotide particles (snRNPs) and proteins which must bind to intron splice sites in order to form a competent spliceosome, the particle that catalyses the splicing process (Green, 1986; Padgett *et al.*, 1986; Maniatis and Reed, 1987; Sharp, 1987).

Computer analysis of plant introns (Brown, 1986) has led to the identification of consensus sequences for splice sites at junctions of introns and exons and also the identification of a putative branch point possibly involved in the splicing process. Many of these are very

similar to those identified in animal systems. This suggests common mechanisms for recognition and excision of introns in eukaryotes. In particular, a feature of all introns in eukaryotic pre-mRNAs appears to be the presence of a GU dinucleotide at the 5' splice site and an AG dinucleotide at the 3' splice site (Csank *et al.*, 1990). However, these two dinucleotides occur elsewhere within exons and introns without being mistaken for splice sites (Goodall and Filipowicz, 1989). This suggests that rather than defining the intron borders within a unique sequence, accuracy of splicing is dependent on the presence and appropriate positioning of various sequence elements to which various snRNP particles and proteins must bind to form a competent spliceosome. The nature and relative importance of these individual sequence elements varies between different organisms (Goodall and Filipowicz, 1989).

Introns in plant nuclear genes are removed by the same lariat-type mechanism as in animals (Padgett *et al.*, 1986; Maniatis and Reed, 1987; Sharp, 1987). However, the sequence requirements for splice site recognition in animal and plant genes are subtly different. In general, plant cells do not process the introns of transcripts expressed from introduced vertebrate genes (Barta *et al.*, 1986; Van Santen and Spritz, 1987; Wiebauer *et al.*, 1988). The requirements for intron recognition in plants also differ from those in both metazoa and yeast (Goodall and Filipowicz, 1989). In contrast, plant introns appear to be faithfully processed in animal systems (Brown *et al.*, 1986; Hartmuth and Barta, 1986; Van Santen and Spritz, 1987; Wiebauer *et al.*, 1988).

Sequence elements involved in intron recognition in vertebrate genes consist of 5' and 3' splice sites, a branch region and a polypyrimidine tract adjacent to the 3' splice site (Goodall and Filipowicz, 1989). Plant genes are characterised by a general absence of an analogous polypyrimidine tract at their 3' ends (Wiebauer *et al.*, 1988). A 3' polypyrimidine tract is not required for splicing in plants (Goodall and Filipowicz, 1989). In the yeast *S. cerevisiae*, sequences near the 5' splice site and the branch point are highly conserved (Green, 1986; Padgett *et al.*, 1986). In yeast, an absolutely conserved UACUAAC branch point sequence (Parker *et al.*, 1987) appears to be the major factor providing specificity for 3' end recognition, with no equivalent to the polypyrimidine tract found in vertebrate introns.

In contrast to these systems, the element characteristic of plant introns is their high A+U content. Goodall and Filipowicz (1989) demonstrated that high A+U content in addition to

appropriate splice-site sequences is sufficient for the processing of pre-mRNA introns in plant cells. Within the intron, it is not the specific sequence or its location, but rather the ratio of A+U content to G+C content that is the most important determinant of splicing efficiency (Goodall and Filipowicz, 1989). These workers observed that in the vicinity of the splice sites of 171 introns from dicotyledonous plants, an average A+U content of 74% contrasted strongly with an average A+U content of only 55% in the flanking exon sequences (Goodall and Filipowicz, 1989). They speculated that AU-richness may be necessary for splicing either because such stretches have minimal secondary structure-forming potential or AU-rich sequences act as binding sites for one or more splicing factors (Goodall and Filipowicz, 1989).

The process of intron splicing represents a possible mode of regulation of eukaryotic gene expression at the posttranscriptional level. In the majority of instances studied to date, each and every one of the exons present in a gene is incorporated into the mature mRNA through the invariant ligation of consecutive pairs of donor and acceptor splice sites, removing every intron. This conventional form of splicing, often referred to as constitutive splicing, yields a single gene product from each transcriptional unit. However, not all genes are spliced constitutively. It is now well established that in many genes, splice sites of nonconsecutive exons are joined in the processing of certain primary transcripts (Andreadis *et al.*, 1987; Breitbart *et al.*, 1987). Such so-called alternative pre-mRNA splicing may exclude either certain individual exon sequences or clusters of exon sequences from the mature mRNA in some transcripts but include them in others (Andreadis *et al.*, 1987; Breitbart *et al.*, 1987). The use of such differential splicing patterns of pre-mRNA transcripts results in the generation of an array of related but structurally distinct isoforms from single genes. Alternative splicing of introns in pre-mRNAs therefore represents a mechanism of generating different products of slightly different enzymatic activities from a single gene.

2.4.3 The 3' end and polyadenylation

In eukaryotes, primary mRNA transcripts (pre-mRNAs) of nuclear genes are longer than the mature and translatable mRNAs present in the cytoplasm. The synthesis of pre-mRNA is followed, therefore, by extensive processing which includes 5' capping with

7-methylguanosine residues and intron splicing. In addition, the precursor mRNA is processed by endonucleolytic cleavage and subsequent addition of a polyadenylate tract to the 3'-OH generated by this cleavage. This tract of polyadenylate is required for subsequent posttranscriptional events and enables the expression of the mature mRNA in the cytoplasm (Hunt, 1994). Polyadenylation is therefore an essential step in the course of the expression of genes in eukaryotes.

The formation of 3' ends of mRNAs appears to be different between plant and mammalian genes. Whereas the conserved hexanucleotide AAUAAA found 10 - 30 bp before the 3' end is the polyadenylation signal in animal systems (Birnsteil *et al.*, 1985; Platt, 1986), the requirement for such a site in plant mRNAs is less stringent (Heidecker and Messing, 1986; Mogen *et al.*, 1990; Kuhlemeier, 1992). An exact match for the canonical AAUAAA motif can be found near polyadenylation sites in only approximately one third of plant genes (Hunt, 1994). An additional 50% of plant gene 3' regions currently characterised contain 4-5 bp out of 6 matches for this sequence. A significant proportion (15%) of plant polyadenylation sites have no AAUAAA-like motif (Hunt, 1994). However, as in mammalian genes, polyadenylation signals in plant pre-mRNAs usually precede the end of the mature message by 18 to 36 nucleotides (Joshi, 1987). Plants have been found to recognise animal gene polyadenylation signals with reduced efficiency and at somewhat different sites (Hunt *et al.*, 1987). This suggests that the spatial relationship between the proposed specificity factor that recognises AAUAAA and the endonucleolytic cleavage site is different in plants from the analogous relationship in animals. Furthermore, unlike mammalian transcripts, mRNAs arising from single plant transcription units are often heterogenous at their 3' ends because of the occurrence of multiple sites of polyadenylation (Dean *et al.*, 1986).

Plant poly(A) mRNA usually contains more than one AAUAAA-like polyadenylation signal, although normally only one is functional (Lefebvre, 1990). The role of multiple polyadenylation signals and what, if anything, controls the choice of which one is used is currently not clear. However, in most cases where two polyadenylation signals are found, the second one is recognised (Heidecker and Messing, 1986). The polyadenylation signals of plant genes may be contained in regions capable of forming hairpin loops (Heidecker and Messing, 1986). In cases where the first signal is not bypassed, the sequence surrounding the first signal affects potential hairpin structure in this region. Analysis of members of the

maize zein gene family indicates that genes in which hairpin structure around the first polyadenylation signal is stabilised appear to use this first signal (Heidecker and Messing, 1986).

In mammalian genes, the AAUAAA signal is generally the only upstream element needed for efficient polyadenylation (Birnsteil *et al.*, 1985; Platt, 1986). Polyadenylation in plants is likely to be directed by other signals in addition to near-upstream regions analogous to the mammalian AAUAAA-like motif. This is supported by the observation that these polyadenylation signals (now more correctly referred to as near-upstream elements; NUEs) have frequently been found to occur unrecognised in the coding regions and introns of several plant genes (Dean *et al.*, 1986). Current evidence suggests that plant poly(A) signals are composed of several *cis* elements that operate in concert in order to ensure cleavage and polyadenylation of pre-mRNAs (Hunt, 1994). Unlike animal genes, no sequences downstream of the polyadenylation site appear to be necessary for determination of the 3' ends of plant mRNA transcripts (Kuhlemeier, 1992). However, elements upstream of AAUAAA-like motifs have been found in several plant genes (Mogen *et al.*, 1990; Sanfacon *et al.*, 1991).

The most distinctive feature of plant polyadenylation signals in comparison with those used by animals and yeast is the requirement of sequences upstream from possible AAUAAA motifs (Hunt, 1994). Besides these polyadenylation signals (NUEs) usually located within 40 nucleotides of the polyadenylation site, polyadenylation in plant genes is also dependent on far-upstream elements (FUEs), which are relatively distant from their associated sites, and the actual sites of polyadenylation themselves (Hunt, 1994).

Far upstream elements can control more than one site within a gene, although not necessarily equally (MacDonald *et al.*, 1991; Mogen *et al.*, 1992). They are not responsible for the actual 3' end profile of the gene. Furthermore, FUEs from different genes appear to be interchangeable (Mogen *et al.*, 1992). At present, no strict consensus sequence that correlates with FUE function has been identified (Hunt, 1994). Furthermore, there are no obvious potential secondary structures involving these elements, or any other part of the region in which they occur, that might explain their role in mRNA 3'-end formation (Mogen *et al.*, 1990). However, FUEs characterised to date appear to possess a decided UG-richness (Joshi, 1987; An *et al.*, 1989; Ingelbrecht *et al.*, 1989). It has been suggested by Hunt

(1994) that this UG-richness alone might suffice for FUE function, much as AU-richness contributes to intron definition in plants (Section 2.4.2.2). Currently, it appears that many different subelements contribute to FUE function, and that the particular combinations of subelements determine the properties of any FUE. These domains possessing FUE activity are usually found between 40 and 160 nucleotides upstream from the sites they control (Hunt, 1994).

As with FUEs, there is no strictly conserved polyadenylation site consensus in plant genes. However, the average base composition near these cleavage sites is distinctive (Joshi, 1987). The defining feature of most polyadenylation sites in plant genes is probably the occurrence of a YA (Y=pyrimidine) dinucleotide in a U-rich region. The site of polyadenylation appears to be an independent *cis* element in plant polyadenylation signals. Mogen *et al.* (1992) have demonstrated that a single cleavage site in the pea *rbcS* gene can function with two different NUEs, but only if the polyadenylation site is brought to an appropriate position by deletion. Based on the known locations of NUEs and polyadenylation sites in several genes, it appears that NUEs and polyadenylation sites usually must occur within 20 and 30 nucleotides of each other in order to permit efficient polyadenylation (Hunt, 1994).

2.4.4 Stress-inducible genes as tools in studies of gene structure and regulation

Despite the vast increase in our appreciation of the structural basis of gene regulation, there are still many unanswered questions. Perhaps the most obvious is how the finely-tuned expression of the full complement of genes is regulated during development and in response to environmental factors.

Any terminally differentiated cell expresses only the array of genes required for its stable functioning. This implies that at any one time, only a fraction of the genome will be expressed. The mechanisms whereby this expression is regulated are largely unknown (Kuhlemeier, 1992). Exposure of plants to severe changes in the environment elicits a rapid and specific response by the genome involving a selective increase in expression of certain genes and a decline in levels of expression of others (Section 2.1.1). This enables

mobilisation of the necessary defenses in response to the environmental challenge. At the molecular level, this response is often characterised not only by its extreme rapidity, but also by its immensity. Such features make stress responsive genes ideal candidates as paradigms for elucidating the mechanisms of finely-tuned gene regulation.

It is likely that most stress-inducible genes have associated DNA sequences capable of perceiving stress signals. Many such *cis*-acting elements have already been identified. For example, the 5' region of a cold-regulated *Arabidopsis* gene has *cis*-acting elements that confer cold-; drought- and ABA-regulated gene expression (Baker *et al.*, 1994). Yamaguchi-Shinozaki and Shinozaki (1994) have identified a different *cis*-acting element in a dehydration-responsive gene of *Arabidopsis* which is involved in responsiveness to drought, low temperature and high salinity. The promoter of the *Arabidopsis* gene encoding alcohol dehydrogenase is inducible by anaerobiosis, dehydration and low temperature stress (Dolferus *et al.*, 1994). Analysis of the promoter region of this gene using 5' deletion and substitution mutants has enabled identification of promoter elements which interact differentially under these different stress conditions. One element is necessary for cold and dehydration induction and is homologous to the abscisic acid responsive element identified in wheat (Guiltinan *et al.*, 1990). However, a general stress responsive element within the promoter of the *Arabidopsis* gene encoding alcohol dehydrogenase has been identified which is responsive to all three stresses (Dolferus *et al.*, 1994).

Molecular analysis of stress-responsive genes is likely to be particularly important in contributing to our overall understanding of gene structure and the molecular events needed to trigger changes in gene expression. In particular, the *cis*- and *trans*- acting factors involved in ABA-induced gene expression have been analysed extensively. A conserved sequence YACGTGGC has been reported to function as an ABA-responsive element in many ABA-responsive genes (Marcotte *et al.*, 1989; Mundy *et al.*, 1990; Bray, 1991). Certain cDNAs encoding DNA-binding proteins that specifically bind to the ABA-responsive element have been cloned and shown to encode products containing the basic domain/leucine zipper (bZIP) structure (Guiltinan *et al.*, 1990; Oeda *et al.*, 1991).

Promoter sequences required for induction of genes by anaerobiosis have also been identified. The anaerobic regulatory element contains two subregions, each with a consensus sequence present in all anaerobically regulated genes examined to date (Ricard *et al.*, 1994).

This suggests a common mechanism of gene activation by anaerobiosis. The anaerobic regulatory element shares properties with other enhancer elements, functioning in both orientations, with subregions interacting cooperatively while multiple elements act additively (Ricard *et al.*, 1994). Activation is believed to be mediated via sequence-specific *trans*-acting DNA-binding proteins.

The induction of proline accumulation by a range of environmental stresses (Section 2.1.3) makes it particularly challenging to elucidate the molecular basis whereby the genes involved in the response are regulated. The mechanism(s) of signal transduction involved in translating an environmental parameter into a physiological response at the genetic level are likely to be extremely complex (Bray, 1994). Elucidation of how signals arising from the disparate range of stresses capable of eliciting proline accumulation cross-react and network in order to induce highly specific and regulated patterns of gene expression remains a challenge for the future. Preliminary characterisation of the structure of the genes encoding proline biosynthetic enzymes is an essential prerequisite for such investigations.

2.5 Fundamentals of protein structure, function and evolution

Although the gene is the primary unit of hereditary, its ultimate expression is dependent on the processes of transcription and translation to produce its corresponding protein product. With the exception of possible minor post-translational modifications, all the information required for specifying the primary structure of the protein product is contained within the nucleotide sequence of the gene. Therefore, having established the nucleotide sequence of a gene, the amino acid sequence of its corresponding product may be deduced. Although protein sequences may be determined by the sequential degradation of the polypeptide chain, or subfragments thereof (Hunkapiller *et al.*, 1974; Walsh *et al.*, 1981), an easier and more accurate method of determining protein sequences is the sequencing of the corresponding gene or cDNA.

Proteins may be classified into three main categories : globular, membrane and fibrous proteins (Fasman, 1990). Globular proteins pack tightly to form distinct tertiary conformations, by association of various secondary structures (usually α-helices and β-strands with intervening β-turns) into various domains (Fasman, 1990). The sequence of amino acids in a polypeptide chain is known as the primary structure.

Proteins are usually biologically active only when folded in their native conformations. Therefore, appreciation of their three-dimensional structures is the key to understanding how they function at the molecular level. Furthermore, in looking at the relatedness of proteins, two proteins may be topologically very similar despite apparently poor conservation of primary structure (Darby and Creighton, 1993). Since structure tends to change more conservatively than sequence, distant evolutionary relationships between proteins may often only by discernable at the level of protein conformation (Doolittle, 1989).

The covalent structure of a polypeptide chain is not sufficient to determine its three-dimensional structure owing to the possibility of different rotations about the many covalent bonds (Darby and Creighton, 1993). Since peptide units are effectively rigid groups that are linked into a chain by covalent bonds at the C_α atoms, the only degrees of freedom they

have are rotations about these bonds (Branden and Tooze, 1991). Each unit can rotate around either the C_α-C' or N-C_α bonds. By convention, the angle of rotation around the N-C_α bond is referred to as phi (Φ) and the angle around the C_α-C' bond from the same C_α atom is referred to as psi (ψ) (Branden and Tooze, 1991). In this way, each amino acid residue is associated with two conformational angles Φ and ψ (Branden and Tooze, 1991). Since these are the only degrees of freedom, the conformation of the whole main chain of the polypeptide is completely determined when the Φ and ψ angles for each amino acid are defined (Branden and Tooze, 1991).

For a single amino acid residue, most combinations of the Φ and ψ torsion angles are not permitted because of steric collisions between the main chain and the side chain (Branden and Tooze, 1991). The permitted values of the Φ and ψ torsion angles depend upon the size and shape of the side chain (Darby and Creighton, 1993). Furthermore, in a polypeptide chain, unfavourable steric interactions with atoms of neighbouring residues may also limit the combinations of Φ and ψ that are permissible.

The combinations of Φ and ψ torsion angles in a polypeptide chain are usually described by Ramachandran plots (Ramachandran and Sasisekharan, 1968). These analyse the configurational energy in terms of the two dihedral angles at the α-carbon atom of the peptide bond.

Analysis of Ramachandran plots constructed from a number of accurately determined protein structures reveals that the range of permitted bond angles Φ and ψ is greatest for glycine residues because of the absence of a bulky side chain in glycine (Branden and Tooze, 1991). This enables glycine residues to provide flexibility in the polypeptide chain and to make tight turns (Darby and Creighton, 1993). Glycine thus plays a structurally very important role; it permits unusual main chain conformations in proteins. Other residues have much less conformational flexibility. Residues with bulky side chains, such as valine and isoleucine, can only adopt a more limited range of Φ and ψ angles (Darby and Creighton, 1993). Proline residues, with their cyclic side chains, are allowed only limited values of Φ, depending upon the extent to which the ring system can be distorted (Darby and Creighton, 1993).

2.5.1 The three-dimensional structure of proteins

In three dimensional terms, proteins may be defined in terms of a structural hierarchy (Blake, 1985). The lowest level of this scale of increasing complexity is represented by the regular and local elements of secondary structure. These include α-helices, β-sheets and random coil structures. Preferred packing patterns of secondary structure (supersecondary structures) constitute the next level of complexity. The third level is represented by domains. Domains have the properties of a folded protein molecule, but are linked covalently through the polypeptide chain to one or more other domains (Blake, 1985). The fourth level, the tertiary structure, is the complete folding of the polypeptide chain to form the protein molecule. In multimeric proteins, quaternary structure describes the three dimensional arrangements of the different subunits (Blake, 1985). Whereas the upper and lower levels of this hierarchy may usually be rigorously defined, supersecondary structures and domains are often more elusive to absolute definition (Blake, 1985).

2.5.1.1 Secondary structure

In globular proteins, the polypeptide chain tends to run back and forth across a protein or individual domain from one surface to another. Each run usually, although not always, contains a helix or is part of a β-pleated sheet. These constituent parts are usually separated from each other by turns or random coil structures (Richardson and Richardson, 1989).

In general, α-helices and β-sheets occur in short segments. Their lengths are primarily limited by the diameter of the domain in which they occur (Richardson and Richardson, 1989). Furthermore, it is often difficult to establish exactly where regions of α-helix or β-sheet start and stop (Richardson and Richardson, 1989). Helices may bend in the middle, to the extent of losing a hydrogen bond, and may have frayed ends. β-sheets are often interrupted by β-bulges, which result when one or two residues break the regular pattern of sheet hydrogen-bonding (Richardson and Richardson, 1989).

Taking all proteins of known structure together, 89% of residues are involved in well-

ordered regions of secondary structure. About 31% of them occur in α-helices, 28% in β-sheets and 30% in coils and turns (Darby and Creighton, 1993).

The α-helix

The α-helix is the most common regular conformation in proteins (Richardson and Richardson, 1989). It is characterised by hydrogen bonding between the backbone carbonyl oxygen atom of each residue and the backbone NH of the fourth residue along the chain (Darby and Creighton, 1993). In an ideal α-helix, this leaves the first three NH groups and the last three carbonyl oxygens of the helix without hydrogen bonds. The average α-helix is 17Å long and contains eleven residues, which corresponds to three turns (Richardson and Richardson, 1989). However, α-helices in globular proteins vary considerably in length and may range from four or five amino acids to over forty residues (Branden and Tooze, 1991). The pitch, or repeat, of an ideal α-helix is 3.6 residues per turn (Richardson and Richardson, 1989). For that pitch, the rise per residue along the helix axis is 1.5Å, or 5.4Å per turn. Although real helices match this value quite well, a difference in the average pitch of as little as 5% (between e.g. 3.5 and 3.7 residues per turn), may produce an offset of an entire residue by the end of a typical four- or five-turn helix (Richardson and Richardson, 1989).

With the exception of proline, which causes the helix to kink, all amino acids are structurally compatible with the formation of α-helices. However, they differ in their tendencies to do so (Darby and Creighton, 1993). Although proline is generally regarded as a helix destabiliser, it is frequently found within the first turn of an α-helix (Richardson and Richardson, 1989). In this position, its lack of backbone hydrogen bonding can be accommodated and its side chain geometry is quite favourable (Richardson and Richardson, 1989).

The stability of any given α-helix depends on what residues are present. For example, glycine residues diminish the stability of the helix because of their flexibility in the unfolded state, while alanine has a particularly high helical propensity (Richardson and Richardson, 1989). Residues with longer side chains are more restricted in their side-chain conformations in the helix and are less likely to be helical (Darby and Creighton, 1993). Negatively and positively charged residues at the N- and C-termini of the α-helix,

respectively, enhance stability of the helix (Darby and Creighton, 1993). The opposite locations destabilise the helix (Darby and Creighton, 1993). Interactions between side-chains may also affect the stability of the α-helix (Darby and Creighton, 1993).

Amphipathic helices have non-polar side chains on one side and polar side chains on the other (Eisenberg *et al.*, 1982). The non-polar surfaces of such amphipathic helices tend to interact with each other or with other non-polar surfaces such as those of membranes. The amphipathic nature of such helices is usually expressed as the hydrophobic moment (Eisenberg *et al.*, 1982). This describes the spatial separation of hydrophobic and polar groups in the α-helix.

The β-sheet

After the α-helix, the second major element of regular secondary structure in proteins is the β-strand, which has the polypeptide backbone extended. Individual β-strands usually consist of between five to ten residues (Branden and Tooze, 1991). The average length of β-strands is 20Å, which corresponds to 6.5 residues (Darby and Creighton, 1993). The extended structure of the β-strand is not stable in isolation and is observed only in β-sheets. These are aggregates of multiple β-strands (Darby and Creighton, 1993). Therefore, whereas α-helices involve repeating patterns of local hydrogen bonding, β-structure involves repeating patterns of hydrogen bonds between distant parts of the polypeptide backbone (Richardson and Richardson, 1989).

The side chains on β-strands extend approximately perpendicular to the plane of hydrogen bonding. Along the strand, they alternate from one side to the other, but on adjacent strands participating in β-sheet formation, they are in register (Richardson and Richardson, 1989). β-strands are cross-linked to other β-strands running either in the same (parallel) or opposite (antiparallel) directions with respect to one another. The linkage is via hydrogen bonds between the carbonyl and NH groups that protrude from the peptide chain at right angles to the side chains (Darby and Creighton, 1993). The β-sheets that are formed from β-strands are "pleated", with C_α atoms occurring successively above and below the plane of the β-sheet (Branden and Tooze, 1991). Both parallel and antiparallel β-sheets have distinctive patterns of hydrogen bonding (Branden and Tooze, 1991).

For antiparallel β-sheets, typically one side is buried within the interior and the other side is exposed to the solvent. As a consequence, the amino acid types tend to alternate between hydrophobic and hydrophilic (Richardson and Richardson, 1989). In contrast, parallel sheets are usually buried on both sides. Their central sequences are usually highly hydrophobic and hydrophilic residues concentrate at the ends (Richardson and Richardson, 1989).

The factors that stabilise the β-strand and β-sheet conformations are not certain, for there is currently no adequate model for studying their formation (Darby and Creighton, 1993). However, as with a proline in an α-helix, there may be disruptions of hydrogen bonding in β-structure. By far the most common is the β-bulge (Richardson *et al.*, 1978). This results from the insertion of an extra residue into one strand so that between a pair of hydrogen bonds there is now one residue on the normal strand but two residues on the bulged strand (Richardson *et al.*, 1978). Proline residues are relatively rare in β-sheets as they cannot participate in the hydrogen bonding network (Darby and Creighton, 1993). Nevertheless, proline is common as a "breaker" residue at the end of a β-strand and in β-bulges (Richardson and Richardson, 1989).

The random coil

Random coils, otherwise referred to as loop regions, are sections of the polypeptide chain that connect regions of regular secondary structure (Branden and Tooze, 1991). Unlike helices and sheets, these are non-repetitive regions and are characterised by possessing no regular secondary structure. Polypeptides assume such structures in the absence of net interactions between atoms distant in the covalent structure and with the solvent. Despite their designation as "random" structures, coils are nevertheless well-ordered, compact and stable pieces of secondary structure (Richardson and Richardson, 1989). Surveys of the known three-dimensional structures of loops have shown that they fall into a rather limited set of structures and are not a random collection of possible structures (Branden and Tooze, 1991).

α-helices and β-sheets usually run across a protein or domain from one surface to another and form the stable hydrophobic core of the molecule. In contrast, coiled structures characteristically appear on the surfaces of proteins (Richardson and Richardson, 1989).

Here the main chain carbonyl and NH groups of loop regions, which in general do not form hydrogen bonds with each other, can hydrogen bond to water molecules. Accordingly, loop regions exposed to solvent are frequently rich in charged and polar hydrophilic residues (Branden and Tooze, 1991). Frequently, coil regions reverse the direction of the polypeptide chain (Richardson and Richardson, 1989). For a short region of polypeptide chain, only three to four residues are needed to reverse direction or enable the chain to fold back on itself (Richardson and Richardson, 1989). The conformations of short loops depend primarily on the position within the loop of special residues (usually glycine, asparagine or proline) that enable the chain to take up an unusual conformation.

Random coil regions often connect successive α-helices or strands of β-sheet that interact with each other to form supersecondary structures. Such structures include β-hairpins (two successive strands of anti-parallel β-sheet connected by a turn) and pairs of anti-parallel packed helices (Branden and Tooze, 1991). A structural motif common to many nucleotide-binding domains is a strand of sheet followed by an anti-parallel helix followed by a parallel strand of β-sheet (Wierenga *et al.*, 1986). This conformation requires two regions of coiled structure.

The β-turn is a common structural feature of most globular proteins (Richardson and Richardson, 1989). In most proteins, segments of α-helices or β-sheets fold back on themselves to yield a series of secondary structure elements. These are usually punctuated by turns (Doolittle, 1989). Chou and Fasman (1978a) found that turns tend to be composed of a definite subset of amino acids. Prominent among these is the frequent occurrence of a proline residue at the second position of a four-residue turn sequence (Chou and Fasman, 1978a).

Turns have a strong tendency to occur at local maxima of hydrophilicity (Rose and Seltzer, 1977). Accordingly, although β-turns may on occasion be buried within the interior of the molecule (Rose *et al.*, 1983), they are usually situated at the surface of the protein where they are exposed to the aqueous environment (Fasman, 1989).

In many enzymes, random coil structures contain residues which are of importance in binding substrates and forming active sites (Branden and Tooze, 1991). Other coil structures have a purely structural role and are frequently subject to amino acid

substitutions. Whereas amino acid substitutions within the hydrophobic core may disrupt the critical packing interactions required for protein stability, substitutions on the protein surface are usually innocuous (Darby and Creighton, 1993).

2.5.1.2 Supersecondary structure

At the atomic level, protein structures are extremely complex and unique. However, if attention is focused on the interactions of the secondary structures found within proteins, it becomes apparent that many protein structures possess topological similarities (Darby and Creighton, 1993). Owing to the limited number of ways in which α-helices, β-sheets and turns can be linked together, there appear to be a finite number of folding patterns found within proteins (Darby and Creighton, 1993).

The term supersecondary structure was introduced by Rao and Rossman (1973) to describe compact elements of secondary structure. In general, supersecondary structures exemplify a number of basic rules of protein topology. In particular, elements of secondary structure that are sequential in the primary structure are frequently adjacent in the tertiary structure and tend to pack in an antiparallel manner (Darby and Creighton, 1993). Furthermore, connections between elements of secondary structure do not cross each other or make knots in the polypeptide chain (Darby and Creighton, 1993).

2.5.1.3 Domains

A domain may be simply described as a part of a molecule whose structure is organised like a complete globular protein (Blake, 1985). It differs from a single molecule in that it is connected by a polypeptide chain to one or more other domains of similar characteristics. The majority of proteins with polypeptide chains longer than 200 residues contain domains, and the longer the chain the larger the number of domains (Blake, 1985). In domain enzymes, individual substrates or effectors are often bound to individual domains, which are usually structurally independent and capable of folding independently (Darby and Creighton, 1993). However, the active site is usually located at the interface between domains (Blake, 1985).

Domains are formed from compact, contiguous chain structures (Wetlaufer, 1973). Generally, domains include only residues that are consecutive in the peptide chain (Darby and Creighton, 1993). However, a number of workers (Wetlaufer, 1973; Richardson, 1981; Rose, 1978) have defined domains in a number of significantly different ways. In particular, these workers have disagreed on how small a domain may be.

It is now clear that proteins fall into distinct structural families, built from an apparently limited set of motifs. Peptide geometry and the limited conformational states of L-amino acids dictate that the α-helix, β-strand and composite secondary structure patterns (supersecondary structures) dominate protein structures. The same structural motifs recur many times in "unrelated" proteins. These recurrent motifs have strengthened the concept of a hierarchy of protein structure comprising building blocks of gradually increasing complexity. Domains have often been experimentally observed to be folding intermediates, with each domain showing its own cooperative transition (Kim and Baldwin, 1982). The folding of domains provides indisputable evidence of hierarchical folding of proteins at the tertiary structure level (Richardson and Richardson, 1989). Besides being proposed as intermediates in process of folding (Kim and Baldwin, 1982), domains have also been proposed as functional units within the native structure (Richardson, 1981; Rossman and Argos, 1981; Kim and Baldwin, 1982) and possibly even as fundamental genetic units (Gilbert, 1978; Blake, 1979).

Domains are usually associated with individual functions (Blake, 1985). Rossman *et al.* (1974) demonstrated that a number of NAD-requiring dehydrogenases which use different substrates are all characterised by a particular βαβ-unit capable of binding the ADP-moiety of the dinucleotide in an identical manner. Although different proteins with this fold bind NAD in very similar ways, they frequently possess very different amino acid sequences within the binding site. Nevertheless, there are some similarities in sequence, for within each βαβ unit, there is a highly conserved sequence, Gly-X-Gly-X-X-Gly (X represents any residue). This makes it possible to identify nucleotide-binding domains from amino acid sequence alone (Wierenga *et al.*, 1986; Scrutton *et al.*, 1990) .

Wierenga *et al.* (1986) derived an amino acid sequence "fingerprint" that may be used to test whether a particular sequence will fold into a βαβ unit with ADP-binding properties. This fingerprint, constructed from analysis of five proteins, is a set of eleven rules

describing the type of amino acid that should occur at a specific position within a peptide fragment forming an ADP-binding βαβ-fold. The total length of this fingerprint varies between 29 and 31 residues (Wierenga *et al.*, 1986).

The first two glycine residues in the Gly-X-Gly-X-X-Gly sequence motif occur in a loop between the β-strand and the following α-helix. The first glycine residue is essential for the tightness of the turn at the end of the first strand of the β-sheet and marks the beginning of the succeeding α-helix. The second glycine residue allows the dinucleotide to be bound without obstruction from an amino acid side chain at this position. The third glycine is located in the α-helix, where the absence of a side chain permits especially close packing of the βαβ unit (Wierenga *et al.*, 1986). There are also six positions that always have hydrophobic residues, because they form the hydrophobic core between the α-helix and the two β-strands. In NAD-binding domains, a conserved glutamic acid or aspartic residue interacts through a hydrogen bond with the 2'-hydroxyl of the adenine ribose of the nucleotide (Wierenga *et al.*, 1986) . The adenine moiety binds in a hydrophobic cleft, while the nicotinamide ring binds so that one face is in a polar environment and the other interacts with non-polar amino acid side-chains (Darby and Creighton, 1993).

Minor alterations to the NAD-binding site permit binding of NADP. The third glycine of the consensus sequence becomes alanine, so that the packing between the β-sheet and α-helix is no longer so close. This enables accommodation of the of the additional 2'-phosphate found in NADP. Furthermore, the negatively charged residue at the C-terminal end of the second strand is replaced, presumably to accommodate the 2'-phosphate group on the coenzyme (Scrutton *et al*; 1990). The negatively charged acidic residue characteristic of NAD-binding proteins is therefore the principle means whereby these enzymes discriminate between NAD and NADP as coenzyme (Branden and Tooze, 1991).

Consideration of the structure of the dinucleotide-binding fold demonstrates that there are strong stereochemical constraints at specific positions in the polypeptide chain that must be respected in order to preserve the structure and function of a domain. The amino acids at these key sites are often diagnostic.

2.5.1.4 Tertiary structure

The polypeptide chains of globular proteins pack tightly to form distinct tertiary conformations, usually producing a hydrophobic core (Fasman, 1989). Usually, approximately 75% of the interior space of globular proteins is filled by atoms (Darby and Creighton, 1993). The compactness of the protein results from the close packing of the elements of secondary structure. Often disulphide cross-links between cysteine residues further stabilise the structure (Richardson and Richardson, 1989). Once the tertiary structure has been assumed, the torsion angles of most peptide bonds are close to those that are allowed (Darby and Creighton, 1993). Therefore, relatively little conformational strain is incorporated into the three-dimensional structure unless it is required for functional reasons (Darby and Creighton, 1993).

Although short segments of α- and β-secondary structures may be mixed in globular proteins, they do so in a relatively well organised manner to form identifiable classes of tertiary structure (Levitt and Clothier, 1976). Since there is a definite recurrence of structural themes among apparently unrelated proteins, it is common to classify them on the basis of topology or "fold".

Most proteins can be classified as belonging to one of four basic classes according to their secondary structural arrangement (Levitt and Chothia, 1976). In the "all-α" class, only α-helices are present as secondary structures and the chain folds to pack the helices together in certain well-defined arrangements. In the "all-β" class, all of the structures are β-strands which hydrogen bond together to form twisted sheets packed face to face (Levitt and Chothia, 1976). In proteins belonging to the "α+β" class, both α- and β-structures are present, but are segregated in different domains within the protein (Levitt and Chothia, 1976).

The final class is the "α/β" class in which α-helices and β-strands alternate along the length of the polypeptide chain (Levitt and Chothia, 1976). Commonly, these proteins have a central β-sheet comprising parallel β-strands, flanked on either side by α-helices running antiparallel to the local β-strands. Alternatively, a central β-sheet of "α/β" proteins may form a barrel, with associated α-helices forming an outer layer to the barrel structure

(Branden and Tooze, 1991).

2.5.1.5 Quaternary structure

Quaternary structure describes the association of multiple subunits in multimeric proteins (Darby and Creighton, 1993). These subunits may be identical or different. Although the subunits are essentially independently folded structures, their entropic advantage of remaining independent may be overcome if they have surfaces that are complementary in shape and physical interactions. Such physical interactions might include non-polar interactions, hydrogen bonds or salt bridges (Darby and Creighton, 1993). Depending on the mode of association, there may be a fixed number of subunits, frequently two or four, or a variable number (Darby and Creighton, 1993).

Interactions between identical subunits are either isologous or heterologous (Darby and Creighton, 1993). Whereas isologous interactions involve the same surfaces on each monomer, heterologous interactions use different surfaces (Darby and Creighton, 1993). Isologous interactions usually produce structures with a fixed number of subunits. In contrast, heterologous interactions between monomers have no inherent limitation to the size of the aggregates that they produce. Frequently, they lead to indefinite polymerisation, as appears to be the case in P5CRs characterised to date (Table 2.3).

The tightness of binding of subunits in multimeric proteins is determined by the number and strengths of the contacts made between the individual subunits. The strength of the interaction between subunits can vary widely (Darby and Creighton, 1993). Whereas certain quaternary structures are extremely difficult to dissociate, in many proteins, the interactions are so weak that they are of dubious physiological significance. As outlined in Section 2.2.1.2, P5CR may be such a protein in which assumption of a specific quaternary structure is not critical for activity.

2.5.2 Prediction of protein structure *ab initio*

Most aspects of protein configuration derive, in the final analysis, from the properties of the particular sequence of amino acids that make up the polypeptide chain. These properties include the characteristics of both the individual side chains and the polypeptide backbone. This is evidenced by the remarkable self-assembly capabilities shown by many proteins for folding *in vitro* (Richardson and Richardson, 1989).

Therefore, a central tenet of protein chemistry is that the three-dimensional structure of a protein is determined by its amino acid sequence. The classical experiments of Anfinsen and co-workers (Anfinsen *et al.*, 1961; Anfinsen, 1973) proved that ribonuclease could be denatured and refolded without loss of enzymatic activity. This demonstrated that the amino acid sequence contains enough information to define the three dimensional structure of a protein in a particular environment. Despite the recent identification of sets of proteins called chaperones that are required for the formation of proper tertiary structure of many proteins (Branden and Tooze, 1991), these proteins merely act as catalysts to increase the rates of the final steps in the folding process (Branden and Tooze, 1991). Therefore, the influence of chaperones on the folding of certain proteins does not in principle violate the central dogma of protein folding.

The acceptance of the tenet that primary structure determines the overall conformation of a protein has led to several attempts by both theoreticians and experimentalists to predict the conformation of proteins based on the consideration of the amino acid sequence alone (Fasman, 1989). Since well-established principles of physics and chemistry must govern the structure and function of proteins, in principle, it should therefore be possible to predict protein structure from sequence. Nevertheless, devising schemes for predicting the folding polypeptide chains into their unique conformations based exclusively on their amino acid sequences remains one of the most persistent and challenging problems in molecular biology (Chou, 1989). Despite knowing the complete three dimensional structure of some three hundred different proteins, it is still not possible to formulate a set of general rules to enable prediction of the three dimensional structure of a protein from the amino acid sequence of its polypeptide chain (Branden and Tooze, 1991). At the present time, the three-dimensional structure of a protein must be determined experimentally, unless the structure of a closely related protein is known.

Several workers (Levitt and Warshell, 1975; McCammon *et al.*, 1977; Weiner *et al.*, 1984) have speculated that, based on the assumption that a protein folds in order to minimise the free energy, the native structure of a protein could be predicted by calculating the total free energy of a protein and finding the global minimum. However, strategies employing energy minimisation have failed to predict chain folding accurately (Hagler and Honig, 1978; Cohen and Sternberg, 1980). The main problem encountered with the simulation of protein folding using energy minimisation approaches is the existence of many conformations corresponding to a minimum in potential energy (Fasman, 1989). Therefore at present, it is still impossible to predict the three dimensional structure of a protein by quantum-mechanical approaches to minimise the free energy of an all-atom representation (Fasman, 1989).

2.5.2.1 Prediction of secondary structure

The local, regular elements of secondary structure have been the subject of most approaches to protein structure prediction. For many years, it has been speculated that the ability to accurately predict secondary structure elements may enable prediction of how they pack together to generate the folded conformation. Since the regular elements of secondary structure simplify the comprehension and description of complex tertiary structures, the ability to make a good prediction of secondary structure is of crucial importance whenever an attempt is made to predict the three-dimensional structure of a protein molecule (Garnier *et al.*, 1978).

To date, over twenty different proposals have been presented for the prediction of the secondary structure of protein conformation from the amino acid sequence (Fasman, 1989). However, the three most widely used methods for prediction of protein secondary structure are those of Chou and Fasman (1974a, 1974b, 1978a, 1978b), Garnier *et al.* (1978) and Lim (1974a, 1974b). The first two methods use an empirical statistical approach using parameters obtained from the analysis of known sequences and structures. In contrast, the predictive strategy of Lim (1974a, 1974b) is based on stereochemical criteria.

Since α-helices, β-sheets and reverse turns are determined not by individual, but by several adjacent residues, a segment of a particular secondary structure is much more probable when several sequential residues tend to prefer that structure. Accordingly, most predictive

methods examine short-range interactions between residues and are based on the premise that the local sequence of a polypeptide chain determines its local structure.

The Chou-Fasman algorithm for the prediction of protein secondary structure is one of the most widely used predictive schemes (Prevelige and Fasman, 1989). This is because of its relative simplicity and its reasonably high degree of accuracy. It is an empirical method based on the correlation between amino acid distribution and conformational settings in proteins of known structure. The frequency with which individual amino acids or short peptides within known structures occur in α-helices, β-sheets or reverse turns can be used to evaluate the probability that these secondary structures may occur in the peptide of interest.

Chou and Fasman (1974a) carefully examined the X-ray determined structures of fifteen proteins containing a total of 2 473 amino acid residues. Thee number of occurrences of a given amino acid in the α-helix, β-sheet and coil conformations used to define the conformational parameters P_α ,P_β and P_c respectively (Chou and Fasman, 1974a). These parameters are essentially a measure of a given amino acid's preference to be found in α-helix, β-sheet or coil. Furthermore, they presumably contain information about many physiochemical parameters defining protein stability, such as hydrophobicity, which are properly weighted for their relative importance (Prevelige and Fasman, 1989). Having computed these conformational parameters, Chou and Fasman formulated a set of empirical rules for predicting protein structure (Chou and Fasman, 1974b). The development of these empirical rules was guided by underlying considerations of protein structure (Prevelige and Fasman, 1989). Amino acids were classified as favouring, breaking, or being indifferent to each type of conformation. An abbreviated set of rules used in applying the Chou-Fasman predictive method is provided in Appendix 2. The method is simple in principle. Using the conformational parameter, one finds nucleation sites within the sequence and extends them until a stretch of amino acids is encountered that has a greater disposition for another type of structure. At this point, the structure is terminated. This process is repeated throughout the length of the polypeptide chain until the entire sequence is predicted.

Chou and Fasman later improved on their method by extending the analysis of α-helix, β-sheet and coil to include 29 proteins of known structure (Chou and Fasman, 1978a, 1978b) and determining the conformational parameter P_t from analysis of amino acid residues found in turns (Chou and Fasman, 1977). In the case of turns, a significant difference was

observed in the frequency of the first, second, third and fourth positions of β turns for all residues. Some residues, such as proline, were found to have a dramatic positional preference. Therefore, this positional preference was considered in the parameter for the prediction of turns (Chou and Fasman, 1979).

Another popular method used in secondary structure prediction is that presented by Garnier *et al.* (1978). This approach is based on information theory and considers the effects of residues from positions *i*-8 to *i*+8 on the conformation of the residue at position *i*. The prediction of the conformational state of each residue involves evaluation of an equation for each conformational state and then selection of the conformation with the highest information content (Garnier *et al.*, 1978).

Analysis of an entire polypeptide chain is facilitated by a computer program. From the information content of the equation for each conformational state, a value can be subtracted before comparison of the information contents. This value is referred to as the decision constant (Garnier *et al.*, 1978). The decision constant is an adjustable parameter chosen with the aim of producing optimal predictions by introducing a slight bias. However, choice of a decision constant prior to secondary structural predictions requires an independent measurement of the percentage composition of α-helical and β-sheet regions within the protein of interest. This may be established by circular dichroism or optical rotatory measurements (Garnier *et al.*, 1978). However, prior choice of a decision constant is usually only critical to improving the prediction of proteins that are exceptionally rich in either α-helical or β-sheet regions (Garnier *et al.*, 1978).

A fundamental difference between the predictive methods of Chou and Fasman (1878a, 1978b) and Garnier *et al.* (1978) is the extent of involvement of the user in making decisions concerning the assignment of secondary structures. The guidelines set out by Chou and Fasman (1978a, 1978b) are qualitative rather than quantitative, and therefore open to interpretation by the user. Since these rules lack the rigorous definition required by a computer algorithm, the data generated must be reduced by hand. In contrast, the computer-generated predictions provided using the algorithm of Garnier *et al.* (1978) assign residues to a single conformational state in a completely unambiguous manner (Garnier *et al.*, 1978). These workers have argued that any secondary structure predictive method should always be presented as an algorithm. Only in this way can ambiguities be resolved and objective predictions guaranteed (Garnier *et al.*, 1978). In defense of this challenge,

Prevelige and Fasman (1990) have argued that in reducing predictive data to a single conformational state, significant information may be lost. Frequently, regions within a polypeptide chain will display significant propensity for more than one type of structure. Such regions may be sites of conformational change and are often likely to be of considerable importance (Fasman, 1989).

The secondary structure assignments generated using the methods of Chou and Fasman (1978a, 1978b) and Garnier *et al.* (1978) cannot be used unquestioningly. Neither of the methods provides accuracy comparable with structural determinations using X-ray crystallography. At present, our ability to predict secondary structure is not better than distinguishing helix, sheet and turn residues with slightly over 60% accuracy (Fasman, 1989). A particularly difficult problem common to all current predictive strategies remains that of defining precisely where pieces of secondary structure begin and end (Richardson and Richardson, 1989).

The concept of a limit to the accuracy of predicting secondary structures has a clear theoretical basis. As pointed out by Garnier *et al.* (1978), in order to achieve a compact globular structure, the tertiary interactions between residues far apart in the amino acid sequence are often likely to override the intrinsic conformational tendencies of many amino acid residues within the polypeptide chain. It is likely that interactions between residues distant in the amino acid sequence, which are not considered in such predictions, have substantial roles to play in determining the secondary structure of any segment.

Nevertheless, in predicting the secondary structure of a protein, there is a distinct advantage to analysing the sequences of homologous proteins which are likely to have similar secondary and tertiary structures (Garnier *et al.*, 1978). Comparison of results obtained from analysis of homologous proteins is therefore a means of verifying the predicted secondary structure. In predicting the secondary structure of a protein, use of more than one predictive method is also advisable. Comparison of the results obtained with different methods is likely to provide some indication of the significance of the results obtained.

2.5.2.2 Class prediction

Despite the comparatively low success rate of predictive methods in accurately assigning

individual regions of sequence to particular conformations (Fasman, 1989; Section 2.5.2.1), contemporary predictive methods can do a fairly good job of distinguishing the basic classes of protein families (Richardson and Richardson, 1989). This enables classification of the protein of interest to one of these four basic classes of proteins (α, β, α+β and α/β) originally defined by Levitt and Chothia (1976). A simple means of assigning a protein of unknown structure to one of these classes is useful as conformational parameters derived from known proteins of that class may then be used to predict the unknown protein conformation (Chou, 1989).

Protein structural classes may also be predicted on the basis of amino acid composition (Chou, 1989). Analysis of the X-ray structures of 64 different proteins, in terms of their α-helical and β-sheet regions as well as their amino acid compositions (Chou, 1979, 1980 cited by Chou, 1989) revealed that the four distinct classes of proteins, *viz.* α, β, α+β and α/β (Levitt and Chothia, 1976; Section 2.5.1.4) had significantly different amino acid compositions. A computerised algorithm is capable of assigning proteins to the correct structural class based on amino acid composition with 80% accuracy (Chou, 1989). Other workers have noticed a similar relationship between the amino acid composition of a protein and the structural class to which it belongs. Assignment of the folding types of 135 proteins of known three dimensional structure in terms of their amino acid compositions, gave an accuracy of 79% (Nakashima *et al.*, 1986 cited by Fasman, 1990).

As outlined in Section 2.5.1.4, although α+β and α/β proteins contain almost identical amounts of helices and β-sheets, they differ in the topological packing of their α- and β-regions (Levitt and Chothia, 1976). In α+β proteins, the α-helical and β-sheet structures do not mix, but tend to segregate in different domains within the protein. In contrast, α/β proteins are comprised of approximately alternating α-helices and β-strands (Levitt and Chothia, 1976).

Chou (1989) has suggested two powerful criteria for distinguishing α+β and α/β proteins. Firstly, analysis of 16 α/β proteins revealed that they are characterised by their large size. Only four of the representative set selected by Chou (1989) contained fewer than 248 amino acids (Chou, 1989). The average size of the α/β proteins was 271 amino acids (Chou, 1989). In contrast, the average chain length of the α+β proteins studied was only 135 amino acid residues (Chou, 1989). The larger size of α/β proteins in comparison with α+β proteins is consistent with the fact that proteins belonging to α/β class are generally

divided into at least two domains (Chou, 1989; Branden and Tooze, 1991).

Furthermore, there is a significant difference in the amino acid compositions of α/β and α+β proteins. In particular, α/β proteins are characterised by a high abundance of hydrophobic residues and a low content of cysteine (Chou, 1989). In the set of proteins belonging to both classes that was studied by Chou (1989), in comparison with the α/β proteins, the α+β proteins had greater amounts of asparagine, tyrosine and cysteine and less valine, leucine, lysine and glutamate (Chou, 1989).

Therefore, although the helix and sheet contents of α+β and α/β proteins are usually almost identical, the two classes may be distinguished on the basis of their chain length and amino acid compositions.

2.5.2.3 Hydrophobicity

Prediction of the secondary structures of proteins provides little conclusive information concerning the actual location of individual amino acid residues within a protein molecule. However, analysis of the relative hydrophobicities of stretches of amino acids within a polypeptide chain provides a rough translation of the one-dimensional property of protein primary structure into the three-dimensional configuration assumed by that protein (Branden and Tooze, 1991).

Since the native structure of any globular protein must exist in the presence of water, the analysis of its interaction with water is central to the prediction of protein structure. A central dogma of protein chemistry is that three-dimensional structures are entirely dictated by primary structures (Doolittle, 1989). However, another fundamental tenet about protein folding is that it is driven by the entropy of removing hydrophobic groups from contact with the aqueous environment (Richardson and Richardson, 1989).

Patterns of hydrophobic *versus* hydrophilic side chains are very important for the prediction of protein structure simply by virtue of their preferential occurrence on the inside *versus* the outside of the folded polypeptide chain (Richardson and Richardson, 1989). In most globular proteins, alanine, glycine, isoleucine, leucine, phenylalanine and valine comprise approximately 63% of the residues found in the protein interior (Darby and Creighton,

1993). In contrast, charged residues, even those paired in salt bridges with net neutrality, are comparatively rare in the interiors of proteins. About 96% of charged groups within globular proteins occur at the surface (Darby and Creighton, 1993).

Water is statistically more ordered next to hydrophobic side chains than in bulk solvent (Richardson and Richardson, 1989). Water molecules cannot hydrogen-bond with hydrophobic side chains and are thereby restricted in their orientation. In contrast, hydrogen bonded ordering occurs next to hydrophilic side chains (Richardson and Richardson, 1989). Therefore, it is more favourable for hydrophobic groups to interact with each other, and waters to interact with waters, than for hydrophobic side chains to interact with waters (Richardson and Richardson, 1989). This is the dominant effect that drives soluble globular proteins towards compact, folded conformations in solution. However, it is important to note that not all proteins exist in the same solvent conditions. In particular, the overall folding constraints are different for membrane-bound proteins and for soluble proteins.

Although there is no doubt about the overall importance of hydrophobicity in predicting the structures of proteins, its detailed characterisation has been complicated by the multiplicity of scales for measuring it (Richardson and Richardson, 1989). Hydrophobicity is related to electrostatic charge, to hydrogen-bonding capability, and to surface area of aliphatic and aromatic carbons. However, there is no consensus as to how these various factors should be weighted (Richardson and Richardson, 1989). This has led to the development of several hydrophobicity scales, each of which assign different hydrophobicity values to each of the twenty amino acids. Hydrophobicity scales have been presented by, amongst others, Janin (1979), Wolfenden *et al.* (1981) and Rose *et al.* (1985).

However, by far the most widely used hydrophobicity scale is that presented by Kyte and Doolittle (1982). This scale is based on an amalgam of experimental observations concerning the hydrophilic and hydrophobic properties of the side chains of the twenty amino acids found in proteins. The two most important parameters considered in the hydrophobicity scale of Kyte and Doolittle (1982) are the transfer free energy of a residue between the water and vapour phases and the interior/exterior distribution of the residue in soluble proteins of known structure (Kyte and Doolittle, 1982).

Using this scale, a hydrophobicity profile may be computed by progressively averaging the hydrophobicity of a protein within a moving window that is stepped along the amino acid

sequence. This window should be larger than a single turn (that is., greater than four residues), but smaller than a complex segment such as a turn-helix (that is, less than twelve residues) (Rose and Dworkin, 1989). A midpoint line is plotted that corresponds to the grand average of the hydropathy of amino acid compositions found in most proteins (Kyte and Doolittle, 1982).

The resultant hydrophobicity profile (a graph of the average hydrophobicity per residue against position in the sequence) provides a simple way of quantifying the concentration of hydrophobic residues along the linear polypeptide chain. Local extrema in hydrophobicity may then be used predictively. In the case of soluble globular proteins, there is a remarkable correspondence between the interior portions of their sequences and the regions appearing on the hydrophobic side of the midpoint line, as well as the exterior portions and the regions on the hydrophilic side (Kyte and Doolittle, 1982). Given the sequence of a protein of unknown structure, hydrophobicity plots have also been used to predict chain turns and antigenic sites (Rose and Dworkin, 1989).

Hydrophobicity analysis is particularly useful for interpreting membrane-associated protein sequences (Finer-Moore *et al.*, 1989). Membrane-incorporated segments that are in contact with lipids must be in an ordered configuration. The reason for this lies in the necessity to saturate the hydrogen bonds intramolecularly, since the lipid hydrocarbon chains cannot participate in hydrogen bonding (Jähnig, 1989). However, they may be in either α-helix or in β-strand conformation (Jähnig, 1989). Helices represent the most stable structures in membranes as the hydrogen bonds can be formed along the helix as intrachain bonds (Jähnig, 1989). A membrane-spanning helix requires at least twenty hydrophobic residues with a mean hydrophobicity of at least 1.6 (Jähnig, 1989).

2.5.2.4 Conclusion

The prediction of the tertiary structure of a protein from its amino acid sequence is possibly the major unsolved problem in structural molecular biology. Despite the wide range of approaches suggested by several workers, our ability to predict secondary structure is not better than distinguishing helix, sheet and turn residues with slightly over 60% accuracy (Fasman, 1989). Kyte and Doolittle (1982) have also pointed out several limitations inherent in their method of predicting the surface probabilities of individual residues.

Analysis of known protein structures indicates that the extent to which residues are buried depends not only on hydrophobicity, but also on steric effects that determine packing within the crowded interior of the macromolecule (Kyte and Doolittle, 1982). Hydrophobic residues frequently occur on the surfaces of proteins, where their nonpolar side chains may be removed from contact with water by being buried between pieces of secondary structure (Kyte and Doolittle, 1982).

The primary limitation of all of the current methods available for the prediction of protein structure is that protein molecules are hierarchical or modular only to a limited extent. Local interactions in the sequence can only partially determine even the local levels of structure (Fasman, 1989). Although short-range interactions appear to play the dominant role in determining the amino acid conformation of a protein, the rest of the protein has an influence even on local secondary structure (Fasman, 1989). These additional interactions required to stabilise the conformation uniquely are not within the predictive powers of current strategies.

Richardson and Richardson (1989) conclude that both cooperativity and hierarchy are important principles in protein folding. They are coexistent and complementary. Whereas hierarchy makes folding possible, cooperativity is vital for the biological functioning of proteins as it confers versatility, responsiveness and functionality, as in the case of allosteric regulation of enzyme activity (Richardson and Richardson, 1989).

2.5.3 The evolution of protein structures

An organism's genome and the immediate products of its expression, namely proteins, probably represent the ultimate record of its evolutionary history (Doolittle, 1989). Most contemporary gene products are the result of past gene duplications and subsequent divergence resulting from gradual amino acid replacement (Doolittle, 1990).

Comparative analysis of molecular sequences has become a powerful tool in reconstructing the process of evolution (Wilson *et al.*, 1977; Felsenstein, 1988). These workers have pointed out that phylogeny based upon a quantitative determination of the very events which permit speciation, namely mutations, must ultimately be capable of providing the most accurate phylogenetic trees.

Fitch and Margoliash (1967) demonstrated that amino acid sequences of cytochrome *c* from different animals produced a phylogeny that corresponded well with phylogenetic relationships developed from data gathered from many disciplines over decades of time. This indicated for the first time that the information contained within a single protein molecule could be used to trace the evolutionary history of organisms. Subsequent to this pioneering study, analysis of many other protein sequences has indicated the value of constructing evolutionary trees on the basis of data gathered from related molecules found in different species (Felsenstein, 1988). Many of these studies have indicated that in most cases the deduced phylogenies bear a strong resemblance to evolutionary tress constructed from classical comparative anatomy.

Besides its obvious taxonomic value, an appreciation of how proteins have changed during evolution may enable one to infer the structure of an unknown protein on the basis of amino acid sequence alone, provided that a structure from a related protein is available (Doolittle, 1989).

One possible mechanism whereby new proteins may arise is by duplication of an existing gene, followed by mutational divergence of the two new genes and their protein products. This enables the original function of the one gene to be maintained while the other may mutate to fulfil a different function. This accounts for the existence of protein families comprising, for example, different dehydrogenases, proteases, phosphatases or kinases (Darby and Creighton, 1993). For example, the NAD-dependent dehydrogenases comprise a family of enzymes that bind NAD through a common structural framework, the dinucleotide-binding fold (Section 2.5.1.3). This framework binds NAD in a very similar way in these enzymes despite very different amino acid sequences. The NAD-binding domains are linked to catalytic domains, which have different structures. These enzymes are likely to have evolved by gene fusion of a common ancestral gene for an NAD-binding protein with different genes encoding metabolite-binding proteins (Branden and Tooze, 1991).

Evolutionary divergence has resulted in many forms of essentially the same protein, all with the same biological function in different species. The only constraint in this divergence is that functionality of the enzyme be maintained. If structural changes do no harm, they have a good chance of passing through the filter of natural selection (Doolittle, 1989). Therefore, the nature of the evolutionary divergence permitted in amino acid sequence reflects

constraints of structure and function on enzyme activity.

The most frequently encountered amino acid replacements during protein evolution are those conserving the type of side chain, for example, arginine for lysine, aspartic acid for glutamic acid or phenylalanine for tyrosine (Darby and Creighton, 1993). However, functionally important residues in enzymes, such as those involved in the binding of substrates or in catalysis, are the most highly conserved (Darby and Creighton, 1993). Therefore, a high degree of conservation of an amino acid residue in homologous enzymes from different sources is likely to indicate that a residue is important in contributing to the active site or participating in the catalytic mechanism. For example, all of the amino acid residues involved in the active site or otherwise participating in the catalytic mechanism of human glutathione reductase (Pai and Schultz, 1983; Pai *et al.*, 1988) are conserved in the *E. coli* enzyme with the exception of a single conservative substitution (Greer and Perham, 1986).

The vast majority of evolutionary change appears to be functionally neutral; natural selection has been primarily negative in selecting against mutations that produce functional change (Darby and Creighton, 1993). Point mutations and insertions are most common in non-functional residues, such as those found in random coil structures at the surface (Section 2.5.1.1). When homologous amino acid sequences from different species are compared, it is frequently found that insertions and deletions of a few residues occur almost exclusively in loop regions (Branden and Tooze, 1991). During evolution, the hydrophobic cores of proteins appear to be much more stable than surface residues. Presumably, mutations in hydrophobic cores are rare because they disrupt critical packing interactions required for conformational stability (Darby and Creighton, 1993). In contrast, substitutions on the surface area are usually innocuous.

Examination of related proteins reveals that while general folding patterns may be preserved, there are distortions in the primary structure which increase in magnitude as the amino acid sequences diverge. These distortions are not uniformly distributed. In any family of proteins, there is invariably a core of structure that retains the same qualitative fold, while other parts of the structure may change conformation radically. Nevertheless, given enough time, relationships between proteins at the level of primary structure can be blurred to the point where recognition is no longer possible (Doolittle, 1989).

Alternatively, the absence of sequence homology between proteins of similar structure may be ascribed to them having arisen independently through convergent evolution (Darby and Creighton, 1993). There are a large number of ways in which the twenty amino acids found in proteins may be assembled into equivalent structures, and two or more very dissimilar amino acid sequences may give rise to very similar three-dimensional structures. Similar sequences may thus result from convergence as a result of selection for particular structures (Doolittle, 1989).

A fundamental principle that has emerged from structural analyses of protein structures is the occurrence of recurring structural motifs in proteins (Branden and Tooze, 1991). The observation that proteins of unrelated functions are often built up from combinations of these recurrent motifs suggests that certain structural themes have been used time and again throughout the course of evolution (Branden and Tooze, 1991). The $(\alpha/\beta)_8$ barrel, originally identified in triose phosphate isomerase, has been proposed as a common protein fold that has arisen by convergent evolution (Branden and Tooze, 1991). Many of the seventeen known examples of this fold have no sequence homology and only slight functional similarities (Darby and Creighton, 1993). The $(\alpha/\beta)_8$ barrel is a plausible candidate for having arisen independently in several proteins as it represents one of the simplest thermodynamically stable structures that an α/β protein can adopt (Branden and Tooze, 1991).

Although many common motifs within proteins may have arisen by convergence, proteins which exhibit similarity over the full course of their lengths must be the result of divergence. The greater the number of residues over which a resemblance extends, the less likely it is to be the result of chance or convergence (Doolittle, 1989).

Probably the most evident effect of evolutionary history on protein structure is the prevalence of approximately duplicated parts within different proteins (Darby and Creighton, 1993). Domains, and perhaps smaller sequences as well, are apparently interchanged within and between proteins (Doolittle, 1989). Doolittle (1985) has noted that many of the proteins which have probably arisen fairly recently in evolutionary terms are mosaics, apparently built up by linking together two or more domains that have been duplicated from other proteins. These proteins appear to have been generated genetically by a process of domain shuffling.

Domains are often encoded at the gene level by individual exons (Blake, 1978, 1979; Gilbert, 1978; Traut, 1986). The evolutionary significance of splitting eukaryotic genes into coding sequences (exons) and noncoding sequences (introns) was first proposed by Tonegawa *et al.* (1978). In this scheme, individual exons correspond to functional or structural units of proteins. Recombination within the intronic regions of genes could reassort these functions into novel protein products. Alternatively, intron deletion may elicit shuffling of a functional unit encoded by a single exon. Point mutations at intron/exon splice junctions could result in the deletion or addition of whole blocks of amino acid sequences, while variable splicing could enable transcription of both the original and new gene products simultaneously (Gilbert, 1978; Tonegawa *et al.*, 1978). Therefore, genes for proteins may have evolved by the genetic shuffling of different exons to generate different proteins composed of different combinations of structural units (Gilbert, 1978; Blake, 1979).

In accordance with this hypothesis, it has frequently been observed that introns are often found in sites of structural genes corresponding to loop regions in the protein structure (Branden and Tooze, 1991). During evolution, hydrophobic cores are generally more stable than residues exposed to the surface (Branden and Tooze, 1991). Craik *et al.* (1983) have speculated that a reason why intron-exon splice junction frequently map to surface loops is that the alterations mediated by the sliding of these junctions can be effected without disrupting the stability of the protein core. This would invariably destroy enzymatic activity.

Some evidence in favour of the hypothesis of domain shuffling exists for certain proteins. For example, the exons of the gene encoding mouse β-crystallin encode the four structural motifs of the corresponding protein. Three introns are present at the junctions between the motifs (Branden and Tooze, 1991). Examination of the genomic sequence of the gene encoding a plasma membrane H^+-ATPase in *Arabidopsis* has revealed partial conservation of exon boundaries with respect to animal (Na^+/K^+)- and Ca^{2+}-ATPases (Pardo and Serrano, 1989). This suggests a degenerate correspondence between exons and the structural motifs found in eukaryotic ATPases.

However, although many proteins appear to have been generated by the shuffling of exons, the generality of this phenomenon is still being debated (Blake, 1985). At present, there is

no well defined universal relationship between exons and protein structure.

3. Materials and Methods

3.1 Reagents

All common laboratory chemicals used were of the highest quality available. Agarose was purchased from Promega (Madison, WI). Sequencing grade polyacrylamide was purchased from Stratagene Inc. (La Jolla, CA). GelSlick™ is a product of AT Biochemicals Inc. (Malvern, PA). Unless otherwise specified, all enzymes, antibiotics, sequencing primers and other reagents used in molecular protocols were obtained from Boehringer Mannheim, South Africa.

The plasmids pBluescriptII® SK⁻ and pBluescriptII® SK⁺ and helper phage M13 R408 are products of Stratagene Inc. (La Jolla, CA). The plasmid pUC18 was obtained from Boehringer Mannheim, S.A. Seeds of *Arabidopsis thaliana* (L.) Heynh. ecotype Columbia were purchased from Lehle Seeds (Tucson, AZ).

The Digoxigenin (DIG) Nucleic Acid Labelling and Detection Kits were obtained from Boehringer-Mannheim, S.A. The Sequenase® Version 2.0 Sequencing Kit is a product of US Biochemical Corp. (Cleveland, OH). The Geneclean DNA Purification System was obtained from Bio 101 Inc. (La Jolla, CA). The ECL™ Gene Detection Kit, Hybond N⁺ nylon membranes, [α-^{35}S]dATP and Hyperfilm-ECL autoradiography film are products of Amersham International (Buckinghamshire, England).

Unless otherwise specified, all reagents were used according to the instructions of the manufacturer.

3.2 Bacterial strains

Escherichia coli strain HB101 (*hsdR hsdM proA2 lacZ24 leuB thi-1 rpsL20 supE44 recA13*) was used for transformation with both YAP057 and FAFJ25 cDNA clones. The *E. coli* strain XL1-Blue (*endA1 hsdR17(rk⁻mk⁺) supE44 thi-1 λ⁻ recA1 gyrA96 relA1 (lac⁻)* [F',

proAB, lacIqZΔM15, Tn10 (tet^r)]) was used as the host for the plasmids bearing the six fragments of the YAP057 insert subcloned for sequencing internal regions within the insert of YAP057. All bacteria were maintained on the appropriate selective media at all times.

3.3 Plant growth

Seeds of *Arabidopsis thaliana* (L.) Heynh. ecotype Columbia were germinated on a mixture of peat-moss, vermiculite and perlite (1:1:1) and fertilised weekly with a complete nutrient solution (Somerville and Ogren, 1982; Appendix 1) in a thermostatically controlled greenhouse. Greenhouse temperatures varied between 18°C and 26°C. The transmittance of the greenhouse to solar radiation was 35%.

3.4 Proline assays

3.4.1. Determination of tissue proline contents

Proline was assayed as described by Bates *et al.* (1973). Depending on the tissue being assayed, 0.5-1.0 g of plant material was frozen in liquid nitrogen and ground to a fine powder with a mortar and pestle. When necessary, frozen material was stored at -70°C until needed. The material was homogenised in 10 ml of 3% (w/v) sulphosalicylic acid and the homogenate centrifuged at 14 000 x *g* for 20 min at 4°C. Two ml aliquots of supernatant solution were transferred to clean tubes and reacted with 2 ml of acid ninhydrin reagent (Appendix 1) and 2 ml of glacial acetic acid for 1 h at 100°C. The reaction was terminated in an ice bath. After warming to room temperature, the reaction mixture was extracted with 4 ml of toluene by vigorous vortexing for 15 s. The toluene phase containing the red chromophore was aspirated and its absorbance read at 520 nm using toluene as a blank. Proline concentration was determined using a standard curve constructed from A_{520} readings obtained from assaying 0.1 to 40 μmol of proline. In cases where proline concentrations were beyond the linear range of the standard curve, samples were diluted and values obtained were multiplied by the appropriate dilution factor.

3.4.2. Osmotic stress treatments

In order to establish whether *A. thaliana* accumulates proline in response to hyperosmotic stress, five week old plants were transferred to a controlled environment chamber (22°C; 70% relative humidity; 200 $\mu E.m^{-2}.s^{-1}$ continuous light) and allowed to acclimate for 3 days. Plants were watered with nutrient solution upon transfer to the growth chamber.

In order to assess whether the rate of imposition of stress affected the amount of proline accumulated, plants were watered with solutions of increasing concentrations of NaCl (50 mM NaCl, 100 mM NaCl and 200 mM NaCl) at 8 h, 12 h and 16 h intervals. Salt solutions were prepared in distilled water (dH_2O). In order to establish whether the nature of the osmoticum used affected proline accumulation, additional plants were watered with solutions of polyethylene glycol (PEG-6000) at 16 h intervals. The concentrations of polyethylene glycol used provided solutions with solute potentials equivalent to those used in the salinisation experiments (Lang, 1967; Michel and Kaufmann, 1973). Solutions of PEG-6000 were prepared in dH_2O. Control plants were watered with dH_2O that was not supplemented with any osmoticum. In all cases, plants were watered until the solution applied began to emerge from the bottom of the pot.

Six h after the last application of osmoticum, basal rosette leaves, stems sections and siliques were harvested, weighed and their proline content determined as outlined above. All determinations were performed in triplicate.

3.4.3. Organ-specific proline levels in unstressed plants

Leaves from the basal rosette, roots, stem segments, siliques, flowers and ripe seeds were assayed for their proline content as outlined above. All determinations were conducted in triplicate.

3.5 Sources of *Arabidopsis thaliana* cDNA clones encoding P5CR

The cDNA clones YAP057 and FAFJ25 were isolated as part of the French *Arabidopsis* cDNA sequencing initiative sponsored by the Centre National de la Recherche Scientifique (Höfte *et al.*, 1993). Both clones are expressed sequence tags (ESTs; Section 2.3). Preliminary identification of these clones was based on homology of a short stretch of their sequences to plant P5CRs of known sequence.

The cDNA clone YAP057 (ATTS0197; EMBL Accession No. Z17623) was obtained from M. Raynal (Laboratoire de Physiologie et Biologie Moléculaire Végétales, Université de Perpignan, France). It was isolated from a library constructed from developing siliques of *A. thaliana* ecotype Columbia (Giraudat *et al.*, 1992). The insert of YAP057 was cloned into the *Eco*RI site of Lambda ZAPII (Stratagene, La Jolla, CA). The phagemid pBluescriptII SK⁻ was subsequently excised from this phage vector following global *in vivo* excision of the λZAPII library (M. Raynal, pers. commun.). Sequence determination of the 355 nucleotides closest to the T7 priming site of this vector revealed the translation product of one open reading frame (ORF) to have high homology to the N-terminal regions of the translation products of the genes encoding P5CR from soybean (Delauney *et al.*, 1990) and pea (Williamson and Slocum, 1992).

The cDNA clone FAFJ25 (ATTS3034; EMBL Accession No. Z33985) was obtained from J. Fleck (Institut de Biologie Moléculaire des Plantes, Strasbourg, France). It was isolated from the Strasbourg-A library constructed from green leaf strips of *A. thaliana* ecotype Columbia incubated in liquid culture medium (Y. Parmentier, pers. commun.*). The insert of FAFJ25 was directionally cloned into the *Eco*RI and *Xho*I sites of pBluescriptII SK⁻. Sequence determination of the 365 nucleotides closest to the T7 priming site of pBluescriptII SK⁻ revealed it to possess high homology to part of the gene encoding P5CR from *Arabidopsis* (Verbruggen *et al.*, 1993).

---

* Y. Parmentier, Institut de Biologie Moléculaires des Plantes, Strasbourg, France

For both YAP057 and FAFJ25, the 2 μg of lyophilised DNA provided was suspended in 20 μl of dH_2O and a 5 μl aliquot used to transform the *E. coli* strain HB101. Cultures of bacteria carrying YAP057 or FAFJ25 were grown in LB broth (Appendix 1) supplemented with 100 $\mu g.ml^{-1}$ of ampicillin and stored at -70°C after dilution with an equal volume of Glycerol Freezer Store Solution (Appendix 1).

3.6 Transformation of *E. coli* with plasmid DNA

Transformation of *E. coli* with plasmid DNA was performed using a modification of the method described by Manniatis *et al.* (1989). An overnight culture of the appropriate strain of *E. coli* was grown in LB broth at 37°C. Thirty ml of fresh LB broth in a 250 ml flask was inoculated with 200 μl of overnight culture and grown at 37°C with gentle agitation until the culture reached mid-log phase (A_{540} = 0.5). One ml aliquots of cells were centrifuged for 1 min (6 000 x *g*) and the supernatant solution decanted. The cell pellet was resuspended in 400 μl of 0.1 M $MgCl_2$ and the cells pelleted by centrifugation (6 000 x *g*, 1 min). After decanting the supernate, the cells were resuspended in 200 μl of 0.1 M $CaCl_2$ and left on ice for 3 h.

A volume of aqueous DNA solution no greater than 20 μl containing at least 100 ng of DNA was added to the competent cells and the contents of the tube mixed gently. The mixture of competent cells and DNA was incubated on ice for 30 min. During this time, a small volume of LB broth was warmed to 37°C. The competent cells were heat-shocked by direct transfer to 42°C for 1 min. One ml of LB broth, pre-warmed to 37°C, was added to the transformed cells. The cells were left to incubate at 37°C for 20 min. Two hundred μl of cell suspension was spread over LB plates containing suitable selective antibiotic(s) and the plates incubated overnight at 37°C.

3.7 Plasmid miniprep

Plasmid minipreps were performed according to the modified alkaline lysis method of Feliciello and Chinali (1993). A single bacterial colony was inoculated into 20 ml of LB

containing the appropriate antibiotic(s) in a 100 ml flask. After overnight growth at 37°C, cells were pelleted by centrifugation in a Sorvall SS34 rotor (10 000 x *g*, 10 min). The supernatant solution was discarded and the cells resuspended in 2 ml of ice-cold STE buffer (Appendix 1).

The cell suspension was halved, with two 1 ml aliquots each being transferred into a 1.5 ml microfuge tube and kept on ice. Following centrifugation in a microfuge (6 000 x *g*, 1 min), the supernatant was removed and the pellet resuspended completely in 250 µl of ice-cold solution I (Appendix 1) by repeated vigorous vortexing. Bacterial suspensions were kept on ice. Five hundred µl of freshly prepared solution II (Appendix 1) was added to each tube. The contents of the tube were mixed immediately by inverting the tube three to four times. Tubes were stored on ice for 3-5 min. Seven hundred and fifty µl of ice-cold 4M potassium acetate-2M acetic acid (Appendix 1) was added to each tube. Tubes were shaken immediately and vigorously by hand to mix the contents thoroughly and left on ice for 10 min.

Following centrifugation (12 000 x *g*, 10 min), 1.4 ml of supernatant solution was transferred to a 2 ml microfuge tube, taking care to avoid collecting any denatured material that had not been pelleted. Seven hundred µl of isopropanol was added to each tube and mixed well by inversion. Samples were left for 5 min at room temperature, prior to centrifugation (12 000 x *g*, 10 min). The supernate was decanted, and each tube placed back in the original orientation. After brief centrifugation (12 000 x *g*, 10 s), all remaining supernatant solution was removed using a 1 ml disposable syringe.

Pellets were resuspended in 250 µl of TE buffer (Appendix 1) containing 100 $\mu g.ml^{-1}$ DNase-free RNase (Appendix 1) and left at room temperature for 15 min. Three hundred µl of an 88% isopropanol-0.2M potassium acetate solution (Appendix 1) was added. After mixing by brief vortexing, tubes were left at room temperature for 10 min. Following centrifugation (12 000 x *g*, 10 min), the supernatant solution was discarded. Tubes were recentrifuged briefly and residual supernate withdrawn using a 1 ml disposable syringe. After washing the pellets in 70% ethanol to remove any residual salt, pellets were dried in a vacuum dessicator and resuspended in 100 µl of TE buffer. Plasmid DNA was stored at -20°C.

3.8 Purification of DNA fragments

Inserts or gene fragments generated by restriction digestion were separated on a 1.1% agarose gel by electrophoresis in 1x TAE buffer prepared from a 50x TAE buffer stock (Appendix 1). The DNA fragments were purified using the Geneclean system (Bio101 Inc., La Jolla, CA). Following visualisation of the DNA on a Spectroline TC-312A UV transilluminator (312 nm), the fragment of interest was excised from the agarose gel. Care was taken to trim off as much of the agarose as possible which did not contain fluorescent DNA. Less than 400 µl of gel (of weight less than 0.4 g) was transferred to a 1.5 ml microfuge tube and exactly three volumes of a NaI solution (Bio101 Inc.) added. The sample was incubated at 50°C with occasional gentle agitation until all the agarose had dissolved. Care was taken not to exceed a 5 min incubation at this temperature.

A solution of glass milk (Bio101 Inc.) was vortexed well for at least 30 s before immediately adding 5 µl of the glassmilk solution to the dissolved agarose and mixing well. The suspension was incubated at room temperature for 5 min. The side of the tube was tapped gently at 1 min intervals to ensure complete suspension of the glass particles.

The glass milk was pelleted by brief centrifugation (6 000 x g, 5 s) and washed with 500 µl of ice cold NEW wash buffer (Bio101 Inc.). The pellet was resuspended in fresh NEW wash buffer and the process of centrifugation and resuspension repeated. Following a third centrifugation (6 000 x g, 5 s), all traces of NEW wash buffer were removed by aspiration using a 1 ml disposable syringe. The DNA bound to the glass milk was eluted into 20 µl of dH_2O by incubation at 50°C for 2 min. The glass milk was pelleted (6 000 x g, 10 s) and the DNA solution transferred to a clean microfuge tube for subsequent manipulation.

3.9 DNA ligation

Plasmid DNA was digested with the relevant restriction enzyme(s) and the resulting fragments fractionated on a 1.5% agarose gel. In cases where the two restriction enzymes did not exhibit optimal activity in the same restriction buffer, the DNA was digested

sequentially with both enzymes. Following the first digestion, DNA was precipitated by addition of 0.1 vol of 3M sodium acetate (pH 5.2) and 1 vol of ethanol. The restricted DNA was then resuspended in an appropriate restriction buffer and digested with the second enzyme. Following precipitation as outlined above, the pellet was washed in 70% ethanol and the DNA dried in a vacuum dessicator prior to resuspension in an appropriate amount of dH_2O. Following electrophoresis (80V) through a 1.1% agarose gel in 1x TAE buffer, the sizes of fragments were determined by comparison of their migration with that of the fragments of Molecular Weight Marker III (Boehringer Mannheim; Appendix 1). Gene fragments of interest were recovered and purified as described in Section 3.8.

The ends of the digested vector DNA were dephosphorylated by adding 1 unit of calf intestinal alkaline phosphatase (Boehringer Mannheim) to 5 μg of vector DNA in a total volume of 20 μl of dephosphorylation buffer (Boehringer Mannheim). Following incubation for 1 h at 37°C, the alkaline phosphatase was inactivated by heating the reaction to 75°C for 15 min. The sample was left to cool to room temperature. The DNA was precipitated by addition of 0.1 vol of 3 M sodium acetate (pH 5.2) and 1 vol of ethanol. Residual salt was removed by washing the pellet with 70% ethanol. The DNA pellet was resuspended in 20 μl of dH_2O and quantified by measuring the absorbance at 260 nm using a Pharmacia GeneQuant RNA/DNA Calculator. The concentration of DNA was calculated as outlined in Appendix 1. The digested vector DNA was diluted to a final concentration of 0.15 $\mu g.\mu l^{-1}$.

All ligation reactions contained 1 unit of T4 DNA ligase (Boehringer Mannheim), 2 μl T4 DNA ligase buffer (Boehringer Mannheim) and 1 μl of 20 mM dATP, prepared by five-fold dilution of a 100 mM dATP stock (Boehringer Mannheim) in dH_2O. Reactions were made up to a final volume of 20 μl with the DNA to be ligated. In each experiment, ratios of vector:insert of 1:15, 7:8 and 15:1 were used. The concentration of the DNA fragment to be inserted was 0.05 $\mu g.ml^{-1}$.

As a negative control, a ligation reaction containing 16 μl of digested vector DNA was set up. In cases where vector DNA was digested with a single restriction enzyme, this indicated whether or not the dephosphorylation step was effective. In ligations involving directional cloning, this indicated whether or not both sites were digested to completion.

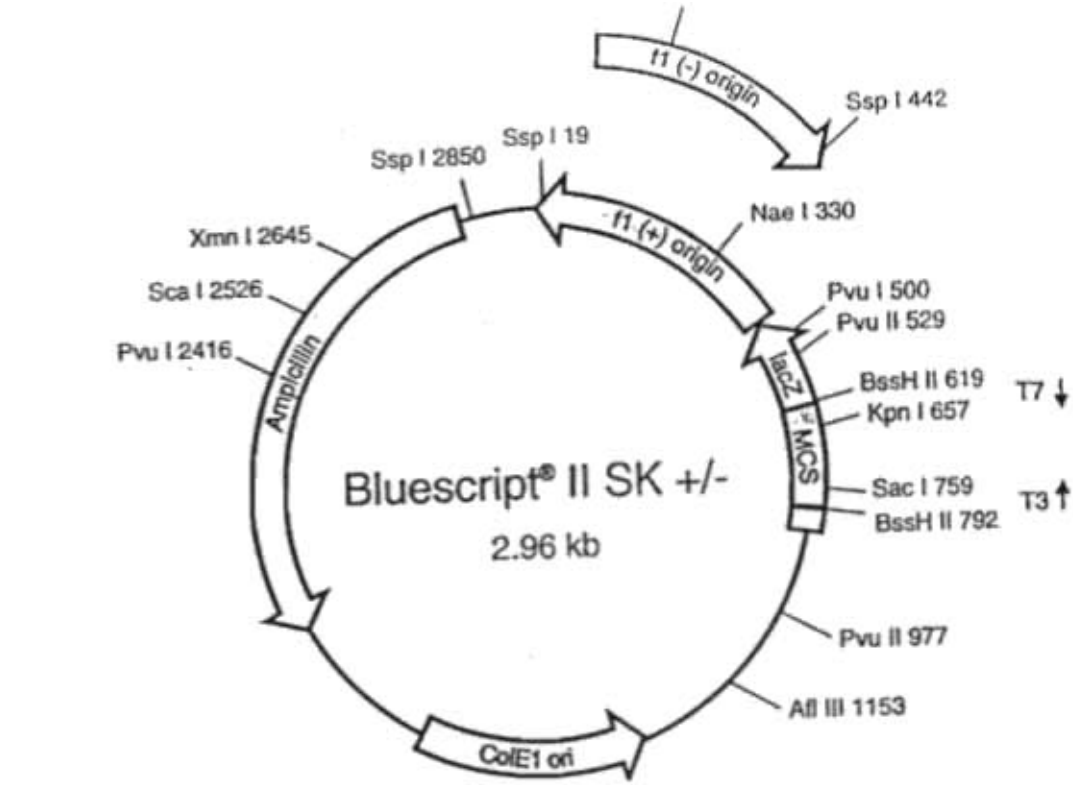
Following overnight ligation at 10°C, 200 µl aliquots of competent *E. coli* XL1-Blue cells (prepared as outlined in Section 3.6) were transformed with the entire ligation reaction (20 µl) and plated onto LB plates containing 100 $\mu g.ml^{-1}$ ampicillin and 12.5 $\mu g.ml^{-1}$ tetracycline. Thirty min prior to plating cells, 50 µl of 2% (w/v) 5-bromo-4-chloro-3-indoyl-β-D-galactoside (X-Gal) dissolved in dimethylformamide and 10 µl of 100 mM isopropylthiogalactoside (IPTG) were spread over the surface of the agar to enable detection of recombinants by blue/white colour selection. Blue colonies assumed their characteristic colouration owing to their ability to cleave the chromogeneic substrate X-Gal in the presence of IPTG, an analogue of lactose. Since insertion of a fragment of foreign DNA into the polycloning site of pBluescript invariably results in production of an incompetent *lacZ'* gene product, recombinant cells were unable to metabolise X-Gal and therefore appeared white.

After overnight growth at 37°C, plates were incubated at 4°C for 10 h to enhance the formation of the blue colour. White colonies were picked and restreaked onto the same media to confirm that they did not contain a functional *lacZ'* gene. In all experiments, at least 15% of the colonies carried a non-functional *lacZ'* gene. Following plasmid isolation from selected white colonies, these were tested for the presence of an appropriately-sized insert by restriction analysis using appropriate restriction enzymes.

3.10 Single-stranded DNA isolation

Single-stranded DNA for sequencing was isolated by rescue from pBluescriptII phagemid as described by Stratagene. The process is dependent on sequences originating from filamentous (f1) bacteriophage that have been placed within the phagemid (Figure 3.1), and on the presence of a variety of f1 bacteriophage-derived proteins provided by f1 helper phage (Stratagene).

A single colony harbouring the phagemid bearing the appropriate insert, was grown overnight in 10 ml LB in the presence of 100 $\mu g.ml^{-1}$ ampicillin and 12.5 $\mu g.ml^{-1}$ tetracycline. Selection for the F' episome in the presence of tetracycline was essential as it contains the genes for expression of the pili required for infection by filamentous phage



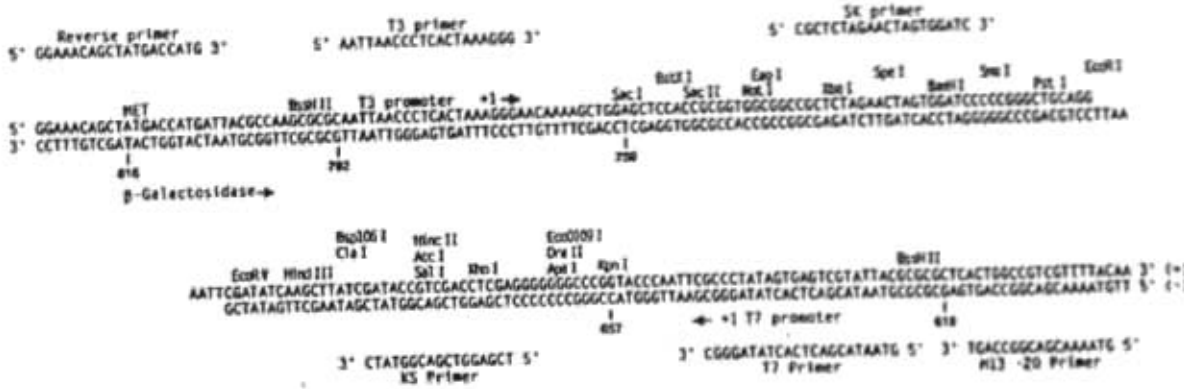


Figure 3.1 : The pBluescript®II SK (+/-) phagemid (Stratagene). The SK designation indicates that the polylinker is oriented such that *lacZ* transcription proceeds from *Sac*I to *Kpn*I. The *lacZ* gene encoding β-glucuronidase provides α-complementation for blue/white colour selection of recombinant phagemids. The f1(+) and f1(-) orgins of replication are found in the phagemids pBluescript SK⁺ and pBluescript SK⁻ respectively. In pBluescript SK⁺, the f1(+) origin of replication enables recovery of the sense strand of the *lacZ* gene when a host strain containing the pBluescriptII phagemid is co-infected with helper phage. In pBluescript SK⁻, the f1(-) origin of replication enables recovery of the antisense strand of the *lacZ* gene when a host strain containing the pBluescriptII phagemid is co-infected with helper phage. The ColE1 origin of replication is used in the absence of helper phage (Stratagene).

The sequence of the multiple cloning site (MCS) containing restriction sites used for subcloning fragments of YAP057 is indicated below the map of the plasmid. Also shown are the priming sites used in sequencing reactions. The upper strand is designated the (+) strand and the lower strand is designated the (-) strand.

(Stratagene). The following morning, 10 ml of Superbroth (Appendix 1) in a 100 ml flask was inoculated with 300 µl of the overnight culture and grown with shaking at 50 strokes per min in a 37°C waterbath. After approximately 2 h, the culture had an absorbance reading of 0.3 at 600 nm.

Helper phage M13 R408 (Stratagene) was added to the culture using a multiplicity of infection (phage:cells) of 20:1. Under the growth conditions described, at A_{600}=0.3, there were approximately 2.5 x 10^8 cells.ml^{-1}. Throughout the period of experimentation, the titre of the stock of M13 R408 used remained approximately 3.0 x 10^{11} pfu.ml^{-1}. After addition of the helper phage, the culture was shaken at 37°C (50 strokes.min^{-1}) for 8 h.

The culture was heated to 65°C for 15 min to kill the cells, centrifuged in a sterile tube (10 000 x *g*, 10 min) and the supernatant solution stored overnight in a sterile flask at 4°C. Heating to 65°C does not affect the viability of the helper phage (Stratagene).

Aliquots of 1.2 ml of supernate were added to 1.5 ml microfuge tubes and 300 µl of M13 precipitation solution (Appendix 1) added to each tube. Tubes were inverted to mix their contents thoroughly, and left for 15 min at room temperature to allow the phage to precipitate. After pelleting the precipitated phage (12 000 x *g*, 15 min), the pellet was resuspended in 300 µl of TE buffer.

An equal volume of equilibrated phenol:chloroform:isoamyl alcohol (25:24:1) was added to each tube and the DNA extracted by vigorous vortexing for 1 min, followed by separation of the organic and aqueous phases by centrifugation (12 000 x *g*, 15 min). Phenol was equilibrated as outlined in Appendix 1. Extraction with phenol:chloroform:isoamyl alcohol was repeated until no white interface between the two phases remained evident. This was followed by an extraction with chloroform to remove any residual phenol.

Single-stranded DNA was precipitated by addition of an equal vol of 7.5 M ammonium acetate (pH 7.5) and 2 vol of cold ethanol. Samples were left on ice for 15 min prior to pelleting by centrifugation (12 000 x *g*, 15 min). Supernatant solution was decanted, the tubes centrifuged briefly (6 000 x *g*, 5 s) and the last remnants of supernate withdrawn using a 1 ml disposable syringe. The precipitated DNA was washed with 80% ethanol prior to vacuum drying and resuspension in 10 µl of TE buffer. The DNA was then quantified

3.11 Sequencing

Sequencing was performed using the dideoxy chain-termination method (Sanger *et al.*, 1977). All sequencing was conducted using the Sequenase® Version 2.0 sequencing kit (US Biochemical Corp.). Sequenase Version 2.0 polymerase is a genetic variant of bacteriophage T7 DNA polymerase created by *in vitro* genetic manipulation (Tabor and Richardson, 1987, 1989).

Both single-stranded and double-stranded sequencing reactions were performed according to the procedures suggested by the manufacturers. However, template preparation for sequencing was by methods different to those recommended by US Biochemical Corp. Single stranded template was prepared as outlined in Section 3.10 and double stranded template was prepared as described in Section 3.7.

The T3 (AATTAACCCTCACTAAAGGG) and T7 (TAATACGACTCACTATAGGG) promoter-specific primers were obtained from Boehringer Mannheim. Clone pPEA was sequenced using the M13 forward primer (GTAAAACGACGGCCAGT) obtained from Boehringer Mannheim. Sequencing reactions were separated on a Hoefer Poker Face II apparatus, connected to a Hoefer PS 2500 DC power supply.

3.11.1 8% Polyacrylamide gel preparation

Seventy five g of urea was dissolved in a mixture of 30 ml 40% acrylamide:*N,N'*-methylene-bis-acrylamide (19:1) mixture, 15 ml 10x TBE (Appendix 1) and 45 ml of dH_2O. The solution was made up to a final volume of 150 ml with dH_2O. After degassing under vacuum for 20 min, 1.5 ml of freshly-prepared 10% ammonium persulphate solution and 23 µl of *N,N,N',N'*-tetramethylethylenediamine (TEMED) were added to the solution. After complete mixing by vigorous swirling, the solution was poured between the two glass plates and left to polymerise at room temperature. Both glass sequencing plates were pre-treated with GelSlick™ (AT Biochemicals Inc.) to facilitate easy removal of the gel following electrophoresis.

3.11.2 Annealing of single-stranded templates to primer

For each set of four sequencing lanes, a single annealing (and subsequent labelling) reaction was used. One µl of primer (0.5 pmol.µl^{-1}), 2 µl of 5x Sequenase buffer (US Biochemical Corp.; Appendix 1) and 7 µl of DNA solution (containing 1 µg single-stranded DNA template) were combined in a microfuge tube. The capped tube was warmed to 65°C for 2 min, and then allowed to cool slowly to room temperature in 250 ml of water that had been pre-heated to 65°C. Once the temperature was below 35°C, the tube containing the template annealed to primer was placed on ice and labelled within 4 h.

3.11.3 Denaturation and annealing of double-stranded templates

Plasmid DNA was purified as described in Section 3.7. Five µg of plasmid was denatured by addition of 0.1 vol of 2M NaOH, 2 mM EDTA and incubation at 37°C for 30 min. The mixture was neutralised by adding 0.1 vol of 3 M sodium acetate (pH 5.5) and the DNA precipitated with 2 vol of ethanol at -70°C for 15 min. After washing the DNA with 70% ethanol, it was redissolved in 7 µl of dH_2O, and 2 µl of 5x Sequenase buffer and 1 µl of the appropriate primer were added. Annealing was performed as described for single-stranded template (Section 3.11.2).

3.11.4 Labelling

The 5x dGTP labelling mix (US Biochemical Corp.; Appendix 1) was diluted five-fold with dH_2O and Sequenase Version 2.0 polymerase (US Biochemical Corp.; Appendix 1) was diluted 1:8 in ice-cold enzyme dilution buffer (US Biochemical Corp.; Appendix 1). Per sequencing reaction, 1 µl of 0.1 M dithiothreitol (US Biochemical Corp.), 2 µl of diluted dGTP labelling mix, 0.5 µl of [α^{35}S]dATP (specific activity > 1 000 Ci.mmol^{-1}; Amersham) and 2 µl of diluted Sequenase polymerase were added to the 10 µl of ice-cold annealed DNA mixture. The contents of the tube were mixed and incubated on ice for 5 min. ^{35}S-labelled dATP was used in preference to ^{32}P-labelled nucleotide as ^{35}S has a longer half life and provides superior autoradiographic resolution (Biggin *et al.*, 1983).

3.11.5 Termination

Labelling reaction (3.5 µl) was transferred to each of 4 tubes each containing 2.5 µl of one of the ddGTP, ddATP, ddTTP and ddCTP termination mixes (US Biochemical Corp.; Appendix 1). The contents of the tubes were mixed and incubated for 5 min at 37°C. Since no problems with compressions were encountered, the dITP termination mixes supplied in the kit were not used.

Reactions were stopped by addition of 4 µl of stop solution (US Biochemical Corp.; Appendix 1). Samples were heated to 75°C for 2 min before loading 3 µl immediately onto a sequencing gel which had been pre-run for at least 30 min at a constant power of 60 W. Sequencing gels were run using 1x TBE buffer.

Gels were run at a constant power of 60 W. Three 3 µl aliquots of each of the four reaction mixes were loaded at 2.5 h intervals. This enabled the determination of up to 265 nucleotides from a single set of sequencing reactions, while still permitting an overlap of at least 30 nucleotides between successive loadings. Under the conditions described, nucleotide sequence could routinely be read to within 30 nucleotides downstream from the priming site. In all sequencing experiments, the interface between the vector and the insert could be discerned from the third loading.

Gels were transferred onto Whatman 3 MM chromatography paper and dried on a Hoefer SE 1160 slab gel drier. The dried gels were exposed to Hyperfilm β-max film (Amersham) for at least 48 h before developing the film.

3.12 Sequencing strategies

3.12.1 YAP057

The names of clones used in sequencing of YAP057 and descriptions of how they were generated are summarised in Table 3.1.

Table 3.1 : Clones used in the sequencing of YAP057. The clone YAP057 was obtained from M. Raynal (Laboratoire de Physiologie et Biologie Moléculaire Végétales, Université de Perpignan, France). All other clones were produced in this study. The phagemid vectors pBluescriptII® SK⁺ and pBluescriptII® SK⁻ (Figure 3.1) are products of Stratagene (La Jolla, CA). The plasmid vector pUC18 is a product of Boehringer Mannheim.

Clone Name	Description
YAP057	999 bp insert cloned into *Eco*RI site of pBluescriptII SK⁻
pEHIC	493 bp fragment generated by double digestion of YAP057 with *Eco*RI and *Hin*dIII cloned into the *Eco*RI and *Hin*dIII sites of pBluescriptII SK⁺
pEHsB	214 bp fragment generated by double digestion of YAP057 with *Eco*RI and *Hin*dIII cloned into the *Eco*RI and *Hin*dIII sites of pBluescriptII SK⁺
pHHA	292 bp fragment generated by digestion of YAP057 with *Hin*dIII cloned into the *Hin*dIII site of pBluescriptII SK⁻; the 5' end of the sense strand of the cDNA encoding *Arabidopsis* P5CR is closest to the T3 promoter
pHHB	292 bp fragment generated by digestion of YAP057 with *Hin*dIII cloned into the *Hin*dIII site of pBluescriptII SK⁻; the 5' end of the sense strand of the cDNA encoding *Arabidopsis* P5CR is closest to the T7 promoter
pHPA	226 bp fragment generated by double digestion of YAP057 with *Hin*dIII and *Pst*I cloned into the *Hin*dIII and *Pst*I sites of pBluescriptII SK⁻
pPEA	267 bp fragment generated by double digestion of YAP057 with *Eco*RI and *Pst*I cloned into the *Eco*RI and *Pst*I sites of pUC18

Both strands of YAP057 were sequenced using a combination of double- and single-stranded templates. The ends of the insert of YAP057 were sequenced using double-stranded template of pBluescriptII SK⁻ in which YAP057 insert had been cloned into the *Eco*RI site. The T3 and T7 promoter-specific primers (Boehringer Mannheim) were used.

A 292 bp fragment generated by digestion of YAP057 with *Hin*dIII was subcloned into pBluescriptII SK⁻ and three recombinant clones sequenced using the T3 primer. Of these, two clones, pHHA and pHHB were found to have inserted in opposite orientations. This enabled sequencing of both strands of this internal fragment using single stranded template and the T3 promoter-specific primer.

A 226 bp fragment generated by double-digestion of YAP057 with *Pst*I and *Hin*dIII was subcloned into compatible sites in pBluescriptII SK⁻ to generate clone pHPA. Single-stranded template obtained from this clone was sequenced using the T3 primer.

Double-digestion of YAP057 with *Eco*RI and *Hin*dIII yielded three fragments of sizes 214 bp, 292 bp and 493 bp. The 214 bp and 493 bp fragments were cloned into pBluescriptII SK⁺ which had been double-digested with *Eco*RI and *Hin*dIII to provide clones pEHsB and pEHlC respectively. These were sequenced with the T7 primer using single stranded template.

The close alignment of *Pst*I and *Eco*RI sites in the pBluescriptII vectors (Figure 3.1) prevents their double-digestion using these enzymes. Therefore, the 267 bp fragment generated by double digestion of YAP057 with *Pst*I and *Eco*RI was subcloned into pUC18 to generate clone pPEA. Double-stranded template was used to sequence this insert using the M13 forward primer (Boehringer Mannheim).

3.12.2 FAFJ25

In order to establish whether the insert of the clone FAFJ25 contained a more complete cDNA than that contained in YAP057, the ends of FAFJ25 were sequenced by priming off the T3 and T7 promoter sites flanking the insert. Double stranded template was used. These termini were sequenced on one strand only.

3.13 Computational sequence analysis

Computer searches of the GenBank, European Molecular Biology Laboratory (EMBL), GenPept, Swiss Prot and Protein Information Resource (PIR) sequence databases were conducted using the BLAST programs (Altschul *et al.*, 1990; Gish and States, 1993) accessed via electronic mail server (blast@ncbi.nlm.nih.gov).

Sequences were edited manually to align overlapping segments between loadings. Vector sequence at the end closest to the priming site was removed after its identification with BLASTN (Altschul *et al.*, 1994) using a cut-off PAM120 score of 150. Sequences were translated in all six reading frames and compared with the non-redundant protein database at the National Centre for Biotechnology and Information (NCBI, Bethesda, MD) using BLASTX (Altschul *et al.*, 1994). The sequences of genes encoding P5CRs and their deduced amino acid products were retrieved from the non-redundant protein database at the NCBI via electronic mail server (retrieve@ncbi.nlm.nih.gov).

The computer programs ALX3 (Gotoh, 1986), CODONS (Lloyd and Sharp, 1992), MACAW (Schuler *et al.*, 1991), PREDICT7 (Càrmenes *et al.*, 1989) and SEQAID II (D.D. Rhoads and D.J. Roufa, Kansas State University) were retrieved from the EMBL electronic mail server (netserv@embl.heidelberg.de).

Secondary structure probabilities were determined according to Chou and Fasman (1978a, 1978b) and Garnier *et al.* (1978) using numerical data generated by the computer programs SEQAID II and PREDICT7 respectively. Assignment of secondary structure using the method of Chou and Fasman (1978a, 1978b) was performed by visual inspection of the numerical data using a set of guidelines provided by Fasman (1985). These rules are outlined in Appendix 2. For Chou Fasman prediction of secondary structure, six residues were averaged for α-helix formation and four residues averaged for determinations of the formation of both β-sheet and β-turn structures. Using the method of Garnier *et al.* (1978), a window of length six amino acid residues was used for all determinations. All decision constants in predictions using the method of Garnier *et al.* (1978) were taken as zero in order to avoid overpredictions of secondary structures. Secondary structure predictions using the method of Garnier *et al.* (1978) were generated by computational interpretations of the

numerical data conducted by the computer program PREDICT7 (Càrmenes *et al.*, 1989).

Hydrophathy profiles were calculated according to Kyte and Doolittle (1982) with a window of length six amino acid residues using PREDICT7. Determinations of codon usage bias in the gene encoding *Arabidopsis* P5CR and amino acid usage in the corresponding gene product were established using the computer program CODONS.

Homology of stretches of amino acid sequences to protein domains of known structure and/or function was assessed by searching the SBASE protein domain sequence library release 3.0 (Pongor *et al.*, 1994) via automated electronic mail server (sbase@icgeb.trieste.it).

Phenogram construction was performed using the SAHN, COPH and XCOMP programs in the package NTSYSpc (Numerical Taxonomy and Multivariate Analysis System, Version 1.80, Exeter Software).

All other analysis of DNA sequences was conducted using the computer package DNAsis (Hibio).

3.14 Genomic copy number determination

3.14.1 Genomic DNA Extraction (D.M. Horvath, pers. commun.*)

Approximately 3 g of leaf tissue from the basal rosette of mature *Arabidopsis thaliana* plants was washed for 1 minute in 70% EtOH and left in a 50% dilution of Jik™ (a 1.75% (w/v) solution of sodium hypochlorite) for 10 min. Following 3 washes in a large excess of water, the tissue was dried, placed in a pre-chilled porcelain mortar and covered with liquid nitrogen. After grinding to a fine powder, the frozen material was added to 6 ml of

---

* D. M. Horvath, Laboratory of Plant Molecular Biology, Rockefeller University, New York

genomic DNA extraction buffer (Appendix 1) in a 40 ml teflon centrifuge tube. The tissue was thawed rapidly by warming the tube in a waterbath at 42°C. The tube was shaken for 10 min at 37°C waterbath. Five ml of a mixture of phenol:chloroform:isoamyl alcohol (100:100:1) was added to the suspension and the mixture vortexed for 30 s and left to shake at 37°C for a further 10 min.

After centrifugation (10 000 x g, 10 min), 5 ml of the aqueous phase was transferred to a new tube. Five hundred µl of 3 M sodium acetate (pH 5.2) and 6 ml of isopropanol were added and the tube inverted several times to ensure complete mixing. After centrifugation (10 000 x g, 10 min), the supernatant was withdrawn and the pellet washed with 70% ethanol. The pellet was resuspended in 500 µl of TE buffer and DNase-free RNase added to a final concentration of 100 $\mu g.ml^{-1}$. Following incubation for 15 min at 37°C, DNA was precipitated by the addition of 0.1 vol of 3 M sodium acetate (pH 5.2) and 1 vol of ethanol. After a brief centrifugation (12 000 x g, 1 min), the supernatant solution was discarded. The pellet was washed with 70% ethanol, dried and resuspended in 500 µl of sterile dH_2O.

3.14.2 Digestion of genomic DNA

Genomic DNA was digested with *Xho*I, *Xba*I, *Kpn*I and *Eco*RI (Boehringer Mannheim). For each digestion, 20 µg of genomic DNA was digested overnight at 37°C with 20 µl (200 units) of the appropriate enzyme in a total volume of 500 µl. Besides the inclusion of 50 µl of the appropriate 10x restriction buffer, all restriction digestions also contained a final concentration of 2.5 mM spermidine (Sigma) and 0.1 $mg.ml^{-1}$ bovine serum albumin (Boehringer Mannheim). After a 12 h incubation at 37°C, the reactions were spiked with an additional 50 units of the appropriate restriction enzyme and left to incubate for a further 2 h at 37°C. The restricted DNA was precipitated by the addition of 50 µl of 3 M sodium acetate (pH 5.2) and 550 µl of ethanol. Following centrifugation (12 000 x g, 10 min), the pellet was washed with 200 µl of 70% ethanol to remove residual salt, then dried and resuspended in 30 µl of dH_2O. Following addition of 5 µl of 10x loading buffer (Appendix 1), samples were loaded onto a 0.8% agarose gel containing 1x TBE. Samples were electrophoresed in 1x TBE at 25 V overnight and stained in a 20 $\mu g.ml^{-1}$ solution of ethidium bromide (Sigma). The gel was examined on a Spectroline TC-312A UV transilluminator to confirm that the DNA was smeared the full length of the lane.

3.14.3 Southern Hybridisation

Southern hybridisation was performed according to the instructions provided with the ECL™ direct nucleic acid labelling and detection system (Amersham).

3.14.3.1 Processing the gel

Following electrophoresis, the DNA in the gel was depurinated by placing the gel in a pyrex dish and covering it with 500 ml of freshly prepared 0.25 N HCl and agitating gently until the bromophenol blue marker had turned completely yellow. This took approximately 10 min. Following depurination, the gel was rinsed briefly with distilled water and covered with 500 ml of denaturation solution (Appendix 1). The gel was agitated gently for 15 min, the solution discarded, 500 ml of fresh denaturation solution added and agitation continued for a further 15 min. The denaturation solution was discarded and the gel rinsed briefly with 500 ml of distilled water. The dH_2O was discarded and the gel immersed in 500 ml of neutralisation solution (Appendix 1). After gentle agitation for 30 min, the neutralisation solution was replaced with an equal volume of fresh neutralisation solution and agitation continued for a further 30 min.

3.14.3.2 Transfer

Blotting was performed with slight modification to the method described by Southern (1975). A pyrex dish was partially filled with 20x SSC (Appendix 1) and a glass plate laid over the dish. This was covered with Whatman 3MM paper that had been pre-soaked in 20x SSC to serve as a wick. The gel was placed on the 3MM paper, taking care to avoid trapping air bubbles. A suitably-sized Hybond N" membrane (Amersham) was placed on top of the gel and any air bubbles removed by rolling a 10 ml glass pipette over the surface of the membrane. Gloves were worn during this and all subsequent steps involving handling of the membrane, as the oils and nucleic acids on bare hands may cause artifacts during transfer or detection (Amersham). A lead pencil was used to mark the position of the wells on the membrane and to distinguish the side of the membrane to which DNA was transferred.

The perimeter of the membrane was lined with Parafilm® (American National Can™) to ensure direct transfer of DNA to the membrane. The membrane was overlaid with three sheets of filter paper which had been soaked in 20x SSC, again taking care to remove all air bubbles. A 15 cm stack of adsorbent tissue paper was placed on top of the 3MM paper and a weight of approximately 750 g applied evenly over the blot to ensure even transfer. The system was left to blot at room temperature for 16 h.

Following blotting, the stack of paper towelling was removed and the membrane lifted off the gel using forceps. The membrane was rinsed in 6x SSC for 1 min with gentle agitation to remove agarose and left to dry DNA face up on a pad of filter paper at the back of a laminar flow bench. DNA was fixed to the membrane by baking the membrane at 80°C for 2 h.

The efficiency of transfer of the DNA was assessed by staining the compressed gel in a 20 $\mu g.ml^{-1}$ solution of ethidium bromide and visualisation on a Spectroline TC-312A UV transilluminator.

3.14.3.3 Hybridisation

Hybridisation was performed using hybridisation buffer supplied in the ECL direct nucleic acid labelling and detection kit (Amersham). Sufficient hybridisation buffer (Amersham) was used to ensure a final ratio of 0.25 ml buffer per cm^2 of membrane. Blocking agent (Amersham) and NaCl were added to final concentrations of 5% (w/v) and 0.5 M respectively. The buffer was immediately mixed thoroughly to ensure that the blocking agent formed a fine suspension. Mixing was continued at room temperature for 1 h on a magnetic stirrer. The buffer was then heated to 42°C and mixed occasionally for a further 1 h.

Hybridisation was performed in a Hybaid mini hybridisation oven. The blot carrying fixed DNA was soaked in 2x SSC and loosely rolled in a sheet of nylon mesh with the DNA surface of the blot facing inwards. The mesh was placed in a Hybaid glass hybridisation tube (4cm diameter, 30cm long) containing a small amount of 2x SSC and unrolled in the opposite direction from the movement of the rotisserie. Care was taken to exclude any bubbles between the membrane and the mesh or the hybridisation tube.

The 2x SSC was poured off and hybridisation buffer added. Following pre-hybridisation for 2 h at 42°C in the rotisserie oven, some of the buffer was withdrawn for mixing with labelled probe. The probe was added to the rest of the buffer. Care was taken not to place probe directly onto the membrane. Hybridisation was performed overnight at 42°C.

3.13.3.4 Labelling of DNA probe

During the course of prehybridisation, DNA was labelled with the enzyme horseradish peroxidase. YAP057 DNA was diluted with dH_2O to a concentration of 10 $ng.\mu l^{-1}$ and sufficient diluted DNA withdrawn to provide a final probe concentration of 10 ng probe per ml of hybridisation buffer. The DNA was denatured in a vigorously boiling water bath for 5 min and then immediately cooled on ice for 5 min. After collecting the contents at the bottom of the tube by brief centrifugation, an equivalent volume of DNA labelling reagent (Amersham) was added to the cooled DNA and the contents of the tube mixed gently but thoroughly. Gluteraldehyde solution (Amersham) was added in a volume equivalent to the volume of the labelling reagent in order to cross-link the enzyme to the probe (Amersham). After mixing by brief vortexing, the contents of the tube were collected at the bottom of the tube by centrifugation, and the mixture incubated at 37°C for 10 min. Probe was immediately diluted in an aliquot of buffer used for prehybridisation and added to the rest of the buffer in the hybridisation tube, taking care to avoid placing it directly on the membrane.

3.14.3.5 Stringency washing

Primary wash buffer (Appendix 1) was pre-heated to 42°C. After removing the hybridisation buffer from the tube, primary wash buffer was added at a volume of approximately 5 $ml.cm^{-2}$ of membrane. This was replaced with fresh primary wash buffer after 20 min washing at 42°C. Neither primary wash exceeded 20 min.

The primary wash buffer was discarded and the blot transferred from the hybridisation tube to a dish. The blot was washed with an excess of secondary wash buffer by gentle agitation at room temperature for 5 min. A further wash in fresh secondary buffer was performed at room temperature for 5 min.

3.14.3.6 Signal generation and detection

The ECL direct nucleic acid labelling and detection system is based on enhanced chemiluminescence (Whitehead *et al.*, 1983).

Equal volumes of detection reagents 1 and 2 (Amersham) were mixed to provide 0.125 ml solution per cm^2 of membrane. Secondary wash buffer was drained from the blot and the detection reagent mixture added directly to the surface carrying the DNA. Following incubation for 1 min at room temperature, excess detection reagent mixture was drained off by gently laying the blot, DNA side up, on Whatman 3MM paper for a few seconds. For optimum sensitivity, it was necessary to work with reasonable speed from this point in the detection procedure. The blot was wrapped in plastic film (ClingWrap™, Multifoil Pty Ltd., S.A.). Care was taken to gently smooth out any air pockets. The blot was placed DNA side up in a film cassette, a sheet of Hyperfilm-ECL film (Amersham) placed over it, and the cassette closed for an exposure period of 120 min.

3.15 Tissue-specific transcription of the *Arabidopsis* gene encoding P5CR

3.15.1 Water stress treatment

Mature *Arabidopsis thaliana* plants were stressed by withholding water until plants displayed symptoms of wilting. The ambient temperature of the greenhouse environment did not drop below 18°C or exceed 26°C. Under the ambient conditions at the time of experimentation, an eight day drought caused severe loss of turgor in the flowering stem. Non-stressed plants were provided with sufficient water during the treatment period to ensure that they displayed no visible signs of water deprivation.

3.15.2 Determination of relative water content (RWC)

The RWC of stem sections 3 cm long were determined by weighing the tissue sections, transferring these sections to dH_2O for 4 h (full turgor), reweighing, drying at 80°C for 60 h and weighing once more. The percentage relative water content was calculated according to the formula:

$$RWC = \frac{(fresh\ weight - dry\ weight)}{(fresh\ weight\ at\ full\ turgor) - (dry\ weight)} X\ 100$$

Four replicates of each determination of RWC were made.

3.15.3 Probe isolation and labelling

The 999 bp insert in YAP057 was isolated by digesting plasmid DNA with *Eco*RI and excising the fragment from a 1% agarose gel. Probe was purified as described in Section 3.8.

Digoxigenin-labelled probe was generated with Klenow polymerase by random primed (Feinberg and Vogelstein, 1983) incorporation of digoxigenin-labelled deoxyuridine triphosphate (DIG-dUTP). In DIG-dUTP, the steroid hapten digoxigenin is covalently linked to dUTP via a spacer arm (Boehringer Mannheim). Three µg of linearised YAP057 insert DNA was dissolved in 15 µl of dH_2O. The DNA was denatured by heating to 100°C for 10 min followed by chilling rapidly on ice. Two µl of 10x hexanucleotide mixture (DIG DNA Labelling Kit, Boehringer Mannheim), 2 µl of 10x DIG dNTP labelling mixture (DIG DNA Labelling Kit, Boehringer Mannheim; Appendix 1) and 2 units of Klenow DNA polymerase (DIG DNA Labelling Kit; Boehringer Mannheim) were added to the denatured DNA. After brief centrifugation (6 000 x *g*, 5 s) in order to settle the contents to the bottom

of the microfuge tube, the sample was incubated at 37°C for 3 h. The labelling reaction was terminated by the addition of 2 µl of 0.2 M Na_2EDTA (pH 8.0). The labelled DNA was precipitated by the addition of 2.5 µl of 4 M LiCl and 75 µl of cold ethanol and left at -20°C for 2 h. After pelleting the DNA by centrifugation (12 000 x *g*, 10 min), the pellet was washed with 50 µl of 70% ethanol, dried under vacuum and resuspended in 50 µl of TE buffer.

The efficiency of labelling was checked by the method of direct detection outlined by Boehringer Mannheim (DIG Nucleic Acid Detection Kit). Ten-fold dilutions (10^{-1} to 10^{-3}) of a 0.1 $mg.ml^{-1}$ stock of unlabelled YAP057 insert DNA were denatured as described above and 2 µl of each dilution spotted onto a Hybond-N$^+$ nylon membrane (Amersham). The DNA was fixed to the membrane by exposure to UV light (312 nm) from a Spectroline TC-312A transilluminator for 3 min. Overnight hybridisation at 42°C in the presence of 25 ng of labelled YAP057 insert DNA per ml of DIG hybridisation buffer (Appendix 1), high-stringency washing and signal detection were performed as described by Boehringer Mannheim (DIG Nucleic Acid Detection Kit). Using the labelled DNA, hybridisation to amounts of target DNA down to 20 pg was evident. This confirmed that efficient labelling of the DNA probe had occurred.

3.15.4 *In situ* hybridisation

Sections were prepared for *in situ* hybridisation using a modification of the method described by Ausubel *et al.* (1987). Unless specified, all steps were conducted at room temperature. At all stages up until detection, care was taken to ensure RNase free conditions.

All reagents were autoclaved wherever possible. Fresh hand-cut sections of stem tissue 3 cm above the basal rosette were transferred immediately to 4% (w/v) paraformaldehyde solution (Appendix 1) and fixed for 20 min at 4°C. Longer fixation times yielded poorer end-results, presumably because extended cross-linking of the tissue prevented penetration of the probe. Shorter fixation times resulted in poor preservation of the tissue. Fixation was stopped by incubation in 3x PBS (Appendix 1) for 2 min at 4°C. Sections were dehydrated by passage through a graded ethanol series. This involved transfer of the sections to aqueous solutions each containing 0.85% (w/v) NaCl and 50% ethanol, 70% ethanol and 85% ethanol, then 95% ethanol (no NaCl) and 100% ethanol. Sections were left to equilibrate in each of these ethanol solutions for 10 min. All dehydration steps were performed at 4°C. Endogenous peroxidases were inactivated by incubation for 30 min in 2% (v/v) hydrogen peroxide in methanol. Following rehydration in the same graded ethanol series outlined above (4°C), sections were incubated for 10 min in 0.2 N HCl in order to remove protein. Following two 15 min washes in 2x SSC at 70°C, and a 5 min wash in 2x SSC, specimens were re-fixed in 4% (w/v) paraformaldehyde solution (20 min). Post-fixation was stopped by transfer to 3x PBS (2 min). This was followed by a 2 min rinse in PBS. At this point, RNase controls were incubated in 100 $\mu g.ml^{-1}$ DNase-free RNase in 2x SSC for 30 min at 37°C. Subsequent to this step, extreme caution was exercised not to contaminate any of the other sections with RNase.

Sections were equilibrated for 2 min in freshly prepared 0.1 M triethanolamine buffer, pH 8.0 (Appendix 1) followed by two successive transfers to the same buffer containing 0.25% (v/v) and 0.5% (v/v) acetic anhydride respectively. Each of these steps involved incubation for 5 min. Acetylation of tissue sections prevents non-specific electrostatic binding of the probe to positively charged amino groups within tissue sections (Wilkinson,

1992).

Hybridisation, stringency washing and signal detection were performed as described by Boehringer Mannheim (DIG Nucleic Acid Detection Kit). After rinsing with 2x SSC, samples were prehybridised in DIG hybridisation buffer for 2 h at 42°C. The high content of formamide within the buffer ensured high stringency of hybridisation at temperatures not deleterious to the tissue sections. DIG hybridisation buffer containing labelled DNA probe (0.3 $\mu g.ml^{-1}$) was added to sections for overnight hybridisation at 42°C. Following post-hybridisation washing (two washes in 2x SSC, each 45 min; one 45 min wash in 1x SSC and one 30 min wash in 0.5x SSC), immunological detection of the hybridised probe using an anti-digoxigenin-peroxidase conjugate was performed as described by Boehringer Mannheim.

Sections were washed by incubation for 5 min at room temperature with 500 µl of Buffer 1 (Appendix 1). Blocking was performed by incubation of samples for 30 min in 200 µl of Buffer 2 (Appendix 1). Anti-digoxigenin antibody conjugated with horseradish peroxidase was diluted 1:20 in 100 mM Tris-HCl, 150 mM NaCl (pH 7.5) to provide a final concentration of 7.5 units Fab fragments per ml. The samples were incubated with the diluted antibody conjugate at room temperature in a moist chamber for 2 h. Unbound antibody conjugate was removed by three successive washes with 200 µl of Buffer 1. Each wash was of 10 min duration. This was followed by two washes in 200 µl of Buffer 3 (Appendix 1), each of 1 min duration. After rinsing the samples in 200 µl of dH_2O, 200 µl of peroxidase substrate (Boehringer Mannheim) was vacuum infiltrated (10^4 Pa) into the sections for 1 min. This was done in order to prevent artifacts arising from differential penetration of substrate into the tissue. Following infiltration of the peroxidase substrate, the

reaction was terminated immediately by aspirating the peroxidase substrate and adding a large excess of water. Samples were viewed immediately using an Olympus BH-2 light microscope.

In situ hybridisation to stem sections from well-watered and water-stressed plants was performed in duplicate. Relative water contents were determined for stem sections used in both experiments. The control to confirm inactivation of endogenous peroxidase activity was performed in both experiments. The RNase control hybridisation was performed once only.

4. Results

4.1 Proline accumulation in *Arabidopsis thaliana* (L.) Heynh. during hyperosmotic stress.

Exposure of mature *Arabidopsis* plants to hyperosmotic stress caused an increase in the free proline content of the leaves, stems and siliques (Figure 4.1). Plants irrigated with solutions of increasing NaCl concentration (50 mM, 100 mM and 200 mM NaCl) at 12 h intervals accumulated more proline in leaves, stems and siliques than those watered with solutions of increasing NaCl concentration at 8 h intervals. Of all the salinisation treatments, irrigation with 50 mM, 100 mM and 200 mM NaCl solutions at 16 h intervals caused the greatest accumulation of proline (Figure 4.1).

Irrigation of plants with solutions of polyethylene glycol (PEG-6000) of equivalent solute potential to the NaCl solutions at 16 h intervals indicated that this induces far greater accumulation of proline in leaves, stems and siliques than salinisation (Figure 4.1). Plants irrigated with solutions of PEG-6000 showed much more severe indications of wilting than those watered with salt solutions.

For all hyperosmotic stresses imposed, the percentage increase in proline was greater for stems and leaves than for siliques (Figure 4.1).

Determination of proline content in non-stressed *Arabidopsis* plants revealed that proline levels varied in different plant parts (Figure 4.2). Amounts of free proline per g of fresh weight were highest for flowers. Ripening siliques and ripe seeds also show high levels of free proline. The free proline content of vegetative tissues (stems, leaves and roots) is approximately an order of magnitude lower than that of reproductive tissues (siliques, flowers and seeds). The lowest proline levels were observed in roots (Figure 4.2).

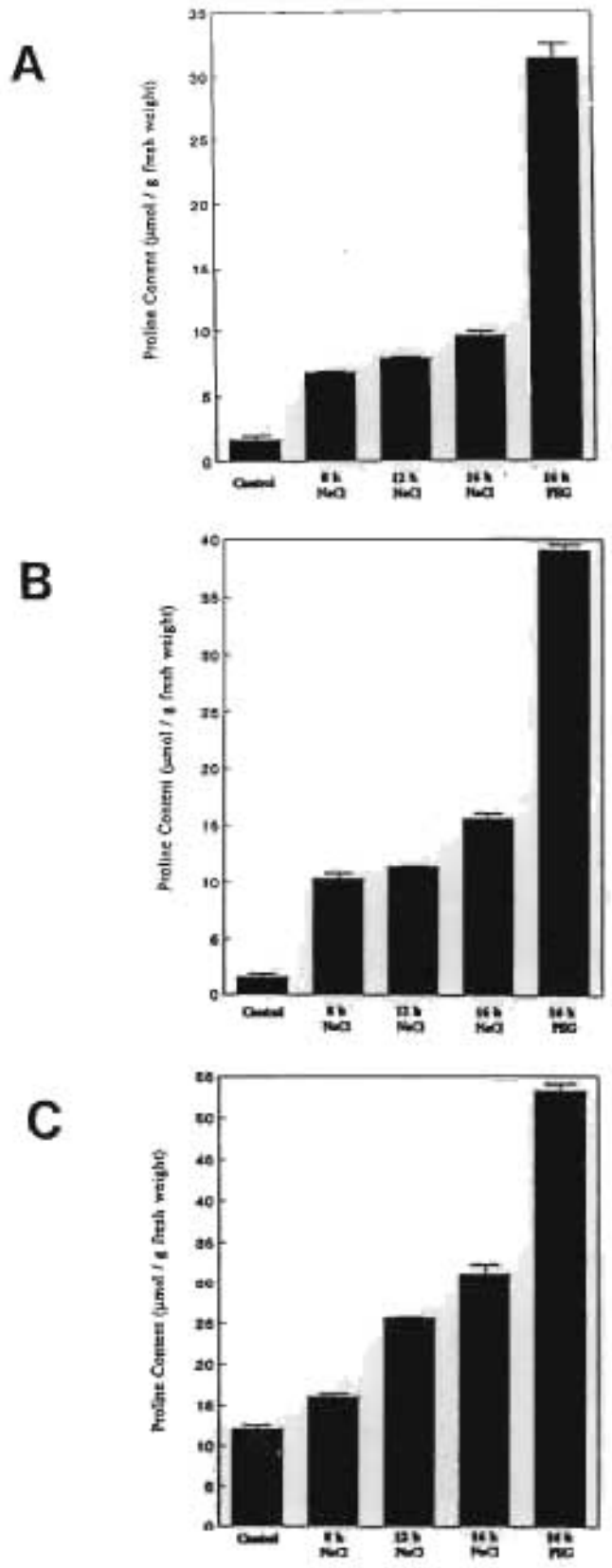


Figure 4.1 : Proline accumulation in leaves, stems and siliques of *A. thaliana* in response to hyperosmotic stress. Mature plants were irrigated with solutions of increasing NaCl concentration (50 mM, 100 mM, 200 mM) at 8 h, 12 h and 16 h intervals, and with solutions of PEG-6000 of osmotic potentials equal to the NaCl solutions at 16 h intervals. As a control, plants were irrigated at 16 h intervals with a volume of water equivalent to that used in the stress treatments. Levels of proline were determined in (A) leaves of the basal rosette, (B) stems and (C) ripening siliques. Values reported are averages of three replicates each. Vertical lines above the bars indicate the 95% confidence limit.

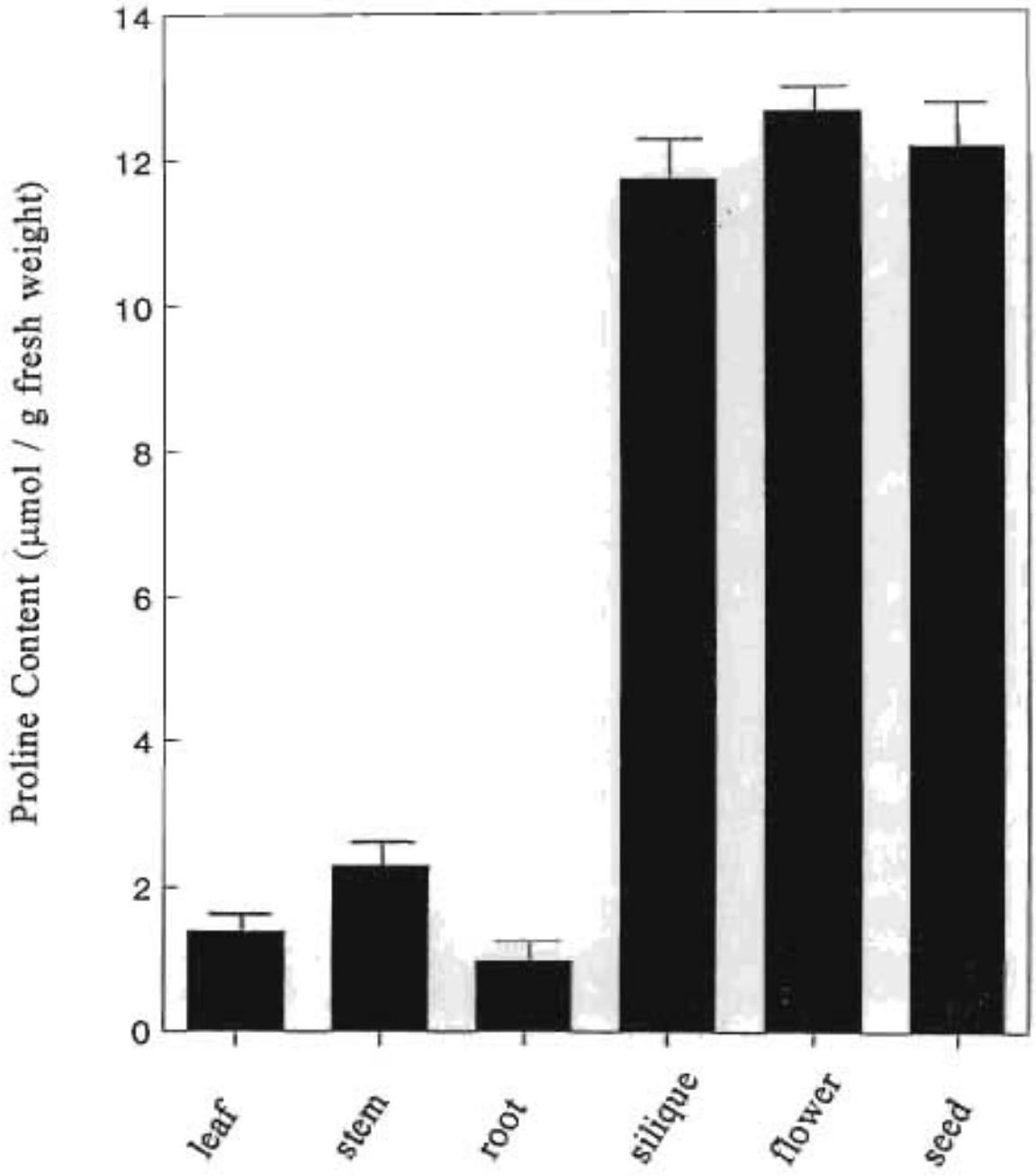


Figure 4.2 : Proline contents of different organs and seeds of *A. thaliana* under non-stressed conditions. The free proline content of leaves, stems, roots, siliques, flowers and ripe seeds were determined. Data represent three replicates. Vertical lines above the bars indicate the 95% confidence limit.

4.2 Nucleotide sequence of YAP057 and deduced amino acid sequence of the YAP057 gene product.

The strategy used to sequence the insert of YAP057 is shown in Figure 4.3. The clones used in each sequence determination have been described in Table 3.1.

Using the ligation and transformation protocols outlined in Sections 3.9 and 3.6 respectively, plating of XL1-Blue cells transformed with the products of ligation reactions routinely yielded at least 400 colonies per plate. Of these, at least 15% of the colonies were white. Plasmids isolated from selected white colonies all contained inserts of a size corresponding to that of the fragment intended to be inserted.

Overall, 87.3% of the insert within YAP057 was sequenced on both strands (Figure 4.3). The results for these sections were unequivocal and confirmatory of one another. Sequencing of YAP057 revealed that it contained an insert of size 999 bp. The nucleotide sequence of this insert is shown in Figure 4.4. Also indicated in Figure 4.4 is the translation product of the largest open reading frame (ORF) within the insert of YAP057. This is not a complete ORF as it lacks a start codon.

The deduced incomplete ORF of the YAP057 insert is terminated by the ochre stop codon TAA (Figure 4.4). A further 253 nucleotides of untranslated sequence follows on the 3' end. The cDNA is terminated by a poly(A) tract 18 nucleotides in length (Figure 4.4). Two potential polyadenylation signals AAAATA and ATAAAA (Dean *et al.*, 1986) are located at the respective positions 19 and 170 nucleotides upstream of the site of polyadenylation (Figure 4.4).

Alignment of the nucleotide sequence of YAP057 with the genomic P5CR sequence from *Arabidopsis* (Verbruggen *et al.*, 1993) revealed the sequence of YAP057 to bear complete homology to the corresponding exons of the P5CR gene (Figure 4.5).

A dot matrix plot (Boswell and Lesk, 1988) representing the alignment of the nucleotide sequence of YAP057 with the genomic clone encoding *Arabidopsis* P5CR (Verbruggen *et al.*, 1993) is shown in Figure 4.6. Regions with no homology indicated by gaps and shifts

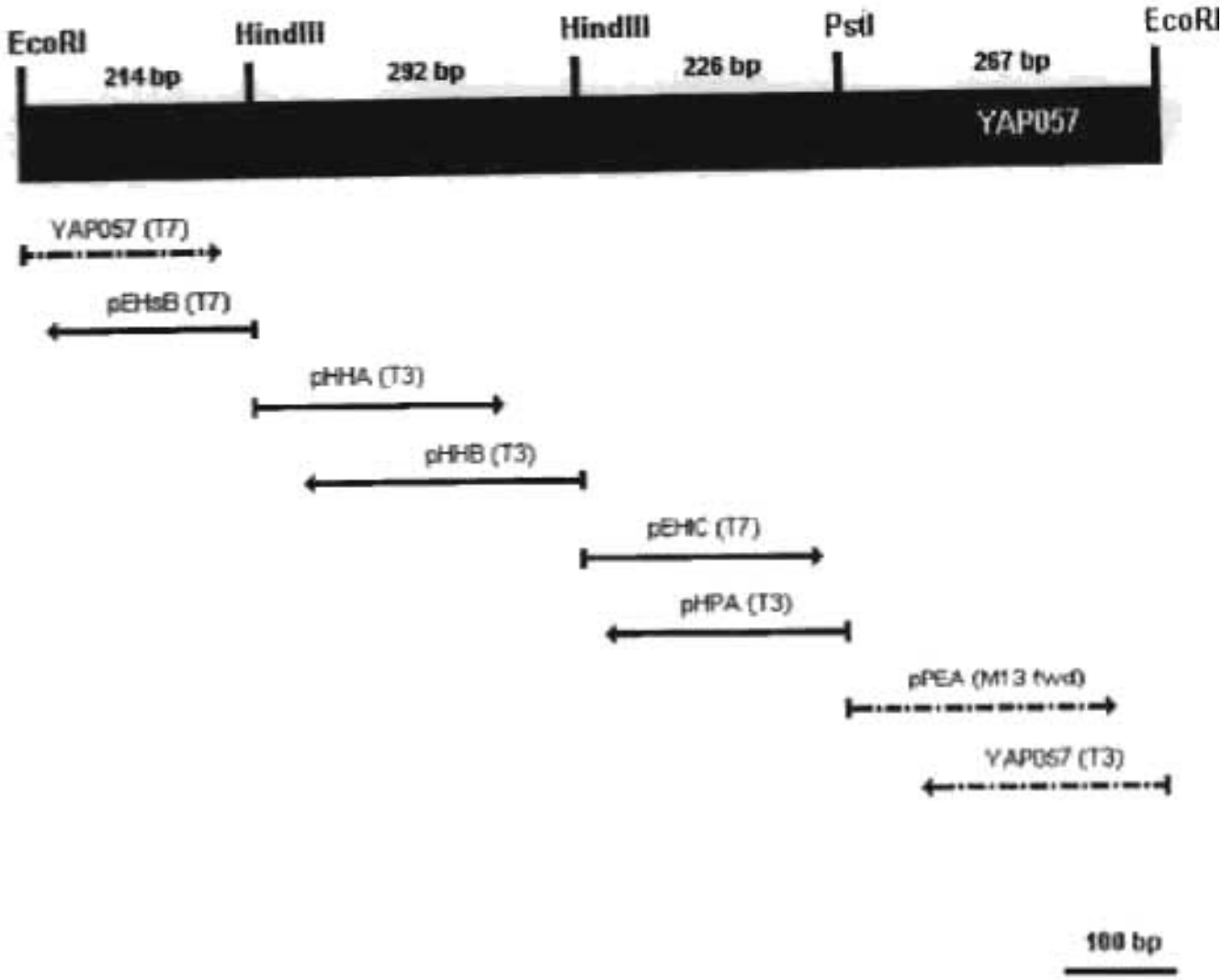


Figure 4.3 : Strategy employed for sequencing of YAP057. Arrows indicate the sequencing direction and the length of sequence obtained in individual sequencing experiments. Restriction sites used to generate the various sequence start sites are indicated. Broken lines indicate soquence data obtained using double-stranded template; solid lines indicate sequence obtained using single stranded template. Also indicated are the names of the clones used in each sequence determination (Table 3.1). The primer used in each reaction is indicated in parentheses. The scale bar represents 100 bp.

```
  1 AGAGTATAGC TAGAGGTGTG GTTGCTTCCG GTGTGCTTCC TCCTAATCGT ATATGCACCG
      S  I  A  R  G  V  V  A  S  G  V  L  P  P  N  R  I  C  T  A
 61 CCGTTCACTC AAATCTCAAT CGCCGTGATG TCTTCGAATC CTTTGGCGTC AATGTCTTCT
      V  H  S  N  L  N  R  R  D  V  F  E  S  F  G  V  N  V  F  S
121 CCACTAGCGA AGAAGTTGTT AAAGAAAGCG ATGTTGTCAT ATTCTCTGTG AAACCTCAAG
      T  S  E  E  V  V  K  E  S  D  V  V  I  F  S  V  K  P  Q  V
181 TTGTTAAGAA GGCTGTCACA GAATTAAAGT CGAAGCTTTC AAAGAATAAG ATTCTGGTTT
      V  K  K  A  V  T  E  L  K  S  K  L  S  K  N  K  I  L  V  S
241 CTGTTGCAGC TGGAATCAAG TTGAATGATT TACAGGAATG GTCTGGTCAA GATCGATTCA
      V  A  A  G  I  K  L  N  D  L  Q  E  W  S  G  Q  D  R  F  I
301 TAAGGGTGAT GCCTAATACA CCTGCCGCTG TTGGTGAGGC AGCTTCAGTT ATGAGCCTTG
      R  V  M  P  N  T  P  A  A  V  G  E  A  A  S  V  M  S  L  G
361 GCACAGGAGC AACGGAAGAG GATGGAGCAA TTGTTGCTAT GTTGTTTGGC GCGGTGGGGA
      T  G  A  T  E  E  D  G  A  I  V  A  M  L  F  G  A  V  G  K
421 AGATATTGAA AGCTGATGAG AAAATGTTTG ATGCTGTCAC TGGTCTCAGT GGAAGTGGAC
      I  L  K  A  D  E  K  M  F  D  A  V  T  G  L  S  G  S  G  P
481 CAGCATACAT ATTTTTAGCA ATTGAAGCTT TAGCCGATGG AGGAGTAGCT GCTGGTTTAC
      A  Y  I  F  L  A  I  E  A  L  A  D  G  G  V  A  A  G  L  P
541 CCCGAGAGCT TGCATTGAGT TTAGCTTCAC AGACCGTTCT TGGAGCTGCA ACGATGGTTA
      R  E  L  A  L  S  L  A  S  Q  T  V  L  G  A  A  T  M  V  S
601 GCAAAACTGG GAAGCATCCA GGTGTGTTGA AAGATGATGT TACCTCACCT GGCGGCACTA
      K  T  G  K  H  P  G  V  L  K  D  D  V  T  S  P  G  G  T  T
661 CAATAGCCGG AGTTCATGAA CTAGAGAAAG GCTCTTTCCG GGCAACACTT ATGAATGCAG
      I  A  G  V  H  E  L  E  K  G  S  F  R  A  T  L  M  N  A  V
721 TTGTTGCTGC AGCTAAACGA AGCCGCGAGC TCTCACAGAG CTAAATGATA CATATGTAGT
      V  A  A  A  K  R  S  R  E  L  S  Q  S  *
781 TGCTGCATTG TTTCACCTCA CAGATTAATC AAAATAAGGG TTATGGGCCT TATGGCATTG

841 CTTGTTTTTA GGCGAGAGTT TTATCCCACT TGTCTTCGAT GGTAGAGGTG AAGATTATTT

901 ATCTAGACTA TGATGTATTA GTTCAGACAG AACTCAGATA CTTTTCTATA ATTCTTAATC

961 TAATAAAAAT CACTTTCAGT TAAAAAAAAA AAAAAAAAA
```

Figure 4.4 : Nucleotide sequence and translation product of the YAP057 insert. The deduced amino acid sequence indicated below the nucleotide sequence represents the translation product of the largest incomplete open reading frame within the insert of YAP057.

The two *Hin*dIII sites (nucleotide positions 213 and 505) as well as the *Pst*I site (nucleotide position 727) used in subcloning of YAP057 fragments (Section 3.12.1) are underlined. Two putative polyadenylation signal sequences (Dean *et al.*, 1986) are italicised and underlined. A poly(A) tail 18 nucleotides in length is italicised.

Figure 4.5 : Alignment of the nucleotide sequence of YAP057 with the genomic sequence of *Arabidopsis* P5CR. Part of the genomic sequence of *Arabidopsis* P5CR deduced by sequencing the clone gP5CR23 (Verbruggen *et al.*, 1993) is shown in the upper line and the corresponding sequence of YAP057 below it. Introns within the genomic sequence are italicised and underlined. The 18 bp poly(A) tail at the 3' end of YAP057 was not included in the alignment.

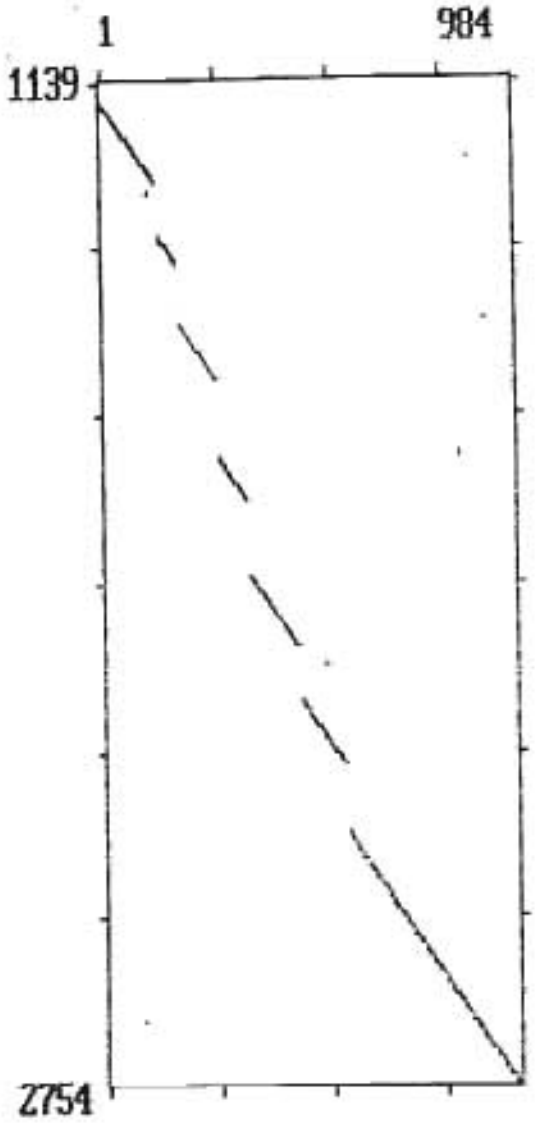


Figure 4.6 : Dot matrix plot of the YAP057 cDNA and *Arabidopsis* P5CR genomic sequence. The nucleotide sequence of YAP057 (x-axis) was aligned with the *Arabidopsis* P5CR genomic sequence (Verbruggen *et al*., 1993) indicated on the y-axis. The calculation was performed with the computer program DNASIS (Hibio), using a window of size nine nucleotides and a stringency of nine.

in the diagonal, represent the six introns present within the genomic copy of the *Arabidopsis* gene encoding P5CR.

Alignment of the deduced amino acid sequence of the YAP057 gene product with the primary structures of P5CRs from *Arabidopsis* (Verbruggen *et al.*, 1993), *Glycine max* (soybean; Delauney and Verma, 1990) and *Pisum sativum* (pea; Williamson and Slocum, 1992) revealed very high homology between the three plant P5CRs (Figure 4.7). The gene product of YAP057 is identical to the corresponding portion of the *Arabidopsis* P5CR sequence (Verbruggen *et al.*, 1993). Alignment with the two legume sequences using BLAST (Altschul *et al.*, 1990) indicated that YAP057 possesses 75.1% and 74.7% identity with the corresponding stretches of sequence of soybean and pea P5CR respectively. No gaps were introduced in alignment with soybean P5CR. Two gaps, each comprising a single nucleotide were inserted in order to facilitate alignment of the YAP057 translation product with the pea P5CR sequence (Figure 4.7). Taking conservative substitutions into account, the product encoded by the YAP057 insert possesses 85.8% and 87.7% similarity to the soybean and pea P5CRs respectively.

Comparison of the deduced amino acid sequence of the most likely translation product of YAP057 with the amino acid sequences of plant P5CRs suggests that the YAP057 cDNA is missing the 5' untranslated leader sequence and the nucleotide stretch comprising the first 23 codons of the *Arabidopsis* P5CR gene (Figure 4.7).

4.3 Nucleotide sequence of the termini of FAFJ25

In order to investigate whether FAFJ25 represented a more complete cDNA than YAP057, the 5' and 3' ends of the cDNA were sequenced using the T3 and T7 primers (Boehringer Mannheim). The sequence determined was not confirmed by sequencing the opposite strand. The sequence of the ends of FAFJ25 is shown in Figure 4.8. It has been aligned with the nucleic acid sequence of YAP057. Both of the ends of FAFJ25 sequenced exhibit complete homology to the corresponding stretches of nucleotides in YAP057 and the two cDNAs sequenced by Verbruggen *et al.* (1993). The fact that FAFJ25 was truncated even further from the 5' end than YAP057 was confirmed by restriction mapping (Figure 4.9). FAFJ25 lacks the *Hin*dIII site present 214 nucleotides from the T7 side of the insert within

```
YAP057      ----------------------SIARGVVASGVLPPNRICTAVHSNL
pcP5CR5     MEILPIPAESFKVGFIGAGKMAESIARGVVASGVLPPNRICTAVHSNL
Soybean     MEI:PIPAES+:+GFIGAGKMAESIARG:V:SGVLPP+RI:TAVH:NL
Pea         MEI-PI:++S+:+GFIGAGKMAESIA+G:::SGVLP:+RI:TAVHSN:

YAP057      NRRDVFESFGVNVFSTSEEVVKESDVVIFSVKPQVVKKAVTELKSKLSKNKIL
pcP5CR5     NRRDVFESFGVNVFSTSEEVVKESDVVIFSVKPQVVKKAVTELKSKLSKNKIL
Soybean     :RR::FESFGV:V::++++VV+ESDVV+:SVKPQ+VK::V++L:::L+K+K+L
Pea         +RR::FES:G+:V:S++++VV+-S+VV+FSVKPQ+VK::V:+LK::L+K+K+L

YAP057      VSVAAGIKLNDLQEWSGQDRFIRVMPNTPAAVGEAASVMSLGTGATEEDGAIV
pcP5CR5     VSVAAGIKLNDLQEWSGQDRFIRVMPNTPAAVGEAASVMSLGTGATEEDGAIV
Soybean     VSVAAG:KL:DLQEW+G:DRFIRVMPNTPAAVG+AASVMSLG::ATEEDG:I+
Pea         VSVAAGIKL:DLQEW+G:+RFIRVMPNTPAAVG+AASVMSLG::ATEED::++

YAP057      AMLFGAVGKILKADEKMFDAVTGLSGSGPAYIFLAIEALADGGVAAGLPRELA
pcP5CR5     AMLFGAVGKILKADEKMFDAVTGLSGSGPAYIFLAIEALADGGVAAGLPRELA
Soybean     A:LFG++GKI:KA+EK:FDA+TGLSGSGPAY++LAIEALADGGVAAGLPR+L+
Pea         +-LFG++GKI:KA+EK:FDA+TGLSGSGPAY++LAIEALADGGVAAGLPR+L+

YAP057      LSLASQTVLGAATMVSKTGKHPGVLKDDVTSPGGTTIAGVHELEKGSFRATLM
pcP5CR5     LSLASQTVLGAATMVSKTGKHPGVLKDDVTSPGGTTIAGVHELEKGSFRATLM
Soybean     LSLASQTVLGAA+MVS+TGKHPG:LKDDVTSPGGTTI:G+HELE:G:FR:TLM
Pea         LSLASQTVLGAA+M:+:+GKHPG:LKDDVTSPGGTTIAGVHELEKG:FR:TLM

YAP057      NAVVAAAKRSRELSQS
pcP5CR5     NAVVAAAKRSRELSQS
Soybean     NAVVAAAKRSRELS--
Pea         NAVVAAAKRSRELS--
```

Figure 4.7 : Alignment of the translation product of the largest incomplete open reading frame within the insert of YAP057 with the primary structures of plant P5CRs. The deduced translation product of YAP057 was aligned with translation products of the *Arabidopsis* cDNA pcP5CR5 (Verbruggen *et al.*, 1993) and cDNAs encoding P5CRs from soybean (Delauney and Verma, 1990) and pea (Williamson and Slocum, 1992). Conservative substitutions are denoted by "+". Non-conservative substitions in the two legume sequences are denoted by colons. Alignment was performed using the BLAST algorithm (Altschul *et al.*, 1990).

```
YAP057    1 AGAGTATAGC TAGAGGTGTG GTTGCTTCCG GTGTGCTTCC TCCTAATCGT ATATGCACCG

YAP057   61 CCGTTCACTC AAATCTCAAT CGCCGTGATG TCTTCGAATC CTTTGGCGTC AATGTCTTCT

YAP057  121 CCACTAGCGA AGAAGTTGTT AAAGAAAGCG ATGTTGTCAT ATTCTCTGTG AAACCTCAAG

YAP057  181 TTGTTAAGAA GGCTGTCACA GAATTAAAGT CGAAGCTTTC AAAGAATAAG ATTCTGGTTT

YAP057  241 CTGTTGCAGC TGGAATCAAG TTGAATGATT TACAGGAATG GTCTGGTCAA GATCGATTCA
PAFJ25    1 CTGTTGCAGC TGGAATCAAG TTGAATGATT TACAGGAATG GTCTGGTCAA GATCGATTCA

YAP057  301 TAAGGGTGAT GCCTAATACA CCTGCCGCTG TTGGTGAGGC AGCTTCAGTT ATGAGCCTTG
PAFJ25   61 TAAGGGTGAT GCCTAATACA CCTGCCGCTG TTGGTGAGGC AGCTTCAGTT ATGAGCCTTG

YAP057  361 GCACAGGAGC AACGGAAGAG GATGGAGCAA TTGTTGCTAT GTTGTTTGGC GCGGTGGGGA
PAFJ25  121 GCACAGGAGC AACGGAAGAG GATGGAGCAA TTGTTGCTAT GTTGTTTGGC GCGGTGGGGA

YAP057  421 AGATATTGAA AGCTGATGAG AAAATGTTTG ATGCTGTCAC TGGTCTCAGT GGAAGTGGAC
PAFJ25  181 AGATATTGAA AGCTGATGAG AAAATGTTTG ATGCTGTCAC TGGTCTCAGT GGAAGTGGAC

YAP057  481 CAGCATACAT ATTTTTAGCA ATTGAAGCTT TAGCCGATGG AGGAGTAGCT GCTGGTTTAC
PAFJ25  241 CAGCATACAT ATTTTTAGCA ATTGAAG::: :::::::::: :::::::::: ::::::::::

YAP057  541 CCCGAGAGCT TGCATTGAGT TTAGCTTCAC AGACCGTTCT TGGAGCTGCA ACGATGGTTA
PAFJ25  301 :::::::::: :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::

YAP057  601 GCAAAACTGG GAAGCATCCA GGTGTGTTGA AAGATGATGT TACCTCACCT GGCGGCACTA
PAFJ25  361 :::::::::: :::::::::: :::::::::: :::::::::: :ACCTCACCT GGCGGCACTA

YAP057  661 CAATAGCCGG AGTTCATGAA CTAGAGAAAG GCTCTTTCCG GGCAACACTT ATGAATGCAG
PAFJ25  421 CAATAGCCGG AGTTCATGAA CTAGAGAAAG GCTCTTTCCG GGCAACACTT ATGAATGCAG

YAP057  721 TTGTTGCTGC AGCTAAACGA AGCCGCGAGC TCTCACAGAG CTAAATGATA CATATGTAGT
PAFJ25  481 TTGTTGCTGC AGCTAAACGA AGCCGCGAGC TCTCACAGAG CTAAATGATA CATATGTAGT

YAP057  781 TGCTGCATTG TTTCACCTCA CAGATTAATC AAAATAAGGG TTATGGGCCT TATGGCATTG
PAFJ25  541 TGCTGCATTG TTTCACCTCA CAGATTAATC AAAATAAGGG TTATGGGCCT TATGGCATTG

YAP057  841 CTTGTTTTTA GGCGAGAGTT TTATCCCACT TGTCTTCGAT GGTAGAGGTG AAGATTATTT
PAFJ25  601 CTTGTTTTTA GGCGAGAGTT TTATCCCACT TGTCTTCGAT GGTAGAGGTG AAGATTATTT

YAP057  901 ATCTAGACTA TGATGTATTA GTTCAGACAG AACTCAGATA CTTTTCTATA ATTCTTAATC
PAFJ25  661 ATCT

YAP057  961 TAATAAAAAT CACTTTCAGT TAAAAAAAAA AAAAAAAAA
```

Figure 4.8 : Alignment of the nucleotide sequence of YAP057 with the ends of FAFJ25. The sequence of a 134 bp stretch within the centre of FAFJ25 was not determined and is indicated by colons. Positions of intron splice sites deduced by comparison with the sequence of a genomic clone encoding P5CR in *Arabidopsis* (Verbruggen *et al.*, 1993) are denoted by arrows.

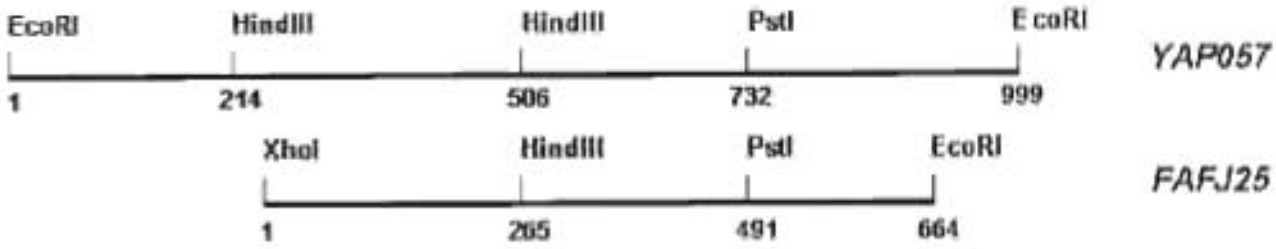


Figure 4.9 : Restriction maps of inserts within YAP057 and FAFJ25. The exact sizes of all fragments of YAP057 were deduced from the complete nucleotide sequence (Figure 4.4). The exact sizes of the fragments generated by double-digestion of FAFJ25 with *Xho*I and *Hin*dIII as well as with *Pst*I and *Eco*RI were determined by sequencing the termini of FAFJ25. Double digestion of FAFJ25 with *Hin*dIII and *Pst*I as well as with *Xho*I and *Eco*RI generated fragments of sizes estimated to be approximately 220 bp and 670 bp respectively (data not shown). Restriction sites indicated at the termini of the inserts are vector-derived cloning sites and are not found within the cDNAs encoding P5CR from *Arabidopsis*.

YAP057. Sequencing of FAFJ25 from the T3 priming site in pBluescript SK⁻ indicated that it lacks a poly(A) tail (Figure 4.8).

4.4 Polyadenylation in mRNA transcripts encoding *Arabidopsis* P5CR

Comparison of the 3' ends of three mRNA transcripts encoding *Arabidopsis* P5CR, deduced by sequencing of the corresponding cDNAs, is shown in Figure 4.10. Analysis of the 3' untranslated region of YAP057 revealed the corresponding mRNA to contain the sequence UAUUUAUC 67 nucleotides upstream of the second putative polyadenylation signal. This bears high homology to the UAUUUGUA sequence shown to possess far upstream element (FUE) activity in the cauliflower mosaic virus polyadenylation signal (Sanfacon *et al.*, 1991). Also present are two UG-rich stretches GUUUU and UUGCUUGUUUUU at the respective positions 104 and 124 nucleotides upstream of the site of polyadenylation. The UG-richness of these regions suggests that they may be subelements involved in FUE activity (Hunt, 1994).

The sequence UUUCAGUU immediately adjacent to the site of polyadenylation in YAP057 contains a YA dinucleotide (Y=pyrimidine) in a U-rich region. This is a defining feature of many plant gene polyadenylation sites (Joshi, 1987).

FAFJ25

513 UCACAGAGCUAAAUGAUACAUAUGUAGUUGCUGCUAUUGUUUCACCUCACAGAUUAAUCAAAAUA
S Q S *
577 AGGGUUAUGGGCCUUAUGGCAUUGCUUGUUUUUAGGCGAGAGUUUUAUCCCACUUGUCUUCGAUG

642 GUAGAGGUGAAGA UAUUUAUC U

YAP057

753 UCACAGAGCUAAAUGAUACAUAUGUAGUUGCUGCUAUUGUUUCACCUCACAGAUUAAUCAAAAUA
S Q S *
817 AGGGUUAUGGGCCUUAUGGCAUUGCUUGUUUUUAGGCGAGAGUUUUAUCCCACUUGUCUUCGAUG

882 GUAGAGGUGAAGA UAUUUAUC UAGACUAUGAUGUAUUAGUUCAGACAGAACUCAGAUACUUUU

946 CUAUAAUUCUUAAUCUAAUAAAAAUCAC UUUCAGUU AAAAAAAAAAAAAAAAAA

pcP5CR9 (Verbruggen *et al.*, 1993)

947 UCACAGAGCUAAAUGAUACAUAUGUAGUUGCUGCUAUUGUUUCACCUCACAGAUUAAUCAAAAUA
S Q S *
1011 AGGGUUAUGGGCCUUAUGGCAUUGCUUGUUUUUAGGCGAGAGUUUUAUCCCACUUGUCUUCGAUG

1077 GUAGAGGUGAAGA UAUUUAUC UAGACUAUGAUGUAUUAGUUCAGACAGAACUCAGAUACUUUU

1141 CUAUAAUUCUUAAUCUAAUAAAAAUCAC UUUCAGUUUU GGGUUAAAAAAAAAAAAAAAAA

gP5CR23 (Verbruggen *et al.*, 1993)

2519 TCACAGAGCTAAATGATACATATGTAGTTGCTGCTATTGTTTCACCTCACAGATTAATCAAAATA
S Q S *
2584 AGGGTTATGGGCCTTATGGCATTGCTTGTTTTTAGGCGAGAGTTTTATCCCACTTGTCTTCGATG

2649 GTAGAGGTGAAGA TATTTATC TAGACTATGATGTATTAGTTCAGACAGAACTCAGATACTTTT

2713 CTATAATTCTTAATCTAATAAAAATCAC TTTCAGTTTT GGGTTCAATATATCTGACCAATAAG

2774 CCAAGCCTCGGAGCTGATTCTGATTCTAAAAAAAATCGACTAAATTAAAGAATCTTTTACTAT

Figure 4.10 : Polyadenylation of the *Arabidopsis* mRNA transcript encoding P5CR. The 3' ends of the mRNAs corresponding to the cDNAs FAFJ25 and YAP057 are aligned with the 3' end of the mRNA corresponding to the cDNA pcP5CR9 (Verbruggen *et al.*, 1993) and the corresponding portion of the genomic clone gP5CR23 encoding P5CR (Verbruggen *et al.*, 1993). Putative polyadenylation signals (Dean *et al.*, 1986) are in bold and underlined. Poly(A) tails are italicised and underlined. Regions upstream of the second polyadenylation site with possible far upstream element (FUE) activity (Hunt, 1994) are underlined. A region immediately adjacent to the site of polyadenylation with similarity to polyadenylation sites in other plant genes (Joshi, 1987) is boxed.

4.5 Codon usage in the *Arabidopsis* gene encoding P5CR

Codon usage in the complete gene encoding P5CR in *Arabidopsis* (Verbruggen *et al.*, 1993) is shown in Table 4.1. Overall, 64% of the codons used have an A or T in the third position. Comparison of codon usage bias in the *Arabidopsis* P5CR gene with the consensus *Arabidopsis* codon usage pattern deduced by analysis of 515 *Arabidopsis* coding sequences (M. Cherry, pers. commun.*) revealed no major deviation of codon usage from the consensus pattern found in *Arabidopsis* genes characterised to date (Table 4.1).

Using the computer program CODONS (Lloyd and Sharp, 1992), a value of 55.65 was computed for the effective number of codons (Wright, 1990) used in the complete *Arabidopsis* gene encoding P5CR (Verbruggen *et al.*, 1993).

4.6 Amino acid composition of *Arabidopsis* P5CR

The amino acid composition of *Arabidopsis* P5CR (Verbruggen *et al.*, 1993) is shown in Table 4.2. *Arabidopsis* P5CR possesses more hydrophobic residues than most of the *Arabidopsis* gene products characterised to date. In particular, the amino acids valine and alanine are found at levels approximately twice those found in most *Arabidopsis* gene products characterised to date.

4.7 Homology of the gene encoding *Arabidopsis* P5CR with other genes encoding P5CR

Searches of the Swiss Prot, GenPept and Protein Information Resource (PIR) databases (National Centre for Biotechnology Information) using the BLAST network service indicated homology of the deduced amino acid sequence encoded by YAP057 to eleven other P5CRs from organisms from several biological kingdoms. These searches were conducted using the

---

* M. Cherry, Department of Molecular Biology, Massachusetts General Hospital, Harvard

Table 4.1 : Codon usage in the *Arabidopsis* gene encoding P5CR. The sequence of the complete open reading frame of the gene encoding *Arabidopsis* P5CR (Verbruggen *et al.*, 1993) was used in the analysis. Values represent the percentage contribution of each codon to the amino acid concerned. The consensus *Arabidopsis* codon usage pattern is provided in parentheses. This was determined by analysis of 515 coding sequences from *Arabidopsis* (M. Cherry, pers. commun.*).

	U		C		A		G		
U	Phe	36.4 (44.0)	Ser	16.0 (27.4)	Tyr	0.0 (41.1)	Cys	0.0 (51.5)	U
	Phe	63.6 (55.9)	Ser	12.0 (14.2)	Tyr	100.0 (58.8)	Cys	100.0 (48.4)	C
	Leu	26.1 (10.5)	Ser	24.0 (18.3)	*** (ochre)	100.0 (34.5)	*** (opal)	0.0 (45.5)	A
	Leu	21.7 (22.6)	Ser	4.0 (10.0)	*** (amber)	0.0 (19.8)	Trp	100.0 (100.0)	G
C	Leu	30.4 (26.8)	Pro	54.5 (35.8)	His	66.6 (52.1)	Arg	20.0 (19.3)	U
	Leu	13.0 (19.5)	Pro	9.1 (13.1)	His	33.3 (47.8)	Arg	20.0 (6.9)	C
	Leu	4.3 (9.6)	Pro	18.2 (34.9)	Gln	40.0 (50.1)	Arg	30.0 (9.5)	A
	Leu	4.3 (10.8)	Pro	18.2 (16.1)	Gln	60.0 (49.8)	Arg	10 (7.9)	G
A	Ile	35.7 (40.6)	Thr	28.6 (35.0)	Asn	100.0 (43.2)	Ser	20.0 (15.2)	U
	Ile	14.3 (41.2)	Thr	21.4 (25.2)	Asn	0.0 (56.7)	Ser	**24.0 (14.6)**	C
	Ile	50.0 (18.1)	Thr	35.7 (26.2)	Lys	47.4 (41.8)	Arg	10.0 (31.9)	A
	Met	100 (100.0)	Thr	14.3 (13.4)	Lys	52.6 (58.1)	Arg	10.0 (24.2)	G
G	Val	56.3 (39.5)	Ala	48.6 (46.1)	Asp	100.0 (62.3)	Gly	29.6 (35.7)	U
	Val	18.8 (22.5)	Ala	11.4 (18.8)	Asp	0.0 (37.6)	Gly	22.2 (13.4)	C
	Val	6.3 (11.4)	Ala	31.4 (22.7)	Glu	50.0 (46.1)	Gly	40.7 (38.4)	A
	Val	18.8 (26.4)	Ala	8.6 (12.2)	Glu	50.0 (53.8)	Gly	7.4 (12.3)	G

* M. Cherry, Department of Molecular Biology, Massachusetts General Hospital, Harvard

Table 4.2 : Amino acid composition of *Arabidopsis* P5CR. The complete amino acid sequence (Verbruggen *et al.*, 1993) was used in the analysis. Values represent the percentage contribution of each amino acid to the polypeptide chain. Also shown are the consensus values for the amino acid composition of *Arabidopsis* gene products.

			***Arabidopsis* Consensus [a]**	**P5CR**
Hydrophilic			**50.6**	**43.9**
	Basic		13.8	11.6
		Lys	6.4	6.9
		Arg	5.3	3.6
		His	2.1	1.1
	Acidic		12.0	10.1
		Asp	5.4	3.6
		Glu	6.6	6.5
	Neutral		24.8	22.2
		Asn	3.9	2.9
		Gln	3.5	1.8
		Cys	1.6	0.4
		Met	2.6	2.9
		Ser	7.7	9.1
		Thr	5.4	5.1
Hydrophobic			**49.0**	**56.3**
	Aliphatic		40.8	51.5
		Gly	8.0	9.8
		Ala	7.3	12.7
		Val	6.7	11.6
		Pro	4.8	4.0
		Leu	8.7	8.3
		Ile	5.3	5.1
	Aromatic		8.1	4.8
		Phe	4.1	4.0
		Tyr	2.8	0.4
		Trp	1.2	0.4

[a] Determined by analysis of 515 *Arabidopsis* coding sequences (M. Cherry, Department of Molecular Biology, Massachusetts General Hospital, Harvard)

BLASTX algorithm (Altschul *et al.*, 1990; Altschul *et al.*, 1994).

The *Arabidopsis* P5CR is homologous to P5CRs from soybean (Delauney and Verma, 1990; Swiss Prot Accession No. P17817), pea (Williamson and Slocum, 1992; Swiss Prot Accession No. Q04708), *Homo sapiens* (human; Dougherty *et al.*, 1992; Swiss Prot Accession No. P32322), *Escherichia coli* (Deutch *et al.*, 1982; Swiss Prot Accession No. P00373), *Saccharomyces cerevisiae* (Brandriss and Falvey, 1992; Swiss Prot Accession No. P32263), *Thermus thermophilus* (Hoshino *et al.*, 1994; PIR Accession No. JC2078), *Pseudomonas aeruginosa* (Savioz *et al.*, 1990; PIR Accession No. JQ0418), *Mycobacterium leprae* (author(s) unknown; GenPept Accession No. U00018), *Treponema pallidum* (F.C. Gherardini, C.R. Moomaw and P.J. Bassford, Swiss Prot Accession No. P277771) and *Methanobrevibacter smithii* (Hamilton and Reeve, 1985; GenPept Accession No. 36534). Homology was also observed between *Arabidopsis* P5CR and the translation product of an ORF from the *Bacillus subtilis* chromosome (Lewis and Wake, 1989; Ahn and Wake, 1991; Swiss Prot Accession No. P14383).

Percentage identities of the ORFs encoding all known P5CRs are presented in Table 4.3. Non-translated regions on the 5' and 3' sides of the ORFs were not considered in the alignment.

4.8 Homology of *Arabidopsis* P5CR with other P5CRs

Values for the percentage amino acid identities of all P5CRs sequenced to date are presented in Table 4.4. These were deduced using a global alignment strategy (Gotoh, 1986). These values were used to construct phenograms demonstrating the relatedness of all known P5CRs (Figure 4.11).

Clustering of values indicated P5CR primary sequence identity using both the UPGMA and complete link algorithms (Sneath and Sokal, 1973) provided very similar results (Figure 4.11). This suggests that the clusters found are rather distinct. The high cophenetic correlation values obtained for the UPGMA (r=0.97719) and complete link (r=0.93232) clustering methods confirm that both phenograms accurately reflect the data (Rohlf, 1992).

Table 4.3 : Percentage identities of the nucleic acid sequences of genes encoding P5CRs. Nucleotide sequences used in the alignment were those of genes encoding P5CRs from *A. thaliana* (Verbruggen *et al.*, 1993), soybean (Delauney and Verma, 1990), pea (Williamson and Slocum, 1992), human (Dougherty *et al.*, 1992), *S. cerevisiae* (Brandriss and Falvey, 1992), *E. coli* (Deutch *et al.*, 1982), *P. aeruginosa* (Savioz *et al.*, 1990), *M. leprae* (author(s) unknown; GenPept Accesion No. U00018), *T. pallidum* (F.C. Gherardini, C.R. Moomaw and P.J. Bassford, Swiss Prot Accession No. P27771) *T. thermophilus* (Hoshino *et al.*, 1994) and *M. smithii* (Hamilton and Reeve, 1985). Also included in the alignments was an open reading frame from *B. subtilis* with high homology to the *E. coli proC* gene (Ahn and Wake, 1991). Alignment was performed using the program ALX3 (Gotoh, 1986). Only open reading frames of the genes were used in alignments. Non-translated regions on the 5' and 3' sides of the open reading frames were not considered in alignments.

Abbreviations used at the tops of columns are : Soy, *G. max*; Pea, *P. sativum*; Hum, *H. sapiens*; Yea, *S. cerevisiae*; Eco, *E. coli*; Bac, *B. subtilis*; Pse, *P. aeruginosa*; Myc, *M. leprae*; Tre, *T. pallidum*, The, *T. thermophilus*, Met, *M. smithii*.

	Soy	Pea	Hum	Yea	Eco	Bac	Pse	Myc	Tre	The	Met
A. thaliana	71.3	73.2	55.2	51.7	56.9	53.9	50.1	54.3	50.7	51.9	52.8
G. max		82.9	57.2	51.6	54.2	54.3	54.9	51.2	49.3	48.7	48.1
P. sativum			56.3	52.9	53.6	51.9	55.6	48.9	53.7	52.7	49.6
H. sapiens				51.0	57.3	53.8	62.4	49.4	54.8	56.3	53.3
S. cerevisiae					51.9	52.3	49.8	46.3	52.7	52.8	58.5
E. coli						55.6	54.0	56.8	61.2	67.8	49.9
B. subtilis							43.9	59.3	63.4	53.1	49.4
P. aeruginosa								45.6	53.9	62.4	44.6
M. leprae									62.7	57.8	51.7
T. pallidum										63.4	59.6
T. thermophilus											54.3

Table 4.4 : Percentage identities of the amino acid sequences of P5CRs. Amino acid sequences used in the alignment were those of P5CRs from *A. thaliana* (Verbruggen *et al.*, 1993), soybean (Delauney and Verma, 1990), pea (Williamson and Slocum, 1992), human (Dougherty *et al.*, 1992), *S. cerevisiae* (Brandriss and Falvey, 1992), *E. coli* (Deutch *et al.*, 1982), *P. aeruginosa* (Savioz *et al.*, 1990), *M. leprae* (author(s) unknown; GenPept Accesion No. U00018), *T. pallidum* (F.C. Gherardini, C.R. Moomaw and P.J. Bassford, Swiss Prot Accession No. P27771), *T. thermophilus* (Hoshino *et al.*, 1994) and *M. smithii* (Hamilton and Reeve, 1985). Also included in the alignments was the translation product of an open reading frame from *B. subtilis* with high homology to the *E. coli proC* gene (Ahn and Wake, 1991). Alignment was performed using the program ALX3 (Gotoh, 1986).

Abbreviations used at the tops of columns are : Soy, *G. max*; Pea, *P. sativum*; Hum, *H. sapiens*; Yea, *S. cerevisiae*; Eco, *E. coli*; Bac, *B. subtilis*; Pse, *P. aeruginosa*; Myc, *M. leprae*; Tre, *T. pallidum*; The, *T. thermophilus*; Met, *M. smithii*.

	Soy	Pea	Hum	Yea	Eco	Bac	Pse	Myc	Tre	The	Met
A. thaliana	77.2	74.3	35.5	30.1	40.4	15.4	38.1	31.9	25.5	34.1	18.8
G. max		85.0	34.3	29.4	36.8	11.3	23.5	28.0	24.2	35.1	16.8
P. sativum			33.8	30.8	36.1	14.1	25.8	32.3	23.6	35.5	18.6
H. sapiens				26.8	36.5	13.3	37.1	30.5	26.1	34.9	18.0
S. cerevisiae					27.6	23.8	23.9	17.3	19.2	25.5	18.8
E. coli						15.4	34.7	33.1	27.7	32.4	18.0
B. subtilis							13.5	11.6	12.6	9.9	11.5
P. aeruginosa								37.4	22.6	29.5	18.6
M. leprae									22.8	31.8	16.5
T. pallidum										21.0	14.8
T. thermophilus											14.5

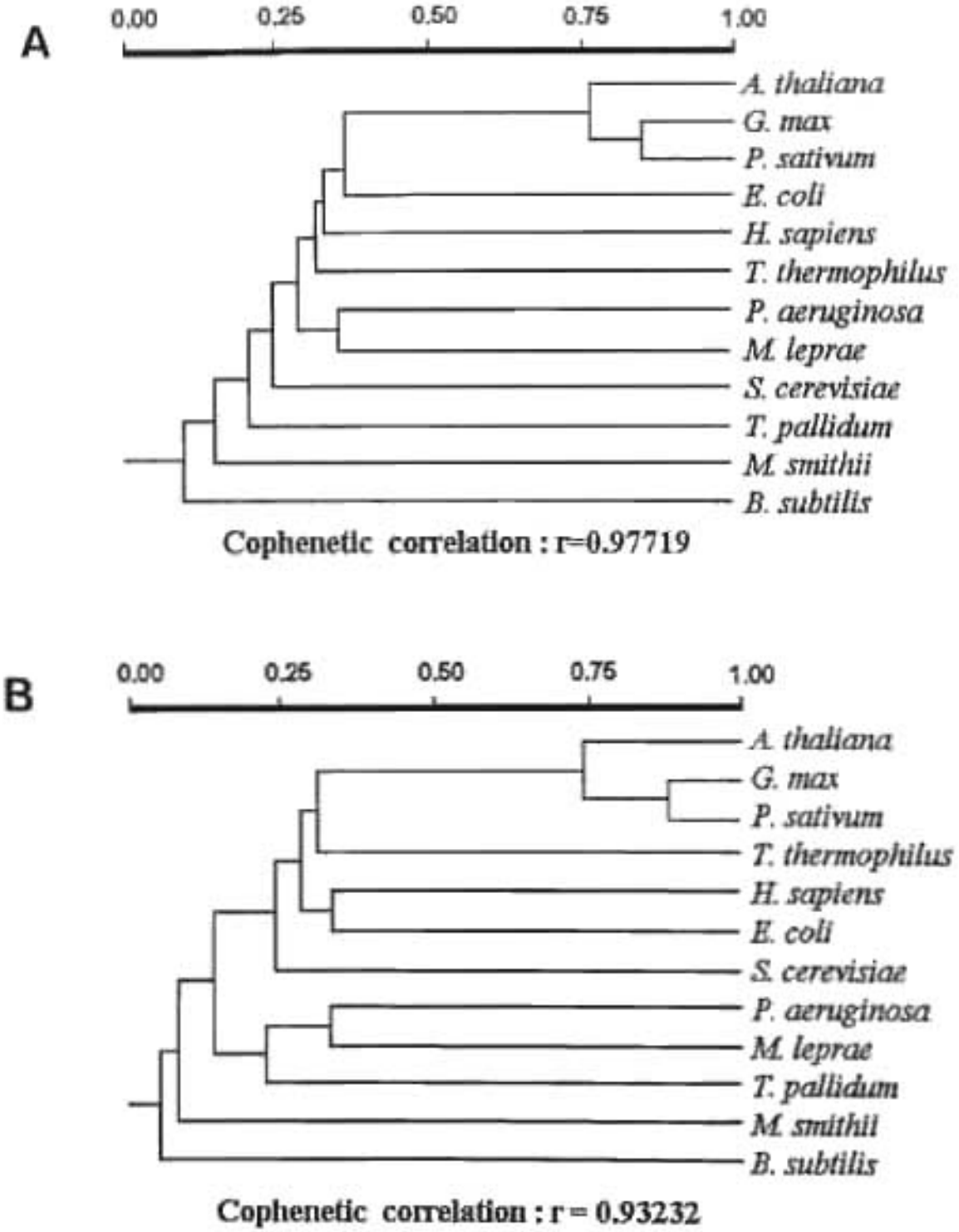


Figure 4.11 : Phenograms demonstrating relatedness of P5CRs from twelve different organisms. Cluster analysis was performed using values of the percentage identities of the twelve P5CR polypeptides to each other (Table 4.4). Clustering was performed using the (A) UPGMA and (B) complete-link clustering methods (Sneath and Sokal, 1973). Cophenetic correlation coefficients for both results are indicated.

Cluster analysis was performed using the computer program SAHN. Cophenetic value matrices were constructed using the computer program COPH. The MXCOMP program was used for comparison of the cophenetic value matrix with the matrix upon which the clustering was based. The programs SAHN, COPH and MXCOMP are part of the computer package NTSYSpc Version 1.80 (F.J. Rohlf, Exeter Software).

Analysis revealed that the P5CR sequences from soybean and pea are very similar, and that the *Arabidopsis* sequence has over 70% identity to both of these sequences. This cluster of plant P5CR sequences has greatest identity to the P5CR sequences from the Gram-negative bacteria *E. coli* and *T. thermophilus* and the human P5CR sequence. The P5CR sequences from the Gram-negative bacterium *P. aeruginosa* and Gram-positive bacterium *M. leprae* clustered together using both methods. Using both clustering methods, the P5CR sequences from *M. smithii* and *B. subtilis* are most dissimilar to all other known P5CR sequences.

Alignment of twelve P5CR sequences was performed using the computer program MACAW (Schuler *et al.*, 1991). Blocks (sets of sequences of the same length taken from several sequences and aligned without the introduction of gaps) were searched for using the BLOSUM62 scoring matrix (Henikoff and Henikoff, 1992). Blocks selected in the alignment of the P5CR sequences are shown in Figure 4.12.

Figure 4.13 shows alignment of the amino acid sequence of *Arabidopsis* P5CR with those from eleven other species. This represents the most extensive alignment of P5CR sequences presented to date. Overall, following alignment, 115 residues were found in identical positions in at least six of the twelve sequences compared. Of these conserved residues, 106 were found in the primary structure of *Arabidopsis* P5CR. Three amino acid residues were invariant throughout all of the sequences. In *Arabidopsis* P5CR, these are Asn_{128}, Gly_{241} and Thr_{243}. In *Arabidopsis* P5CR, the residues Gly_{17}, Gly_{19}, Pro_{229}, Leu_{232} and Pro_{239} are present at identical positions in eleven of the twelve sequences aligned (Figure 4.13).

Homology is distributed throughout the lengths of the P5CRs aligned. However, there are also distinct regions of high homology of the *Arabidopsis* P5CRs with other P5CRs in an N-terminal region (Gly_{14} to Glu_{23}), a central portion (Arg_{124} to Gly_{134}) including an invariant asparagine residue (Asn_{128} in *Arabidopsis* P5CR) and a C-terminal region (Pro_{229} to Gly_{246}). Within this C-terminal region, there are two invariant residues (Gly_{241} and Thr_{243} in *Arabidopsis* P5CR), a highly conserved leucine (Leu_{232} in *Arabidopsis* P5CR) and two highly conserved proline residues (Pro_{229} and Pro_{239} in *Arabidopsis* P5CR). Within this C-terminal domain, percentage sequence identity with *Arabidopsis* P5CR ranged from 94% for pea P5CR to 38% for the enzyme from *M. smithii*.

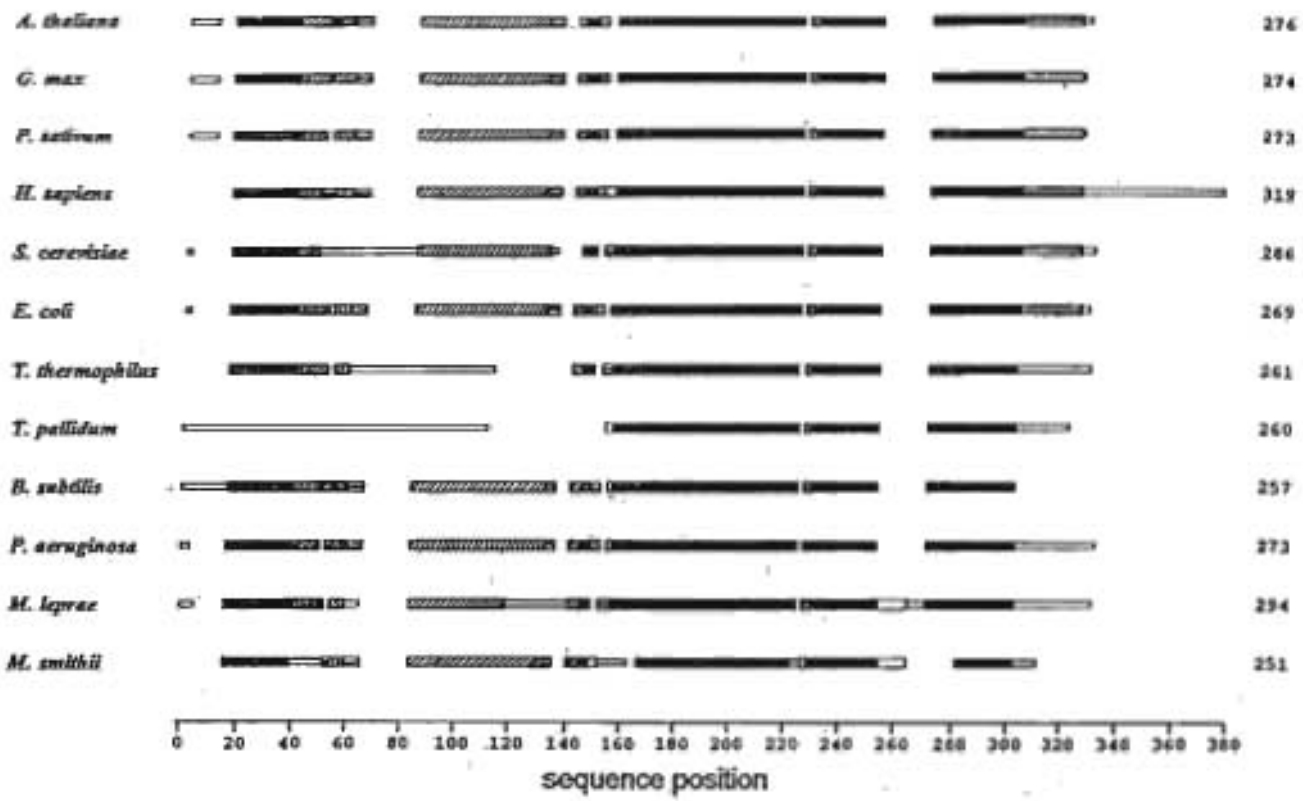


Figure 4.12 : Blocks selected in the alignment of twelve P5CRs. Sequences used in the alignment were those of P5CRs from *A. thaliana* (Verbruggen *et al.*, 1993), *G. max* (Delauney and Verma, 1990), *P. sativum* (Williamson and Slocum, 1992), *H. sapiens* (Dougherty *et al.*, 1992), *S. cerevisiae* (Brandriss and Falvey; 1992), *E. coli* (Deutch *et al.*, 1982), *T. thermophilus* (Hoshino *et al.*, 1994), *T. pallidum* (F.C. Gherardini, C.R. Moomaw and P.J. Bassford; Swiss Prot Accession No. P27771), *B. subtilis* (Ahn and Wake, 1991), *P. aeruginosa* (Savioz *et al.*, 1990), *M. leprae* (author(s) unknown; GenPept Accession No. U00018) and *M. smithii* (Hamilton and Reeve, 1985). Analysis was performed using the computer program MACAW (Schuler *et al.*, 1991). Linked blocks are represented as thickened bars. Breaks represent gaps inserted in the sequence following linkage in order to optimise alignment. Blocks are shaded on the basis of the number of residue links in which they participate.

Blocks were searched for using the segment pair overlap method with the BLOSUM62 scoring matrix (Henikoff and Henikoff, 1992). In searching for blocks, a pairwise cutoff score of 32 was used. Twelve blocks were selected on the basis of their mean pairwise scores and the number of sequences in which they participated. All blocks selected participate in at least six of the sequences. The lowest mean pairwise score (Schuler *et al.*, 1991) of all of the blocks selected is 44.9. The largest block chosen (108 residues in length) has a mean pairwise score of 165.2. Sum of the pairs scores (Schuler *et al.*, 1991) of the blocks selected range from 9088 to 789. All blocks selected represent regions of significant similarity.

Comparison of all twelve P5CR sequences with the contents of the protein domain database SBASE (Pongor *et al.*, 1994) revealed ten of the twelve P5CRs to possess N-terminal regions with homology to known NAD-binding sites (Figure 4.14). No corresponding match for an NAD(P)H binding site was found for P5CRs from *T. pallidum* or *S. cerevisiae*. However, P5CR from *S. cerevisiae* possesses a C-terminal region with homology to an FAD-binding site from *E. coli* NADH dehydrogenase (Figure 4.14). No other significant matches with domains likely to be of importance in P5CR were reported.

Comparison of the sequences of this N-terminal domain with keys for NADPH- (Hanukoglu and Gutfinger, 1989) and NADH- (Branden and Tooze, 1991) binding sites (Figure 4.15) revealed that the sequences from the plant and human N-terminal domains bear perfect homology to the key for NADPH-binding sites defined by Hanukoglu and Gutfinger (1989). In contrast, the corresponding domains of P5CRs from *S. cerevisiae*, *M. smithii*, *T. thermophilus* and *M. leprae* display greater homology to the NADH-binding site consensus defined by Branden and Tooze (1991). A single mismatch for the NADPH binding site consensus (Hanukoglu and Gutfinger, 1989) was observed in the sequence of P5CR from *P. aeruginosa* and two mismatches for the same site occur in the N-terminal stretches of P5CR sequences from *E. coli* and *B. subtilis*.

4.9 Predicted secondary structures of P5CRs from *Arabidopsis*, soybean, pea, human and *E. coli*

The predicted secondary structures of P5CRs from *Arabidopsis*, soybean, pea, human and *E. coli* are presented in Tables 4.5 to 4.9. Secondary structure was predicted using both the methods of Chou and Fasman (1978a, 1978b) and Garnier *et al.* (1978). Also shown for each region of secondary structure is the rationale underlying assignment of that structure. Numerical data used in the prediction of secondary structure using the methods of Chou and Fasman (1978a, 1978b) and Garnier *et al.* (1978) is presented in Appendix 2.

Using the predictive strategy of Garnier *et al.* (1978), 54.3% of *Arabidopsis* P5CR (Verbruggen *et al.*, 1993) was predicted to comprise α-helix, with 16.3% β-sheet structure and 8.3% occurring in coils.

A) Rat 3-hydroxyisobutyrate dehydrogenase (NAD-binding)

```
                    3 PVGFIGLGNMGNPMAKNLIKHGYP   25

P. sativum         13 +GFIG G+M++++AK            27
G. max             13 VGFIG G+M++++A+ ++  G      33
A. thaliana        13 VGFIG G+M++++A+ ++  G      33
H. sapiens          2+VGFIG G ++ ++AK            17
T. thermophilus     3 ++F+GLG+MG ++ K  + +G+     24
B. subtilis        19 V+FIG G+M++ M   +++         37
M. smithii          3 +G IG GN+G+ ++ N+I H +     24
E. coli             5 +GFIG GNMG+++   LI  G      25
P. aeruginosa       6 ++FIG GNM+ ++   L  +G P    28
```

B) *P. aeruginosa* 3-hydroxyisobutyrate dehydrogenase (NAD-binding)

```
                    2 IAFLGLGNMGGPMAANLLKAGHRVN  26

M. leprae           8 IA +G G++G ++ + LL+AG++V+   32
P. aeruginosa       6 IAF+G GNM++++ + L          18
E. coli             5 I+F+G GNMG ++ + L+ +G+     26
B. subtilis        19+AF+G G+M+  M + +++A        38
```

C) *E. coli* NADH dehydrogenase (FAD-binding)

```
                    6 GGGAGGLEMATQLGHKLGRKKK 27

S. cerevisiae     265 +G   G+E A++++  LG+KKK 286
```

D) *E. coli* NADH dehydrogenase (NAD-binding)

```
                    3 IAIVGGGATGVELSAELHNAVKQL 26

M. leprae           8 IAI+GGG+ G  L + L  A +Q+ 31
```

Figure 4.14 : Regions of homology of P5CRs with known binding sites of adenine cofactors. Homology of stretches of amino acid sequences to protein domains of known structure and/or function was assessed by searching the SBASE protein domain sequence library release 3.0 (Pongor *et al.*, 1994) via electronic mail server (sbase@icgeb.trieste.it). Sequences of P5CRs used in the search were those from *Arabidopsis* (Verbruggen *et al.*, 1993), soybean (Delauney and Verma, 1990), pea (Williamson and Slocum, 1992), human (Dougherty *et al.*, 1992), *P. aeruginosa* (Savioz *et al.*, 1990), *E. coli* (Deutch *et al.*, 1982), *T. thermophilus* (Hoshino *et al.*, 1994), *M. leprae* (author(s) unknown; GenPept Accession No. U00018); *S. cerevisiae* (Brandriss and Falvey, 1992), *M. smithii* (Hamilton and Reeve, 1985), *B. subtilis* (Ahn and Wake, 1991) and *T. pallidum* (F.C. Gherardini, C.R. Moomaw and P.J. Bassford; Swiss Prot Accession No. P27771).

Regions of eleven of the twelve P5CRs studied displayed homology to the NAD-binding sites of 3-hydroxyisobutyrate dehydrogenase from rat and *P. aeruginosa* and to the FAD- and NAD-binding sites of *E. coli* NADH dehydrogenase. Sequences accessed from SBASE are indicated in bold. The positions of amino acid residues in each sequence are indicated. Conservative substitutions in P5CR sequences are indicated by "+". Blank spaces in P5CR sequences indicate no similarity to the entry in SBASE. No match was found for an NAD(P) or FAD-binding site for P5CR from *T. pallidum*.

A) NADPH-binding site consensus
(Hanukoglu and Gutfinger, 1989)

	+XXXXGXGXXAXXXAXXXXXXGX+XXXX	
A. thaliana	KVGFIGAGKMAESIARGVVASGVLPPNR	(12-39)
G. max	TLGFIGAGKMAESIARGAVRSGVLPPSR	(12-39)
P. sativum	TLGFIGAGKMAESIAKGASRSGVLPSSR	(12-39)
H. sapiens	SVGFIGAGQLAFALAKGFTAAGVLAAHK	(2-29)
P. aeruginosa	RIAFIGAGNMAASLIGGLRAQGVPAAQI	(5-32)
B. subtilis	KVAFIGAGSMAEGMISGIVRANKIPKQN	(18-45)
E. coli	KIGFIGCGNMGKAILGGLIASGQVLPGQ	(4-31)

B) NADH-binding site consensus
(Branden and Tooze, 1991)

	XXXXXGXGXXGXXXXXXXXXXXXXX-	
S. cerevisiae	TLAILGCGVMGQALLSAIYNAPKAAD	(4-29)
M. smithii	NLGIIGYGNIGELLSQNI----ISHD	(2-23)
T. thermophilus	RLAFVGLGKMGRSILKGALERGFLRPE	(2-28)
M. leprae	RIAIIGGGSIGEALLSGLLRAGRQVKD	(12-33)

Figure 4.15 : Amino acid sequences of N-terminal NAD(P)H-binding sites in P5CRs. N-terminal sequences of P5CRs from *Arabidopsis* (Verbruggen *et al.*, 1993), soybean (Delauney and Verma, 1990), pea (Williamson and Slocum, 1992), human (Dougherty *et al.*, 1992), *P. aeruginosa* (Savioz *et al.*, 1990), *E. coli* (Deutch *et al.*, 1982), *T. thermophilus* (Hoshino *et al.*, 1994), *M. leprae* (author(s) unknown; GenPept Accession No. U00018), *S. cerevisiae* (Brandriss and Falvey, 1992) and *M. smithii* (Hamilton and Reeve, 1985) were aligned with NADPH- and NADH-binding site consensus sequences defined by Hanukoglu and Gutfinger (1989) and Branden and Tooze (1991) respectively. The N-terminal region corresponding to the 5' end of an open reading frame from *B. subtilis* (Ahn and Wake, 1991) is also included in the alignment.

Invariant residues in the consensus sequences are indicated in bold. In the consensus sequences, X represents any amino acid. Hydrophobic residues found at the same positions as hydrophobic residues in consensus sequences are underlined. Positively charged residues in consensus sequences are denoted + and negatively charged residues as -. Positive residues at the N-termini of P5CRs with NADPH binding sites are italicised. The amino acid residues in each sequence which contribute to the NAD(P)H binding site are indicated in parentheses. Using these keys, no homology to an NAD(P)H-binding site was found in the P5CR sequence from *T. pallidum* (F.C. Gherardini, C.R. Moomaw and P.J. Bassford; Swiss Prot Accession No. P27771).

Table 4.5 : Secondary structure prediction for *Arabidopsis* P5CR. The complete amino acid sequence of *Arabidopsis* P5CR (Verbruggen *et al.*, 1993) was used in the analysis.

Chou and Fasman (1978)			Garnier *et al.* (1978)		
Residues	Structural assignment	Justification	Residues	Structural assignment	Justification
1-9	coil		1-6	coil	
10-16	β-sheet	[Pb] > [Pa]	7-17	α-helix	Ph > Ps
17-19	coil		18-19	coil	
20-27	α-helix	[Pa] > [Pb]	20-27	α-helix	Ph > Ps
28-33	β-sheet	[Pb] > [Pa]	28-33	β-sheet	Ps > Ph
34-37	β-turn	[Pa]<[Pt]>[Pb]	34-36 37-39	coil turn	
38-45	β-sheet	[Pb] > [Pa]	40-45	β-sheet	Ps > Ph
46-49	β-turn	[Pa]<[Pt]<[Pb]	46-48 49-51	coil turn	
50-54	α-helix	[Pa] > [Pb]	52-55	β-sheet	Ps > Ph
55-58	β-turn	[Pa]<[Pt]<[Pb]	56-58	turn	
59-98	α-helix	[Pa] > [Pb]	59-64 65-98	coil α-helix	Ph > Ps
99-115	β-sheet	[Pb] > [Pa]	99-106 107-114	β-sheet coil	Ps > Ph
116-119	β-turn	[Pa]<[Pt]>[Pb]	115-121	turn	
120-124	β-sheet	[Pb] > [Pa]	122-125	β-sheet	Ps > Ph
125-128	β-turn	[Pa]<[Pt]>[Pb]	126-131	coil	
129-139	α-helix	[Pa] > [Pb]	132-141	α-helix	Ph > Ps
140-143	β-turn	[Pa]<[Pt]>[Pb]	142-146	coil	
144-150 151-165 166-176	α-helix β-sheet α-helix	[Pa] > [Pb] [Pb] > [Pa] [Pa] > [Pb]	147-176	α-helix	Ph > Ps
177-178 179-182	coil β-turn	[Pa]<[Pt]>[Pb]	177-183	coil	
183-192	α-helix	[Pa] > [Pb]	184-188	β-sheet	Ps > Ph
193-196	β-turn	[Pa]<[Pt]>[Pb]	189-198	α-helix	Ph > Ps
197-205	coil		199-202	coil	
206-225	α-helix	[Pa] > [Pb]	203-223	α-helix	Ph > Ps
226-229	β-turn	[Pa]<[Pt]>[Pb]	224-226 227-229	turn coil	
230-236	β-sheet	[Pb] > [Pa]	230-237	β-sheet	Ps > Ph
237-240	β-turn	[Pa]<[Pt]>[Pb]	238-242	turn	
241-245	β-sheet	[Pb] > [Pa]	243-246	β-sheet	Ps > Ph
246-272	α-helix	[Pa] > [Pb]	247-272	α-helix	Ph > Ps
273-276	β-turn	[Pa]<[Pt]>[Pb]	273-276	turn	

Table 4.6 : Secondary structure prediction for soybean P5CR. The amino acid sequence of soybean P5CR (Delauney and Verma, 1990) was used in the analysis.

Chou and Fasman (1978)			Garnier *et al.* (1978)		
Residues	Structural assignment	Justification	Residues	Structural assignment	Justification
1-9	coil		1-9 10-11	coil turn	
10-16	β-sheet	[Pb] > [Pa]	12-15	β-sheet	Pa > Pb
17-19	coil		16-19	coil	
20-27	α-helix	[Pa] > [Pb]	20-28	α-helix	Pb > Pa
28-33	β-sheet	[Pb] > [Pa]	29-34	coil	
34-37	β-turn	[Pa]<[Pt]>[Pb]	31-32 33-38	turn coil	
38-46	β-sheet	[Pb] > [Pa]	39-46	β-sheet	Pa > Pb
47-61	coil		47-55 56-58	coil turn	
62-65	β-turn	[Pa]<[Pt]>[Pb]	60-66	coil	
66-115	α-helix	[Pa] > [Pb]	67-90 91-117	β-sheet α-helix	Pa > Pb Pb > Pa
116-119	β-turn	[Pa]<[Pt]>[Pb]	118-121	turn	
120-124	β-sheet	[Pb] > [Pa]	122-125	β-sheet	Pa > Pb
125-128	β-turn	[Pa]<[Pt]>[Pb]	126-131	coil	
129-140	β-sheet	[Pb] > [Pa]	133-137 138-141	α-helix β-sheet	Pb > Pa Pa > Pb
140-144	β-turn	[Pa]<[Pt]>[Pb]	142-148	coil	
147-150	β-turn	[Pa]<[Pt]>[Pb]	149-150	turn	
150-156	β-sheet	[Pb] > [Pa]	151-157	coil	
157-160	β-turn	[Pa]<[Pt]>[Pb]	158-160	turn	
161-164	coil		161-164	coil	
165-177	α-helix	[Pa] > [Pb]	165-175	α-helix	Pb > Pa
178-181	β-turn	[Pa]<[Pt]>[Pb]	176-183	coil	
182-192 193-196	α-helix β-turn	[Pa] > [Pb] [Pa]<[Pt]>[Pb]	184-188 189-198	β-sheet α-helix	Pa > Pb Pb > Pa
197-206	β-sheet	[Pb] > [Pa]	199-205	β-sheet	Pa > Pb
204-209	α-helix	[Pa] > [Pb]	206-210	coil	
210-219 220-225	α-helix coil	[Pa] > [Pb]	211-216 217-223 224-226	β-sheet coil turn	Pa > Pb
226-229	β-turn	[Pa]<[Pt]>[Pb]	228-231	turn	
230-236	coil		232-237	β-sheet	Pa > Pb
237-240	β-turn	[Pa]<[Pt]>[Pb]	238-243	turn	
241-248	β-sheet	[Pb] > [Pa]	244-248	β-sheet	Pa > Pb
251-254	β-turn	[Pa]<[Pt]>[Pb]	251-255	turn	
258-274	α-helix	[Pa] > [Pb]	260-274	α-helix	Pb > Pa

Table 4.7 : Secondary structure prediction for pea P5CR. The amino acid sequence of pea P5CR (Williamson and Slocum, 1992) was used in the analysis.

Chou and Fasman (1978)			Garnier *et al.* (1978)		
Residues	**Structural assignment**	**Justification**	**Residues**	**Structural assignment**	**Justification**
1-6	β-sheet	[Pb] > [Pa]	1-6	β-sheet	Ps > Ph
7-10	β-turn	[Pa]<[Pt]>[Pb]	7-11	turn	
11-16	β-sheet	[Pb] > [Pa]	12-16	β-sheet	Ps > Ph
17-19	coil		17-19	coil	
20-27	α-helix	[Pa] > [Pb]	20-27	α-helix	Ph > Ps
28-31	β-turn	[Pa]<[Pt]>[Pb]	28-37	coil	
34-37	β-turn	[Pa]<[Pt]>[Pb]			
38-43	coil		38-43	β-sheet	Ps > Ph
44-47	β-turn	[Pa]<[Pt]>[Pb]	44-50	coil	
48-54	α-helix	[Pa] > [Pb]	51-61	β-sheet	Ps > Ph
55-60	β-sheet	[Pb] > [Pa]			
61-64	β-turn	[Pa]<[Pt]>[Pb]	62-65	turn	
65-79	β-sheet	[Pb] > [Pa]	66-78	β-sheet	Ps > Ph
80-116	α-helix	[Pa] > [Pb]	80-120	α-helix	Ph > Ps
124-127	β-turn	[Pa]<[Pt]>[Pb]	121-123	turn	
128-138	coil		124-130	coil	
			131-136	α-helix	Ph > Ps
139-142	β-turn	[Pa]<[Pt]>[Pb]	138-140	β-sheet	Ps > Ph
			141-144	coil	
143-155	α-helix	[Pa] > [Pb]	145-156	α-helix	Ph > Ps
156-159	β-turn	[Pa]<[Pt]>[Pb]	157-165	coil	
160-167	α-helix	[Pa] > [Pb]	166-172	turn	
168-176	β-sheet	[Pb] > [Pa]			
176-180	β-turn	[Pa]<[Pt]>[Pb]	173-182	coil	
182-191	α-helix	[Pa] > [Pb]	183-188	β-sheet	Ps > Ph
192-195	β-turn	[Pa]<[Pt]>[Pb]	189-197	α-helix	Ph > Ps
199-202	β-turn	[Pa]<[Pt]>[Pb]	198-202	β-sheet	Ps > Ph
203-215	α-helix	[Pa] > [Pb]	203-223	α-helix	Ph > Ps
224-227	β-turn	[Pa]<[Pt]>[Pb]	224-229	turn	
228-235	β-sheet	[Pb] > [Pa]	230-236	β-sheet	Ps > Ph
236-239	β-turn	[Pa]<[Pt]>[Pb]	237-241	turn	
240-247	β-sheet	[Pb] > [Pa]	242-248	β-sheet	Ps > Ph
250-253	β-turn	[Pa]<[Pt]>[Pb]	249-253	turn	
254-257	coil		254-258	coil	
258-273	α-helix	[Pa] > [Pb]	259-273	α-helix	Ph > Ps

Table 4.8 : Secondary structure prediction for human P5CR. The amino acid sequence of human P5CR (Dougherty et al., 1992) was used in the analysis.

Chou and Fasman (1978)			Garnier et al. (1978)		
Residues	Structural assignment	Justification	Residues	Structural assignment	Justification
1-6	β-sheet	[Pb] > [Pa]	1-6	β-sheet	Ps > Ph
7-31	α-helix	[Pa] > [Pb]	9-32	α-helix	Ph > Ps
32-35	β-turn	[Pa]<[Pt]>[Pb]	33-36	coil	
36-48	α-helix	[Pa] > [Pb]	37-51	α-helix	Ph > Ps
51-54	β-turn	[Pa]<[Pt]>[Pb]	52-56	coil	
55-71	α-helix	[Pa] > [Pb]	57-71	α-helix	Ph > Ps
72-79	β-sheet	[Pb] > [Pa]	74-79	β-sheet	Ps > Ph
80-112	α-helix	[Pa] > [Pb]	80-90 91-96 97-98 99-114	α-helix β-sheet turn coil	Ph > Ps Ps > Ph
113-119	β-sheet	[Pb] > [Pa]	115-116 117-120	turn β-sheet	Ps > Ph
120-123	β-turn	[Pa]<[Pt]>[Pb]	121-124	turn	
124-130	α-helix	[Pa] > [Pb]	125-136	β-sheet	Ps > Ph
135-138	β-turn	[Pa]<[Pt]>[Pb]	137-141	coil	
139-151	α-helix	[Pa] > [Pb]	142-153	α-helix	Ph > Ps
152-160	β-sheet	[Pb] > [Pa]	154-160	turn	
161-171	α-helix	[Pa] > [Pb]	161-171	α-helix	Ph > Ps
172-174	β-turn	[Pa]<[Pt]>[Pb]	172-179	coil	
178-187 188-191	α-helix β-turn	[Pa] > [Pb] [Pa]<[Pt]>[Pb]	180-191	α-helix	Ph > Ps
195-198 199-205	β-turn β-sheet	[Pa]<[Pt]>[Pb] [Pb] > [Pa]	193-205	β-sheet	Ps > Ph
206-220	α-helix	[Pa] > [Pb]	206-222	α-helix	Ph > Ps
221-224	β-turn	[Pa]<[Pt]>[Pb]	223-226	turn	
230-233	β-turn	[Pa]<[Pt]>[Pb]	227-231	turn	
234-245	coil		232-237 238-245	coil β-sheet	Ps > Ph
246-249	β-turn	[Pa]<[Pt]>[Pb]	246-250	turn	
250-255	β-sheet	[Pb] > [Pa]	251-255	β-sheet	Ps > Ph
256-291	α-helix	[Pa] > [Pb]	256-288 289-292	α-helix β-sheet	Ph > Ps
295-299	coil		293-302	coil	
300-303	β-turn	[Pa]<[Pt]>[Pb]	303-305	turn	
304-313	coil		306-311	β-sheet	Ps > Ph
314-317	β-turn	[Pa]<[Pt]>[Pb]	314-319	turn	

Table 4.9 : Secondary structure prediction for P5CR from *Escherichia coli*. The amino acid sequence of P5CR from *E. coli* (Deutch *et al.*, 1982) was used in the analysis.

Chou and Fasman (1978)			Garnier et al. (1978)		
Residues	Structural assignment	Justification	Residues	Structural asignment	Justification
4-8 9-12	β-sheet β-turn	[Pb] > [Pa] [Pa]<[Pt]>[Pb]	2-12	turn	
13-34	β-sheet	[Pb] > [Pa]	15-35	β-sheet	Ps > Ph
35-39	β-turn	[Pa]<[Pt]>[Pb]	36-40	coil	
40-49	α-helix	[Pa] > [Pb]	41-49	α-helix	Ph > Ps
51-72	α-helix	[Pa] > [Pb]	50-53 54-71	coil α-helix	Ph > Ps
75-84	α-helix	[Pa] > [Pb]	75-83	β-sheet	Ps > Ph
85-88	β-turn	[Pa]<[Pt]>[Pb]	86-90	turn	
89-106	β-sheet	[Pb] > [Pa]	91-99	β-sheet	Ps > Ph
107-111	coil		100-113	α-helix	Ph > Ps
112-117	β-sheet	[Pb] > [Pa]	114-118	β-sheet	Ps > Ph
118-121	β-turn	[Pa]<[Pt]>[Pb]	119-123	coil	
122-132	β-sheet	[Pb] > [Pa]	124-133	β-sheet	Ps > Ph
131-135 136-143	β-turn coil	[Pa]<[Pt]>[Pb]	134-141	coil	
144-155	β-sheet	[Pb] > [Pa]	142-151 152-155	α-helix turn	Ph > Ps
156-162	α-helix	[Pa] > [Pb]	156-165	α-helix	Ph > Ps
163-170	β-sheet	[Pb] > [Pa]	166-170	β-sheet	Ps > Ph
171-174	β-turn	[Pa]<[Pt]>[Pb]	171-177	coil	
175-190	α-helix	[Pa] > [Pb]	178-192	α-helix	Ph > Ps
191-194	β-turn	[Pa]<[Pt]>[Pb]	193-195	coil	
195-209 210-217 218-221	α-helix β-sheet β-turn	[Pa] > [Pb] [Pb] > [Pa] [Pa]<[Pt]>[Pb]	196-226	α-helix	Ph > Ps
225-229	β-sheet	[Pb] > [Pa]	227-230	β-sheet	Ps > Ph
230-233	β-turn	[Pa]<[Pt]>[Pb]	231-236	turn	
234-269	α-helix	[Pa] > [Pb]	237-269	α-helix	Ph > Ps

The predicted secondary structures of the five P5CR sequences studied are very similar. The propensities of different stretches of the polypeptide chains from these enzymes to form local regions of α-helix or β-sheet are represented graphically in Figure 4.16.

4.10 Hydrophobicity profiles of P5CRs from *Arabidopsis*, soybean, pea, human and *E. coli*

Comparison of the hydrophobicity profiles of P5CRs from *Arabidopsis*, soybean, pea, human and *E. coli* indicates some similarity in the tendency for certain stretches of the polypeptide chain to be concealed within the hydrophobic core of the enzyme (Figure 4.17).

4.11 Genomic DNA analysis

As shown in Figure 4.18, P5CR has a relatively simple genomic pattern. The YAP057 probe hybridised to single fragments of sizes 4 230 bp and 4 335 bp generated by digestion of *Arabidopsis* genomic DNA with *Xho*I and *Kpn*I respectively. Hybridisation of YAP057 with *Xba*I-digested genomic DNA identified two fragments of sizes 4 185 and 2 915 bp.

A strong hybridisation signal was associated with only a single 9 010 bp fragment when genomic DNA was digested with *Eco*RI. Much weaker hybridisation occurred to an *Eco*RI-generated genomic fragment of size 7 130 bp.

4.12 Tissue-specific transcription of the *Arabidopsis* gene encoding P5CR under non-stressed and water-stress conditions

At the time of sectioning of tissues used for the first *in situ* hybridisation water-stressed stem tissue had a mean RWC of 45.4% (SE = 2.59) and non-stressed tissue a mean RWC of 89.6% (SE = 1.66). The corresponding values for RWCs from water-stressed and non-stressed tissues during the replication of the experiment were 56.8 % (SE = 3.88) and 83.4% (SE = 2.79) respectively.

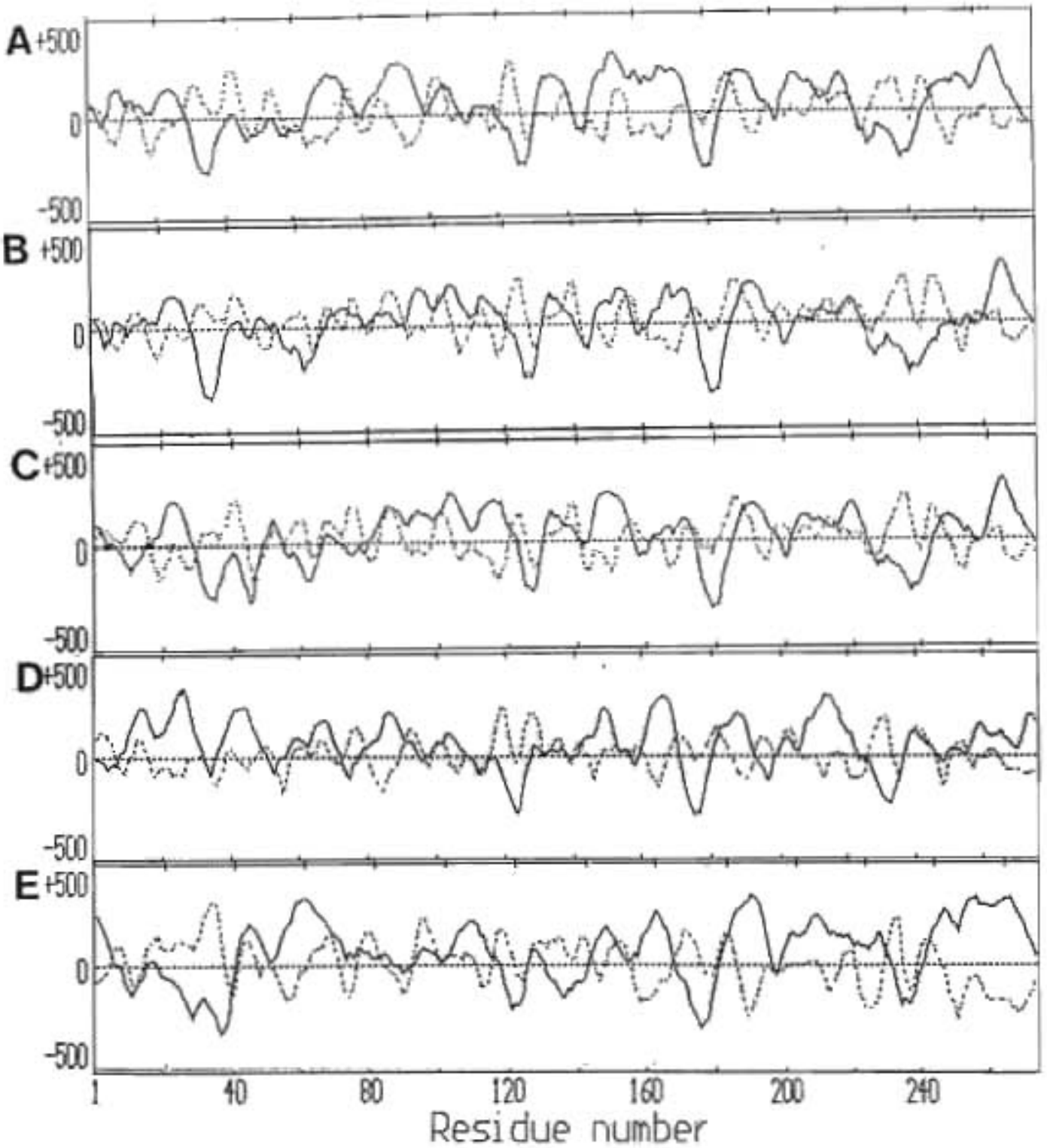


Figure 4.16 : Secondary structures of P5CRs from *Arabidopsis*, soybean, pea, human and *E. coli*. Secondary structure of P5CRs from (A) *Arabidopsis* (Verbruggen *et al.*, 1993), (B) soybean (Delauney and Verma, 1990), (C) pea (Williamson and Slocum, 1992), (D) human (Dougherty *et al.*, 1992) and (E) *E. coli* (Deutch *et al.*, 1982) was predicted according to Garnier *et al.* (1978) using the computer program PREDICT7 (Cármenes *et al.*, 1989). A window of length six residues was used in all analyses. Decision constants used for helix and sheet formation were zero. Helical and sheet probability profiles are drawn in solid and broken lines respectively. In order to facilitate alignment, the 44-residue C-terminal extension of P5CR from human with no homology to other P5CRs characterised to date was not included in the analysis. This did not affect the profile obtained in the 275 amino acid residues analysed (data not shown).

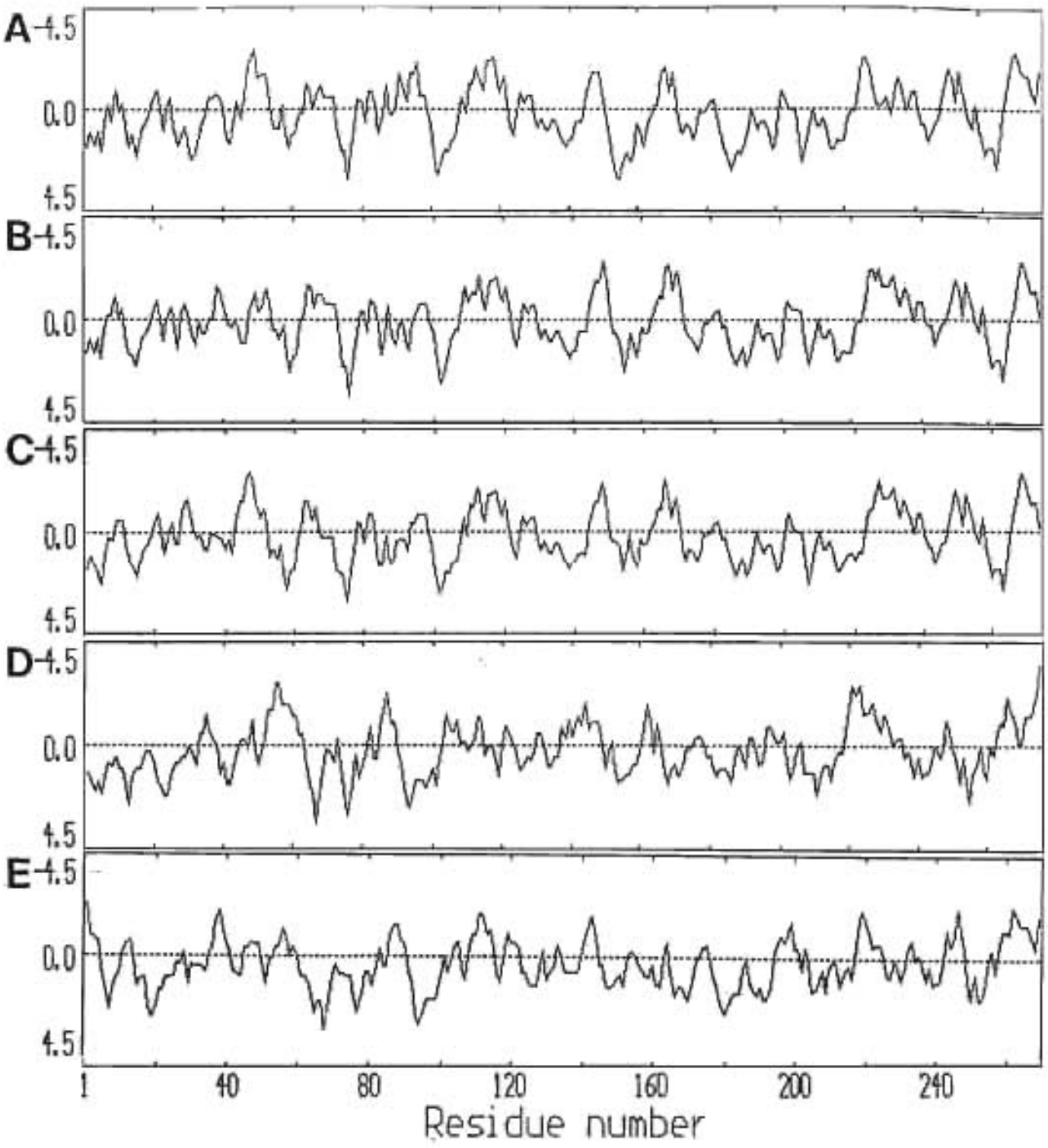


Figure 4.17 : Hydrophobicity profiles of P5CRs from *Arabidopsis*, soybean, pea, human and *E. coli*. Hydrophobicity of P5CRs from (A) *Arabidopsis* (Verbruggen *et al.*, 1993), (B) soybean (Delauney and Verma, 1990), (C) pea (Williamson and Slocum, 1992), (D) human (Dougherty *et al.*, 1992) and (E) *E. coli* (Deutch *et al.*, 1982) was calculated according to Kyte and Doolittle (1982) using the computer program PREDICT7 (Cármenes *et al.*, 1989). A window of length six residues was used in all analyses.

In order to facilitate alignment, the 44-residue C-terminal extension of P5CR from human with no homology to other P5CRs characterised to date was not included in the analysis. This did not affect the profile obtained in the 275 amino acid residues analysed (data not shown).

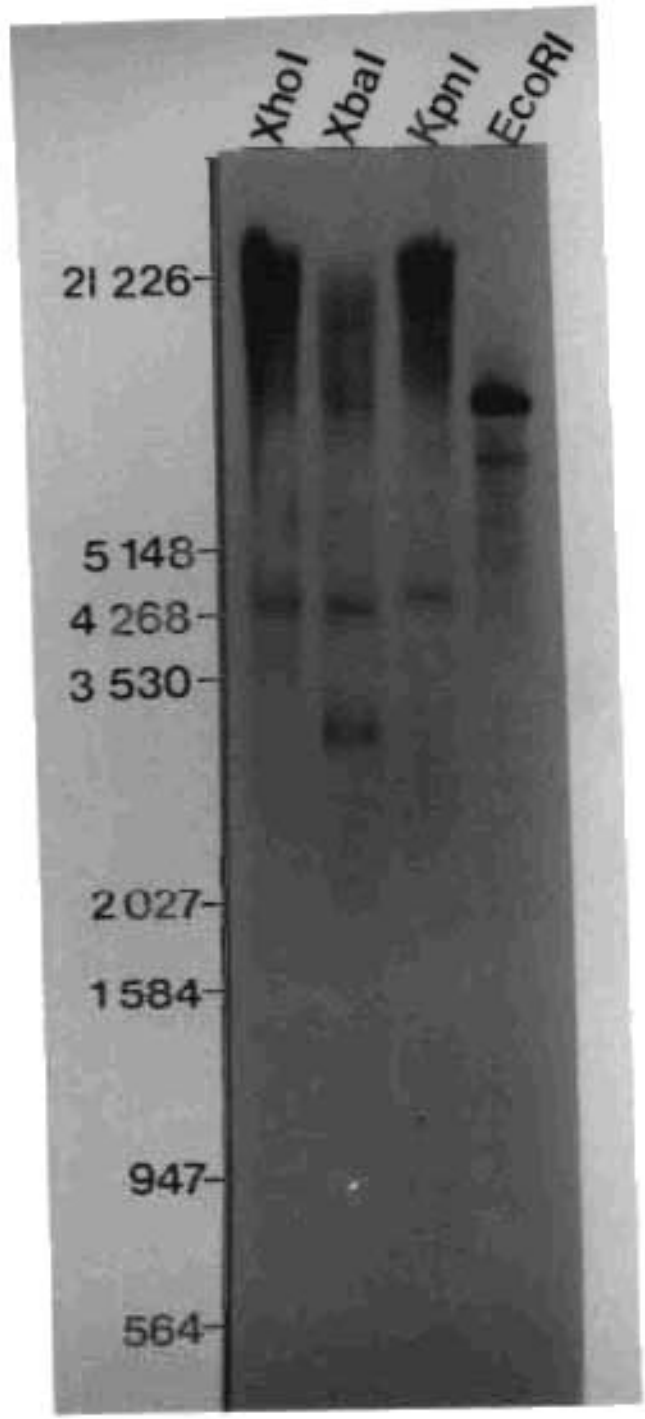


Figure 4.18 : Southern blot analysis of *Arabidopsis* genomic DNA hybridised with the YAP057 insert. Samples of genomic DNA (20 µg) were digested with *Xho*I, *Xba*I, *Kpn*I and *Eco*RI and separated by electrophoresis through a 0.8% agarose gel. The DNA was transferred onto a nylon membrane which was probed with YAP057 insert labelled using the ECL kit (Amersham). Sizes of DNA fragments in Molecular Weight Marker III (Boehringer Mannheim) are indicated in bp.

The tissue-specific differences in levels of transcript of the gene encoding P5CR in non-stressed and water-stressed stem cross-sections are shown in Figures 4.19 and 4.20. These represent results obtained in both determinations of the tissue-specific accumulation of transcript encoding P5CR in cross-sections of the *Arabidopsis* stem under well-watered and water-stressed conditions.

Under non-stressed conditions (Figure 4.19), visible presence of P5CR transcript is indicated in the cortex, vascular cambium and pith parenchyma abutting the protoxylem. A low level of transcript is present in the phloem. An overall increase in the intensity of the signal generated in these tissues is evident in water-stressed tissue (Figure 4.20). No signal was generated in the epidermal layer of stems from non-stressed plants (Fig. 4.19). However, an increase in levels of transcript encoding P5CR was evident in the epidermis of plants subjected to eight days of water deprivation (Figure 4.20).

Figure 4.21 shows that hybridisation to sections pre-treated with RNase A failed to generate any signal. This confirms that the pattern observed in the stressed and non-stressed specimens (Figures 4.19 and 4.20) represents hybridisation to mRNA transcripts.

Treatment of fresh hand-cut sections with peroxidase substrate (Boehringer Mannheim) suggested that most of the endogenous peroxidase activity was localised in the cortical parenchyma (result not shown). Incubation of sections in 2% (v/v) H_2O_2 in methanol for 30 min eliminated most of the endogenous peroxidase activity (Figure 4.22). This confirms that subsequent to their inactivation, endogenous peroxidases did not make a substantial contribution to the total signal generated.

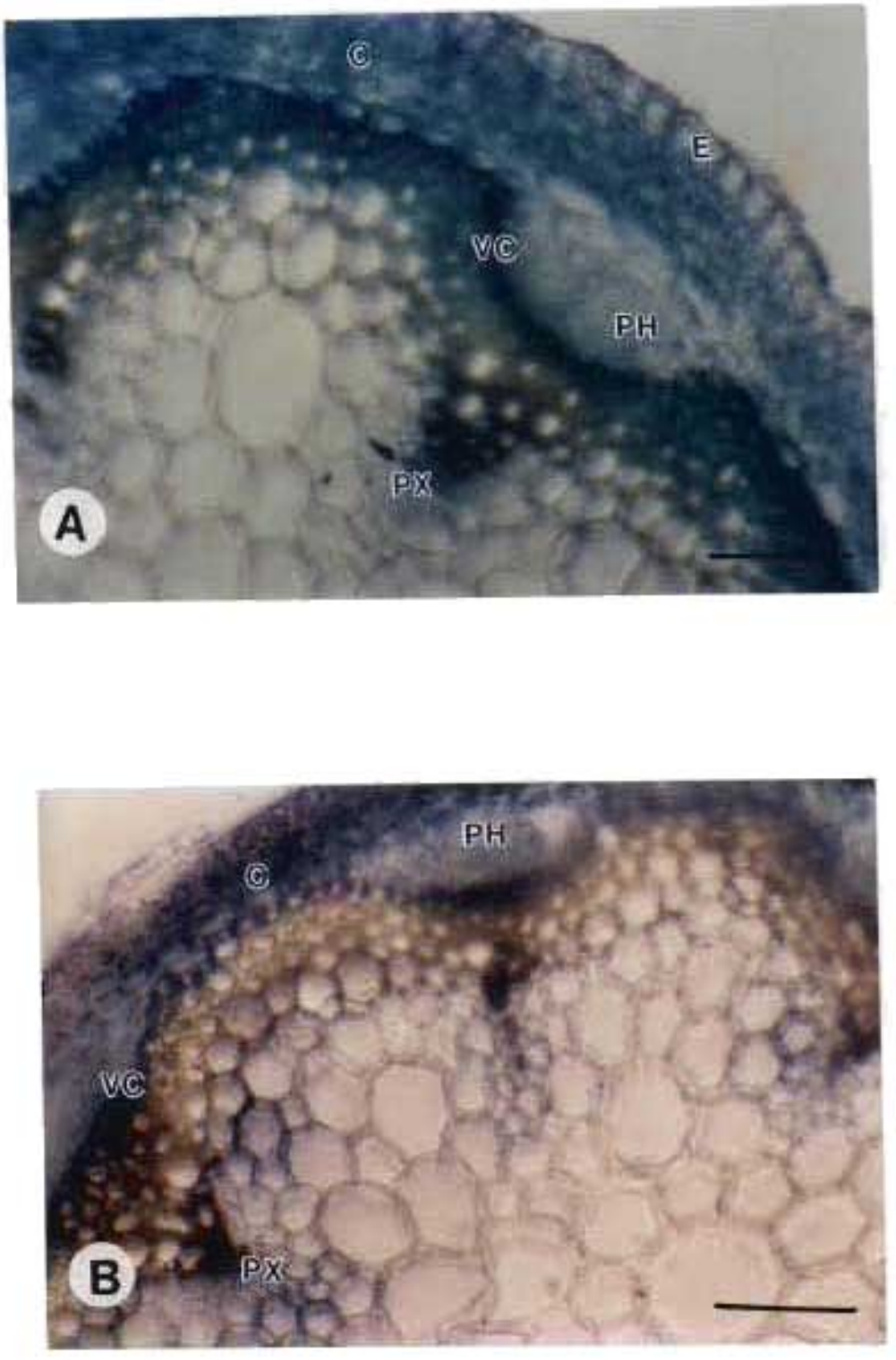


Figure 4.19 : Localisation of transcript encoding P5CR in the flowering stem of *Arabidopsis* under well-watered conditions. Following inactivation of endogenous peroxidase activity, stem sections were hybridised to double-stranded YAP057 probe. The experiment was performed in duplicate with tissue of mean RWC 89.6% (A) and 83.4% (B). E, epidermis; C, cortex; PH, phloem; VC, vascular cambium; PX, protoxylem. The scale bar represents 80 μm.

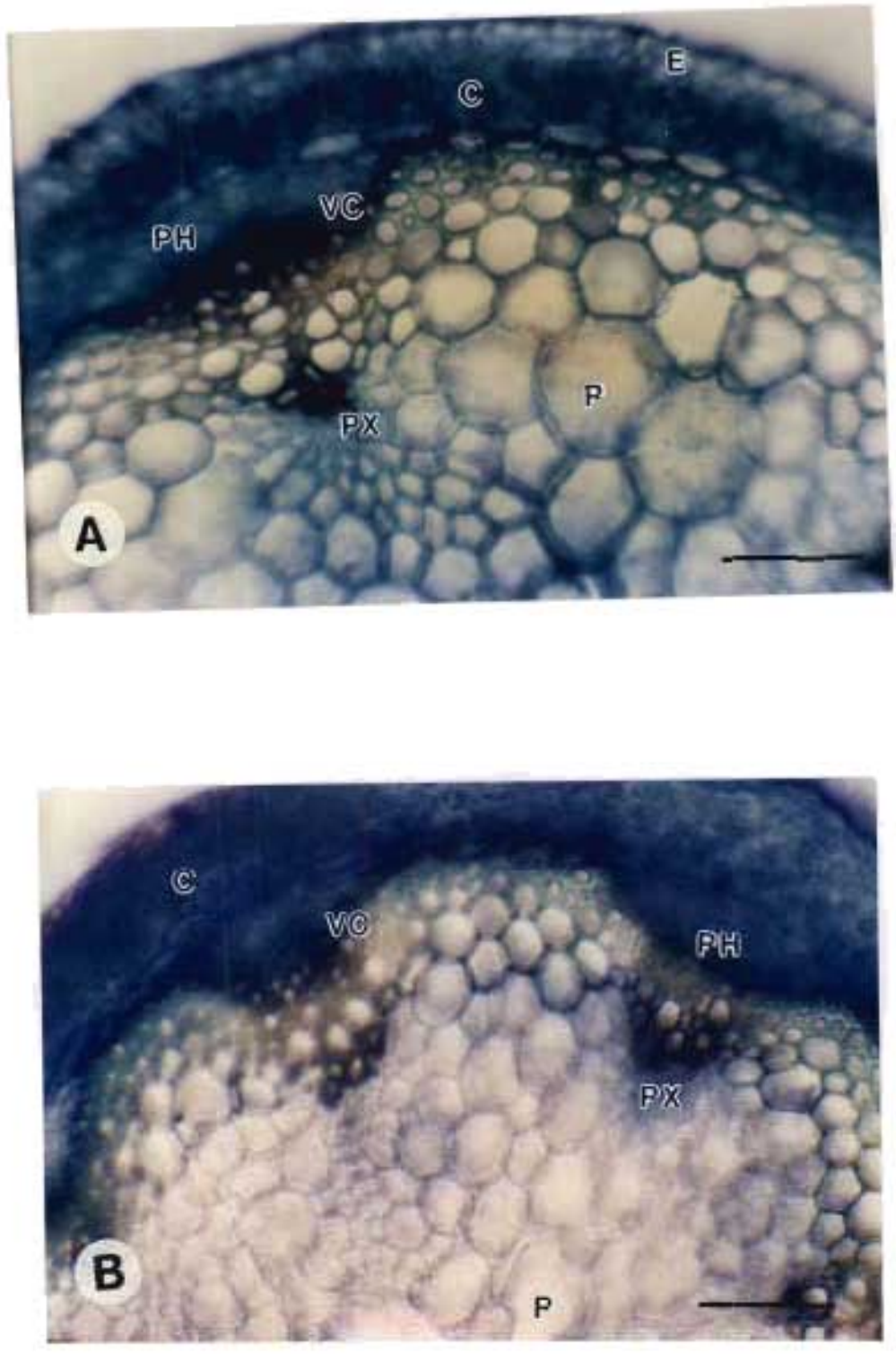


Figure 4.20 : Localisation of transcript encoding P5CR in the flowering stem of *Arabidopsis* following water deprivation . Following inactivation of endogenous peroxidase activity, stem sections were hybridised to double-stranded YAP057 probe. The experiment was performed in duplicate with tissue of mean RWC 45.4% (A) and 56.8% (B). E, epidermis; C, cortex; PH, phloem; VC, vascular cambium; PX, protoxylem; P, pith. The scale bar represents 80 μm.

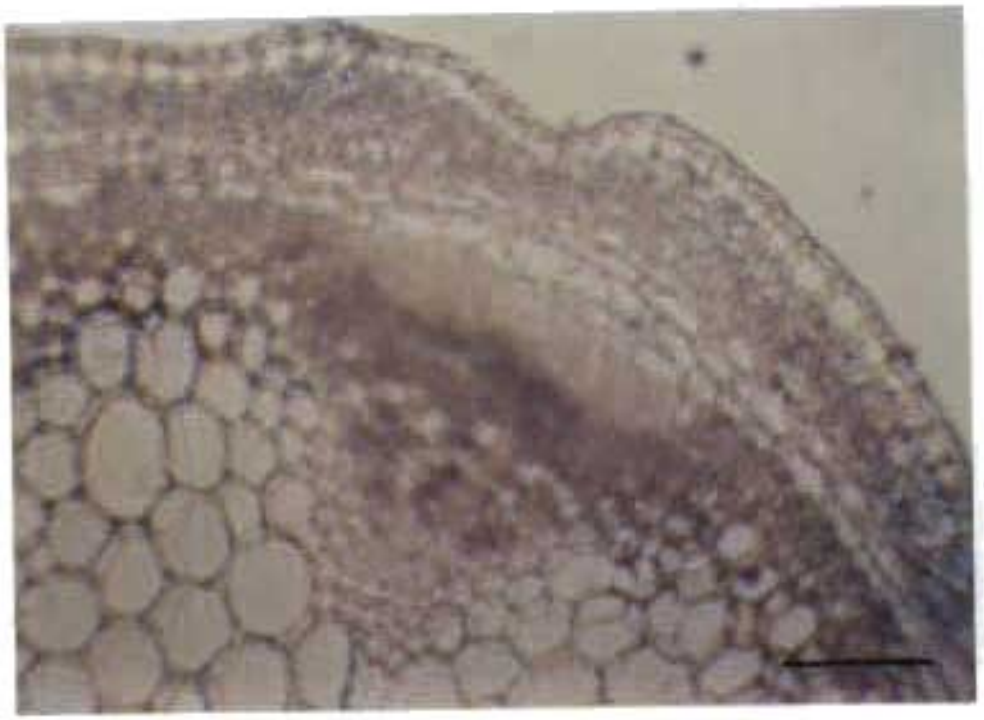


Figure 4.21 : *In situ* hybridisation to a stem section from water-stressed *Arabidopsis* (RWC = 45.4%) pre-treated with 0.1 mg.ml^{-1} RNase. The scale bar represents 80 μm.

Figure 4.22 : Non-hybridised section of a water-stressed *Arabidopsis* stem (RWC = 45.4%) incubated with peroxidase substrate (Boehringer Mannheim) following inactivation of endogenous peroxidase activity. The scale bar represents 80 μm.

5. Discussion

Free proline has been reported to accumulate in a number of plants in response to environmental stresses such as water deprivation, salinity stress, temperature extremes, heavy metal stress, anaerobiosis and pathogen infection (Section 2.1). However, controversy continues as to whether the response has any real adaptive value or is merely an incidental consequence of stress-induced changes in metabolism (Section 2.1.2.5). It is likely that use of the current techniques of molecular biology may help to resolve much of the controversy arising from purely physiological approaches to elucidate the functional significance of proline accumulation in plants exposed to environmental stress. Furthermore, it is possible that manipulation of genes encoding enzymes involved in stress-induced proline accumulation may facilitate the development of transgenic crops which exhibit enhanced environmental tolerance.

In plants, the enzyme Δ^1-pyrroline-5-carboxylate reductase (P5CR) catalyses the final step in proline biosynthesis from glutamate and ornithine, where Δ^1-pyrroline-5-carboxylate (P5C) serves as an intermediate (Section 2.2.1.2). In this study, two *Arabidopsis* cDNAs were sequenced in order to characterise the P5CR gene and the corresponding translation product. A critical piece of information lacking in previous studies of the expression of plant P5CR genes is detail of the sites of expression of the gene at the tissue-level. The tissue-specific expression of the *Arabidopsis* gene encoding P5CR was investigated as this may be of significance in attempts to increase expression of P5CR in transgenic plants.

As shown in Figure 4.1, proline accumulates in the leaves, stems and siliques of mature *Arabidopsis thaliana* (L.) Heynh. plants in response to hyperosmotic stress. Proline accumulation in other plant parts was not investigated. The increase in proline levels that accompanies imposition of hyperosmotic stress suggests that *Arabidopsis* is suitable as a model system in which to elucidate the role of proline in osmotic adjustment under conditions of water deprivation or high salinity. The more extensive accumulation of proline in response to exposure to PEG-6000 than to NaCl (Figure 4.1) may be attributed to the fact that, unlike NaCl, polyethylene glycol is a non-absorbable osmoticum (Cress and Johnson, 1987).

The observation that the rate of imposition of salt stress affected the total end amount of proline accumulated (Figure 4.1) is in keeping with the findings of Naidu *et al.* (1990). These workers demonstrated that rapid imposition of hyperosmotic stress on wheat seedlings resulted in the accumulation of glutamine, asparagine and valine in amounts equal to or greater than that of proline. In contrast, a gradual imposition of stress resulted in proline being the dominant nitrogen-containing solute accumulated (Naidu *et al.*, 1990).

In accordance with the findings of Naidu *et al.* (1990), *Arabidopsis* plants irrigated with NaCl solutions of increasing concentrations (50 mM, 100 mM and 200 mM NaCl) over 24 h showed less proline accumulation than those exposed to the same increments in salinity, but over 36 h. Even greater proline accumulation was observed over a 48 h stress period, during which plants were irrigated with NaCl solutions of the same concentrations at 16 h intervals. In the light of the findings of Naidu *et al.* (1990), it is possible that in plants irrigated with NaCl solutions at 8 h intervals, other nitrogen-containing compounds may have accumulated instead of proline.

This observation may be of considerable significance in future molecular studies of proline accumulation in plants. To date, most studies of the accumulation of mRNA transcripts encoding enzymes related to proline biosynthesis have used salt shock treatments (Delauney and Verma, 1990, Hu *et al.*, 1992; Williamson and Slocum, 1992). Gradual imposition of stress is more akin to what is experienced by plants in the field. It is therefore physiologically more meaningful than the imposition of rapid osmotic shocks. Increases in transcription of the gene encoding P5CR during salt-stress were not monitored in this study. However, it is possible that gradual imposition of stress may result in higher end-levels of transcript encoding P5CR than those observed following an abrupt osmotic shock as used by Delauney and Verma (1990) and Williamson and Slocum (1992).

These results also suggest that proline accumulation is dependent on changes in cellular metabolism other than a simple upregulation of the proline biosynthetic pathway. There are very few physiological and metabolic parameters which do not change when plants are exposed to environmental stress (Hsiao *et al.*, 1976). Of these metabolic responses to stress, probably the best characterised are those in which changes are effected in an "all-or-nothing" fashion. Metabolic switching elicited by environmental stress is normally associated with on/off effects. In this respect, proline accumulation appears to differ from many other responses to stress.

In the case of proline accumulation, such environmentally-induced switching occurs in a pathway that is already operational under normal conditions. In particular, it is likely that a number of metabolic changes need to be effected in order to meet the increased demand for energy, carbon and reducing equivalents which are needed to drive proline biosynthesis. For example, removal of 2-oxoglutarate from the tricarboxylic acid (TCA) cycle for synthesis of the proline precursor glutamate is likely to necessitate the replenishment of TCA cycle intermediates via anaplerotic reactions such as the one catalysed by phosphoenolpyruvate carboxylase (Turpin and Weger, 1990). Several workers have shown that proline accumulation is dependent on an adequate pool of carbohydrate (Stewart *et al.*, 1966; Hsiao, 1973; Singh *et al.*, 1973; Stewart, 1978). The observation that proline accumulation is dependent on the rate of imposition of stress therefore provides indirect evidence to suggest that creation of transgenic plants capable of elevated levels of proline accumulation may also need to be modified in other pathways involved in intermediary metabolism in order to support the metabolic demands of increased proline biosynthesis.

In the light of possible roles for plant growth regulators such as ABA and cytokinins in mediating proline accumulation in response to hyperosmotic stress (Section 2.1.3), it also seems possible that a gradual imposition of stress may be necessary in order to ensure modification of the overall hormonal balance within the stressed plant. Gradual alterations in the relative levels of ABA and cytokinin may be necessary in order to facilitate proline accumulation in stressed plants.

Proline levels vary between different organs of *A. thaliana* (Figure 4.2). Under normal environmental conditions, the levels of free proline in reproductive structures (siliques, flowers and seeds) are approximately ten-fold higher than those found in vegetative plant parts (leaves, stems and roots). The high levels of proline in flowers (Figure 4.2) may be associated with the often observed very high proline content of anthers and pollen (Fukusawa; 1954; Khoo and Stinson, 1957, Linskens and Schrauwen, 1969; Zhang *et al.*, 1982). More recently, Mutters *et al.* (1989) and Walton *et al.* (1991) have suggested a role for proline in flower development.

The high levels of proline in ripe seeds of *Arabidopsis* may be related to the fact that these structures need to exhibit dessication tolerance. Although a number of late embryogenesis abundant (*lea*) proteins have been identified as being important in the development of

dessication tolerance of plant embryos (Dure, 1993), it is possible that high levels of free proline within seeds may contribute to their capacity to survive the extreme water loss that characterises the dessication of seeds.

Alternatively, the high level of proline in seeds may constitute an energy reserve (Section 2.1.2.4) for use upon seed germination or prime activity of the oxidative pentose phosphate pathway (Section 2.1.2.4) during germination. Activity of the two rate-limiting enzymes of the oxidative pentose phosphate pathway, namely glucose-6-phosphate dehydrogenase and 6-phosphogluconate dehydrogenase has been shown to increase in the early stages of germination of oat seeds (Mayer and Poljakoff-Mayber, 1989). Perino and Côme (1991) have presented evidence that activity of the oxidative pentose phosphate pathway is important in germination of non-dormant apple embryos. Venekamp and Koot (1984) suggested that the high levels of free proline in the seeds of *Vicia faba* may serve a specific function in seed development. Alternatively, seed storage proteins are generally characterised by a high proline content (Mayer and Poljakoff-Mayber, 1989). High levels of free proline in seeds may contribute to the synthesis of seed storage proteins. Storage proteins are synthesised from assimilates transported into the seed from other parts of the plant (Mayer and Poljakoff-Mayber, 1989). This might account for the correspondingly high levels of proline in ripening siliques (Figure 4.2).

It is interesting to note that the mean percentage increase in the proline content of siliques with hyperosmotic stress (Figure 4.1) is much less than that observed in vegetative tissues such as stems and leaves. Following the PEG-6000 treatment, proline levels in siliques increased approximately four-fold. In leaves and stems, the same treatment caused 18-fold and 27-fold increases in proline levels respectively (Figure 4.1). Nevertheless, following any of the hyperosmotic stress treatments, free proline levels were always higher in siliques than in leaves or stems (Figure 4.1).

These findings are consistent with the observations of Venekamp and Koot (1988). These workers demonstrated that prior to the imposition of water-stress on *Vicia faba* plants, flowers and pods had levels of free proline approximately twenty-fold in excess of those found in leaves, stems and roots. However, following two days of water-deprivation, the increase in proline levels in the vegetative parts was 10- to 25-fold, but was not more than three-fold in the generative parts (Venekamp and Koot, 1988). Nevertheless, following

water-deprivation, the end-levels of free proline in flowers and pods were still higher than those found in vegetative plant parts (Venekamp and Koot, 1988).

Since the fruits of *Arabidopsis* possess constitutively higher levels of proline than leaves, stems or roots under non-stressed conditions (Figure 4.2), this suggests that there is a maximum threshold for proline accumulation needed to provide osmotic adjustment during water stress. The initial proline content of a tissue prior to the imposition of stress appears to influence the magnitude of the increase in proline content following the imposition of stress. In the leaves, stems and siliques of *Arabidopsis*, the increase in proline during hyperosmotic stress is almost inversely proportional to the initial proline content in these tissues prior to stress (Figure 4.1).

Westgate and Boyer (1985) reported that the threshold level of water deficit required to stop growth in maize varies in an organ-specific manner. It increases from stem (-0.5 MPa) to leaves (-1.0 MPa) to roots (-1.4 MPa). Differences in the capacity of these organs for osmotic adjustment have been suggested to be the reason for this variation. However, assuming that proline makes a significant contribution to osmotic adjustment in *Arabidopsis*, an analogous situation does not appear to apply in the case of this species (Figure 4.2). Both Williamson and Slocum (1992) and Verbruggen *et al.* (1993) have also reported considerably lower proline levels in roots in comparison with other organs in pea and *Arabidopsis* respectively under normal environmental conditions. Nevertheless, Venekamp and Koot (1988) demonstrated that in water-stressed *Vicia faba*, the increase in proline levels in roots exceeded that in all other plant parts. It therefore seems possible that following the imposition of hyperosmotic stress, the free proline content of roots may increase to a level comparable to that found in other tissues.

The inserts of the cDNA clones YAP057 and FAFJ25 were sequenced in order to characterise the *Arabidopsis* gene encoding P5CR and the corresponding translation product. At present, P5CR is generally accepted to catalyse the final and only committed step of proline biosynthesis in plants (Section 2.2.2). The insert within YAP057 was sequenced in its entirety, with 87.3% of the sequence being confirmed by sequencing of both the sense and antisense strands. The ends of FAFJ25 were sequenced on one strand only.

Translation of all three reading frames on both strands of YAP057 failed to reveal a complete open reading frame (ORF) capable of encoding a product of sufficient size to function as a protein. However, as shown in Figure 4.5, the largest incomplete ORF bears complete homology to the corresponding exons of a genomic clone encoding *Arabidopsis* P5CR (Verbruggen *et al.*, 1993). The ORF is incomplete as it lacks a start codon at its 5' end. The cDNA pcP5CR5 sequenced by Verbruggen *et al.* (1993) is capable of complementing an *Escherichia coli proC* mutant (Verbruggen *et al.*, 1993). The complete homology of the deduced translation product of YAP057 to the translation product of the corresponding portion of pcP5CR5 (Figure 4.7) confirms that the insert within YAP057 comprises part of a *bona fide* P5CR gene from *Arabidopsis*. This conclusion is supported by the high degree of homology of the deduced amino acid sequence encoded by YAP057 with the translation products of P5CR genes from soybean and pea (Figure 4.7). In particular, the deduced translation product of YAP057 possesses 75.1% and 74.7% identity to the corresponding regions of P5CRs from soybean and pea respectively. Taking conservative substitutions into account, the product encoded by the YAP057 insert possesses 85.8% and 87.7% similarity to the soybean and pea P5CRs respectively (Figure 4.7).

The genomic and cDNA clones sequenced by Verbruggen *et al.* (1993) were isolated from libraries constructed from the *Arabidopsis* ecotype Landsberg *erecta*. In contrast, YAP057 was isolated from a library constructed from the ecotype Columbia (Giraudat *et al.*, 1992). This difference in genetic background is not reflected in the nucleotide sequences of cDNAs encoding P5CRs from both ecotypes. Despite a number of well established differences between these two ecotypes of *Arabidopsis* (Section 2.3), the structure of the apoprotein-coding region of the gene encoding P5CR appears to have been completely conserved. However, since the cDNA YAP057 does not encode the 23 N-terminal amino acid residues of *Arabidopsis* P5CR (Figure 4.7), it is not possible to conclude on the basis of these results that the sequence of the apoprotein-coding portion of the P5CR gene has been completely conserved in both races of *Arabidopsis*.

The genes from the ecotypes Landsberg *erecta* and Columbia which encode the amino acid biosynthetic enzyme aspartate kinase-homoserine dehydrogenase have recently been sequenced by Ghislain *et al.* (1994). These workers reported 1.95% amino acid divergence between the enzymes found in these two ecotypes. This was reported to be typical of race variation in *Arabidopsis* (Ghislain *et al.*, 1994). The absolute conservation of P5CR

sequence over 253 amino acid residues (Figure 4.7) contrasts with these observations made by Ghislain *et al.* (1994). It is appealing to speculate that the importance of P5CR in environmental tolerance may have selected strongly for conservation of this sequence in *Arabidopsis*.

The complete homology between the nucleotide sequence of YAP057 and the *Arabidopsis* genomic clone encoding P5CR (Verbruggen *et al.*, 1993) is evident in the dot matrix plot shown in Figure 4.6. This indicates that the apoprotein-coding region of the *Arabidopsis* gene encoding P5CR is interrupted by six introns. Analysis of the genomic sequence of *Arabidopsis* P5CR (Verbruggen *et al.*, 1993) reveals that all six introns are flanked by the consensus sequences typical of all eukaryotic genes (Figure 4.5). These are the GT and AG dinucleotides at the 5' and 3' ends respectively (Csank *et al.*, 1990). As shown below, the nucleotides surrounding these twelve splice sites also conform fairly closely to the consensus sequences of intron splice junctions in plants (Brown, 1986).

	G	/	G	T	A	A	G	T	...	...	...	T	/	C	A	G	/	G	
%	67		100	100	67	33	33	50				83		50	100	100		50	(P5CR: Verbruggen *et al.*, 1993)
%	72		100	99	70	55	65	49				68		67	100	100		60	(consensus; Brown 1986)

The introns within the *Arabidopsis* P5CR gene range in size from 78 to 125 nucleotides in length (Figure 4.5). The exons encoding the P5CR apoprotein range in size from 49 to 201 nucleotides in length (Figure 4.5). The exons encoding the N-terminal (201 nucleotides) and C-terminal (189 nucleotides) portions of the enzyme are the largest (Figure 4.5). Although small introns are a characteristic feature of most *Arabidopsis* genes, an intron of length 2 985 nucleotides has been reported to occur within the *agamous* gene from *Arabidopsis* (Yanofsky *et al.*, 1990).

In comparison with most other *Arabidopsis* genes characterised to date, the gene encoding P5CR is fairly rich in introns. Of 41 *Arabidopsis* genes whose complete genomic sequence was known in 1989, only 24 had one or more introns (Meyerowitz, 1992). Using this randomly selected and partial set, the average number of introns per gene in the *Arabidopsis* genome is only 2.3. The mean number of introns in genes containing introns is 3.9 (Meyerowitz, 1992). Nevertheless, the presence of six introns in the gene encoding *Arabidopsis* P5CR is not an unusually large value. A gene encoding a plasma membrane proton pump ATPase in *Arabidopsis* contains fifteen introns (Pardo and Serrano, 1989).

Comparison of the nucleotide sequence of YAP057 with the two cDNAs sequenced by Verbruggen *et al.* (1993) failed to indicate any variation in the pattern of intron splicing in *Arabidopsis* P5CR transcripts. Sequence analysis of the ends of FAFJ25 also failed to implicate variation in splicing of the third, fourth and fifth introns in the *Arabidopsis* gene encoding P5CR (Figure 4.8). Dougherty *et al.* (1992) implicated alternative splicing of introns (Section 2.4.2.2) as a possible mechanism which might account for the kinetically distinguishable P5CR isozymes that have been characterised in humans (Table 2.4). Although the complete sequence homology observed in four cDNAs encoding *Arabidopsis* P5CR (Verbruggen *et al.*, 1993; this study) does not disprove the possibility of alternative intron splicing occurring during processing of *Arabidopsis* P5CR pre-mRNA transcripts, it is not consistent with the possibility of this being a regulatory mechanism capable of providing different P5CR activities in plants.

Analysis of the nucleotide composition of the genomic sequence of *Arabidopsis* P5CR (Verbruggen *et al.*, 1993) reveals that the A+T content of the intronic DNA is 67.8%, while the A+T content of the ORF is 53.5%. This correlates well with a value of 71% of A+T content for the introns and 52% A+T for the ORF of the *Arabidopsis* gene encoding alcohol dehydrogenase (Chang and Meyerowitz, 1986). As outlined in Section 2.4.2.2, the importance of localised regions of A+U richness of pre-mRNA transcripts in intron definition in plant genes has been established by Goodall and Filipowicz (1989). The ratio of A+T content of intronic DNA to the G+C content is an important determinant of splicing efficiency (Goodall and Filipowicz, 1989). The A+T content of introns with the *Arabidopsis* gene encoding P5CR (Verbruggen *et al.*, 1993) does not vary substantially between the six introns. This suggests that they are probably spliced with equal efficiency. The ratio of T to A (T:A) within the intronic regions of the *Arabidopsis* gene encoding P5CR is 1.51. This approximates the consensus values of T:A found in introns from dicotyledonous plants (T:A is 1.45) and monocotyledonous plants (T:A is 1.47) (Goodall and Filipowicz, 1989).

Although the YAP057 cDNA insert is truncated at its 5' end, it is complete at the 3' end. The ochre stop codon TAA is followed by a stretch of 235 nucleotides of 3' untranslated sequence (Figure 4.4). This untranslated stretch includes two putative polyadenylation signal sequences AAAATA and ATAAAA (Dean *et al.*, 1986) at nucleotide positions 811 and 963 respectively, and is terminated by a poly(A) tract 18 nucleotides in length (Figure 4.4). Analysis of the genomic sequence of the *Arabidopsis* P5CR gene (Verbruggen *et al*, 1993)

to the 3' side of the site of polyadenylation (Figure 4.10) failed to reveal any of the other possible polyadenylation signals defined by Dean *et al.* (1986) or Joshi (1987).

As shown in Figure 4.10, polyadenylation in YAP057 commenced 7 bp earlier than in the partial cDNA pcP5CR9 sequenced by Verbruggen *et al.* (1993). The occurrence of heterogenous 3' ends of mRNAs arising from a single transcriptional unit is not uncommon in plants and has been reported by several workers (Dean *et al.*, 1986; Hernandez-Lucas *et al.*, 1986; Sachs *et al.*, 1986; Hunt *et al.*, 1987; Joshi, 1987; Hunt, 1988; Graham *et al.*, 1989; Hunt and McDonald, 1989; Mogen *et al.*, 1990; Montoliu *et al.*, 1990; Sanfacon *et al.*, 1991). At least five different polyadenylation sites have been observed in the *Arabidopsis* LEAFY gene (Weigel *et al.*, 1992).

Alternative sites of polyadenylation in transcripts of the gene encoding P5CR from pea have been reported by Williamson and Slocum (1992). These workers reported that of three pea P5CR cDNA clones sequenced, all differed from each other slightly over a short stretch of sequence at the extreme ends of their 3' untranslated regions, only four nucleotides before the site of polyadenylation (Williamson and Slocum, 1992). However, since only one of the two cDNA clones sequenced by Verbruggen *et al.* (1993) was complete at the 3' end, sequencing of the 3' end of YAP057 (Figure 4.10) has provided the first evidence of multiple sites of poly(A) addition in pre-mRNA transcripts encoding P5CR in *Arabidopsis*.

Although there is no strictly conserved polyadenylation site consensus in plant genes, the average base composition near these sites is distinctive. The defining feature for most polyadenylation sites in plant genes is probably the occurrence of a YA dinucleotide (Y=pyrimidine) in a U-rich region (Joshi, 1987). The sequence UUUCAGUUUU immediately adjacent to the site of polyadenylation of the YAP057 transcriptional product (Figure 4.10) may serve such a role in defining at least one of the addition sites for polyadenylation of P5CR pre-mRNA transcripts in *Arabidopsis*.

It is possible that differences in polyadenylation of mRNA transcripts may serve a regulatory role by affecting their stability and/or translational efficiency (Sullivan and Green, 1993). On this basis, it seems reasonable to speculate that certain sites of polyadenylation in transcripts from the gene encoding P5CR might be used preferentially under normal conditions when proline biosynthesis is required for basic housekeeping functions (proline

is needed for protein and cell wall biosynthesis), while other sites might be preferred under conditions of environmental stress when an elevated level of proline biosynthesis is likely to be of protective value.

However, studies conducted to date have failed to implicate the differential regulation of poly(A) site selection under different growth and environmental conditions (Hunt, 1994). For example, an extensive range of experiments involving 3' end mapping of RNAs of several pea genes encoding the small subunit of ribulose-1,5-bisphosphate carboxylase (*rbcS*) revealed that no significant differences could be observed in the 3' end profiles in light- and dark-grown plants, in etiolated and mature leaves, in plants exposed to blue, red or far-red light, in leaves of different age or even in transgenic petunia or tobacco when the *rbcS* genes were placed under the control of foreign promoters (Nagy *et al.*, 1985; Fluhr and Chua, 1986; Poulsen *et al.*, 1986; Kuhlemeier *et al.*, 1987; Kuhlemeier *et al.*, 1988a; Kuhlemeier *et al.*, 1988b; Lam and Chua, 1990). Furthermore, the observation that virtually all plant genes have multiple polyadenylation sites (Hunt, 1994) makes it unlikely that this lack of fidelity in 3' end formation may be of any regulatory significance. Nevertheless the possibility that use of alternative polyadenylation sites in the P5CR transcripts of plants (Williamson and Slocum, 1992; Figure 4.10) may be of importance in controlling levels of expression of P5CR, cannot yet be excluded completely.

In the majority of plant mRNAs the distance between the polyadenylation signal and the addition site is 27 +/- 9 nucleotides (Joshi, 1987). On this basis, it appears that the second polyadenylation signal AUAAAA (nucleotide position 963 in YAP057) has been recognised in the mRNAs from which both YAP057 and pcP5CR9 originated (Figure 4.10). Verbruggen *et al.* (1993) reported that Northern analysis of total mRNA from *Arabidopsis* seeds, roots, stems, flowers and leaves revealed the presence of only a single 1.4 kb transcript with homology to a gene probe encoding P5CR. A single mRNA transcript was observed in both non-stressed and salt-stressed plantlets. The size of this transcript suggests that the first putative polyadenylation signal (nucleotide position 811 in YAP057) is never recognised (Verbruggen *et al.*, 1993).

This pattern of polyadenylation of *Arabidopsis* P5CR transcripts contrasts with the observed transcriptional pattern of the gene encoding P5CR in pea (Williamson and Slocum, 1992). These workers reported the presence of two mRNA transcripts (1.3 kb and 1.1 kb) encoding

P5CR in mature pea leaves. These sizes are consistent with recognition of both polyadenylation signals in the pea gene encoding P5CR (Williamson and Slocum, 1992). Only the 1.3 kb transcript is expressed in pea seedlings (Williamson and Slocum, 1992). This suggests that polyadenylation of pre-mRNA transcripts encoding P5CR is regulated in pea. In contrast, analysis of cDNAs encoding soybean P5CR suggested that the corresponding gene possesses only a single possible polyadenylation signal (Delauney and Verma, 1990). This is consistent with the observation of only a single P5CR mRNA transcript following Northern hybridisation to total mRNA isolated from soybean root (Delauney and Verma, 1990).

As outlined in Section 2.4.3, polyadenylation signals (near upstream elements; NUEs) are not the only sequences involved in defining the cleavage and poly(A) polymerase recognition sites within mature mRNAs. Within the region of 125 nucleotides upstream of the second polyadenylation signal in YAP057, there are at least three elements with potential far upstream element (FUE) activity (Figure 4.10). These include the sequence UAUUUAUC (nucleotides 895 to 902) and two UG-rich regions (UUGCUUGUUUUU at nucleotides 838 to 849 and GUUUU at nucleotides 858 to 862). Hunt (1994) has suggested that UG-richness may suffice for FUE function, much as AU-richness contributes to intron definition, as discussed above. The sequence UAUUUAUC bears some homology (5 matches out of 8) to the UAUUUGUA motif shown to possess FUE activity in the cauliflower mosaic virus (CaMV) polyadenylation signal (Sanfacon *et al.*, 1991).

It may be of significance to note that, with the exception of the soybean cDNA sequence (Delauney and Verma, 1990), all cDNAs encoding eukaryotic P5CRs have relatively short poly(A) tails. Poly(A) tails of up to 200 nucleotides in length have been reported for plant gene mRNA transcripts (Grierson and Covey, 1988). In *Arabidopsis*, mature mRNA transcripts encoding P5CR which have been characterised to date have poly(A) tracts of 18 bp (YAP057, this study) or 19 bp (Verbruggen *et al.*, 1993). An 18 bp poly(A) tract was reported for the transcript encoding pea P5CR (Williamson and Slocum, 1992). The poly(A) tail of the mRNA corresponding to the human P5CR cDNA sequenced (Dougherty *et al.*, 1992) is also 18 nucleotides in length. In contrast, a slightly longer poly(A) tract 42 nucleotides in length was observed in the cDNA encoding soybean P5CR (Delauney and Verma, 1990). No indication of the most likely polyadenylation pattern in *Saccharomyces cerevisiae* is available as only a genomic sequence is available (Brandriss and Falvey, 1992).

There is some evidence that the stability of a mRNA is related to the length of its poly(A) sequence. It is possible that 'older' plant mRNAs have shorter poly(A) tails (Gallie, 1993). The comparatively short poly(A) tails of transcripts encoding P5CRs therefore suggests a fairly rapid rate of turnover of mRNAs encoding P5CR *in vivo*. Furthermore, Gallie *et al.* (1989) found that the presence of a poly(A) tail on the 3' end of a plant mRNA greatly enhanced its translational efficiency. The 5' cap and poly(A) tail appear to act synergistically to increase translational efficiency (Gallie, 1991 cited by Sullivan and Green, 1993). It therefore seems feasible to suggest that the relatively short poly(A) tails found in transcripts encoding P5CRs may affect the rate at which they are translated *in vivo*.

The cDNA clone FAFJ25 is truncated at both the 5' and 3' ends (Figures 4.8 and 4.9). It is unusual for the 3' terminus of a cDNA to be missing, since synthesis of oligo-dT primed cDNAs usually starts at the poly(A) tail of the mRNA. At least two possible explanations exist for the absence of a poly(A) tail in FAFJ25. Firstly, the cDNA used to construct the library from which FAFJ25 was isolated may have been synthesised by random priming rather than by the use of an oligo-dT primer. Alternatively, partial degradation of the cDNA may have occurred prior to it being cloned into the vector. Analysis of the region adjacent to the 3' terminus of FAFJ25 (Figure 4.8) did not reveal the presence of any A-rich stretches which could have served as a priming site for an oligo-dT primer.

Analysis of codon usage in the gene encoding P5CR in *Arabidopsis* indicates a strong preference for A or T in the third codon position. Overall, 64% of the codons used have either an A or T in the third codon position. This bias is in agreement with the observed distribution of codon frequency for dicotyledonous genes (Murray *et al.*, 1989; Campbell and Gowri, 1990). It is also in keeping with the high A+T content (58.6% A+T) in the *Arabidopsis* genome (Leutwiler *et al.*, 1984). The ORF of the gene encoding *Arabidopsis* P5CR (Verbruggen *et al.*, 1993) displays a pattern of codon usage similar to that observed for the combined codon distribution pattern of 515 *Arabidopsis* coding sequences (Table 4.1).

Unusual codon usage frequency in a gene might suggest a form of translational control of the abundance of the product. In many organisms, translational rate of mRNA transcripts has been shown to be affected by whether the transcript requires tRNAs with rare or prevalent anticodons (Section 2.4.2.1). In *E. coli* (Bennetzen and Hall, 1982), *S. cerevisiae*

(Bennetzen and Hall, 1982; Ikemura, 1982) and *Dictyostelium discoideum* (Sharp and Devine, 1989), highly expressed genes have codon usage that is strongly biased towards a subset of "optimal" codons that are most efficiently and/or accurately translated by the most abundant tRNA species. Lowly expressed genes are influenced by mutational biases (Section 2.4.2.1). Similar selection for differential codon usage appears to have occurred amongst plant genes. Highly expressed genes such as those encoding ribulose-1,5-bisphosphate carboxylase and chlorophyll a/b binding protein are more restricted in their codon usage than plant genes in general (Murray *et al.*, 1989).

The high conformity of codon usage in the *Arabidopsis* gene encoding P5CR with the consensus codon usage pattern in *Arabidopsis* (Table 4.1) suggests that the gene is moderately expressed *in vivo*. A value of 55.65 for the effective number of codons (Wright, 1990; Section 2.4.2.1) used in the *Arabidopsis* gene encoding P5CR does not suggest strong codon usage bias. It is therefore unlikely that the gene encoding P5CR in *Arabidopsis* is highly expressed. This is consistent with the comparatively low level of expression of most plant genes encoding amino acid biosynthetic enzymes (Matthews *et al.*, 1988). Analysis of the coding regions of 100 dicotyledonous genes by Campbell and Gowri (1990) revealed that the mean number of codons used in genes belonging to this set was 52. This is slightly lower than the value of 55.65 determined for the effective number of codons used in the *Arabidopsis* gene encoding P5CR. However, it is important to bear in mind that pattern of codon usage within a gene is not the only factor that determines the level of output of the corresponding product (Bennetzen and Hall, 1982). The relative strength of *cis*-acting regulatory regions upstream of the gene (Section 2.4.1) is also likely to affect gene expression by altering the rate of transcription. Other alternative methods of ensuring a high output of a gene product might include measures taken to increase stability of the corresponding mRNA transcript or an increase in the number of gene copies within the genome.

Despite the relatively high similarity of the P5CRs from *Thermus thermophilus* and *Arabidopsis* (Table 4.4; Figure 4.11; Figure 4.13), codon usage in the *Arabidopsis* gene encoding P5CR contrasts starkly with that found in the homologous *T. thermophilus* gene sequenced by Hoshino *et al.* (1994). The percentage use of G or C in the third codon position in the gene encoding P5CR in *T. thermophilus* is 95.4%, whereas 64% of the codons used in the *Arabidopsis* P5CR gene have an A or T in the third codon position.

Considering the greater strength of GC triple bonds in comparison with AT double bonds in the DNA helix, this is in keeping with the codon usage bias expected of a thermophile (Kagawa *et al.*, 1984). This suggests that a number of silent nucleotide substitutions have occurred in the evolution of P5CRs. These have not altered the amino acids specified in regions critical for enzymatic activity. For example, of 91 identical amino acid residues in the *T. thermophilus*/*A. thaliana* pair of P5CRs, 72 result from silent mutations. Of these 72 conservative mutations, 61 single base substitutions (84.7%) occur in the third codon position and involve replacement of an A or T in the *Arabidopsis* gene with a G or C in the corresponding codon in *T. thermophilus*.

Analysis of the amino acid composition of *Arabidopsis* P5CR reveals it to be more hydrophobic than most *Arabidopsis* gene products currently characterised (Table 4.2). An isoelectric point (pI) of 8.64 estimated on the basis of the complete amino acid sequence of *Arabidopsis* P5CR (Verbruggen *et al.*, 1993) indicates that overall the enzyme is basic and most probably carries a net positive charge at physiological pH. This is consistent with a greater percentage of basic residues in comparison with acidic residues (Table 4.2). The estimated pI of *Arabidopsis* P5CR compares fairly favourably with a pI of 7.8 for P5CR from etiolated pea shoots (Rayapati *et al.*, 1989). This was determined using the technique of isoelectric focusing (Rayapati *et al.*, 1989).

In particular, *Arabidopsis* P5CR is rich in hydrophobic amino acid residues with aliphatic side chains (Table 4.2). Both valine and alanine are found at levels approximately twice those found in most *Arabidopsis* gene products. However, the enzyme possesses only single residues of cysteine, tryptophan and tyrosine (Figure 4.7; Table 4.2).

The complete absence of cysteine in P5CRs from soybean, pea and *T. thermophilus* and *Mycobacterium leprae* contrasts with the presence of four cysteine residues in each of the P5CRs from human, *E. coli* and *Pseudomonas aeruginosa*, seven cysteine residues in each of the P5CRs from *S. cerevisiae* and *Methanobrevibacter smithii* and eleven cysteine residues in P5CR from *Treponema pallidum* (Figure 4.13). However, the positions of the cysteine residues in P5CRs containing cysteine are not highly conserved (Figure 4.13). Therefore, assuming that the tertiary structures of P5CRs are conserved, it is unlikely that disulphide bridges contribute to maintenance of the tertiary structures of P5CRs.

Several enzymatic studies conducted on P5CRs from plants (Mazelis and Fowden, 1971; Rena and Splittstoesser, 1975; Miler and Stewart, 1976) have indicated that sulfhydryl group-blocking agents such as cysteine, sodium bisulphite, 2-mercaptoethanol and dithiothreitol inhibit P5CR activity. These reagents have a stimulatory effect on thiol enzymes by keeping their free sulfhydryl groups in the reduced form. However, their inhibitory effect on P5CR activity led to the conclusion that P5CR is not a thiol enzyme, but requires disulphide bonds to maintain the proper molecular conformation for activity (Mazelis and Fowden, 1971).

Nevertheless, analysis of the amino acid composition of the three plant P5CRs for which sequences are available (Figure 4.7) reveals that the soybean and pea P5CRs are completely deficient in cysteine and *Arabidopsis* P5CR contains only a single cysteine residue (Cys_{41}). If the native *Arabidopsis* P5CR is a homopolymer, it is unlikely that Cys_{41} participates in disulphide bonding between subunits as the multimeric structure of plant P5CRs appears to be rather unstable (Table 2.3; Section 2.2.1.2). The presence of only a single cysteine residue in *Arabidopsis* P5CR eliminates the possibility that disulphide bonding may occur within the *Arabidopsis* P5CR polypeptide. Therefore, analysis of sequence data suggests that the proposal originally put forward by Mazelis and Fowden (1971) that plant P5CRs may possess disulphide bonds is not tenable.

It is doubtful that Cys_{41} is of any functional significance in *Arabidopsis* P5CR. The absence of cysteine in soybean and pea P5CRs suggests that Cys_{41} of *Arabidopsis* P5CR is unlikely to provide an active site sulfhydryl group. The high similarity between plant P5CRs (Figure 4.7) suggests that the catalytic mechanism is highly conserved among these enzymes. Furthermore, of the three different types of cysteine found in proteins : free sulfhydryl, ligand sulfhydryl and disulphides (cystines), cysteine with a free sulfhydryl is the most uncommon (Richardson and Richardson, 1989). It is unlikely that Cys_{41} serves as a ligand to a metal or prosthetic group such as Fe, Zn, Cu or heme as dependence of P5CR activity on such ligands has never been reported. Furthermore, typically two or four cysteine residues need to be clustered together in three dimensions to bind such a group (Richardson and Richardson, 1989). It is possible that thiol reagents may target an intermediate in the reaction mechanism itself. Such an intermediate might be an activated double bond in the substrate.

Comparison of known P5CR sequences at the nucleic and deduced amino acid sequence levels reveals that except between the three plant P5CRs studied, P5CRs show greater identity at the nucleic acid level (Table 4.3) than at the amino acid level (Table 4.4). However, in discerning evolutionary relationships between organisms, protein-protein comparison is preferable to comparison of the corresponding nucleic acid sequences (Crawford, 1990). Distinct evolutionary relationships which appear to be merely incidental at the nucleotide sequence level may be meaningfully discerned at the protein sequence level.

For the purposes of phylogenetic classification, the amino acid alphabet (comprising twenty letters) is superior to the nucleic acid alphabet (comprising four letters). The possibility of an identical amino acid residue occurring by chance at an identical position in two proteins is lower than that of an identical nucleotide occurring in two gene sequences. Furthermore, natural selection is likely to operate primarily at the level of the gene product and not at the level of the gene. The degeneracy of the genetic code also suggests that molecular approaches towards establishing relatedness between organisms are better suited to protein-protein comparisons than to comparisons between nucleic acid sequences.

As outlined in Section 2.2.1, a number of inter-species complementation studies (Tomenchok and Brandriss, 1987; Delauney and Verma, 1990; Gherardini *et al.*, 1990; Dougherty *et al.*, 1992; Verbruggen *et al.*, 1993) using genes encoding P5CR have confirmed that this step in proline biosynthesis is common to all living organisms in which proline metabolism has been characterised (Adams and Frank, 1980). However, this does not necessarily imply similarity of P5CRs from different sources at the molecular level. Several studies have indicated that the homology of a plant enzyme to the corresponding bacterial enzyme has no bearing on the ability of the plant enzyme to assemble and function in a bacterium. For example, despite extremely limited homology between plant glutamine synthetase and *E. coli* glutamine synthetase (Tingey *et al.*, 1988), cDNAs encoding plant glutamine synthetases can complement an *E. coli* glutamine auxotroph (DasSarma *et al.*, 1986; Snustad *et al.*, 1988). Similarly, although bacterial acetolactate synthase (an enzyme involved in biosynthesis of leucine, isoleucine and valine) is a heterodimer (Umbarger, 1987), a single cDNA encoding the corresponding plant enzyme is capable of complementing an *E. coli* mutant deficient in the large subunit of acetolactate synthase (Smith *et al.*, 1989). Ability of a plant cDNA to restore prototrophy to a bacterial amino acid auxotroph is therefore no indication of the

similarity of the amino acid biosynthetic enzymes in both organisms. Overall similarity can only be deduced by analysis of sequence data.

A local alignment strategy such as the one used in Figure 4.12 and Figure 4.13 is well-suited to discerning relationships between P5CRs from distantly related organisms. Comparison of the P5CRs from several phylogenetically distinct organisms reveals some conservation in certain regions of the polypeptides (Figure 4.13). This indicates that all P5CRs are likely to share essentially the same modular structure. As shown in Figure 4.13, homology amongst the P5CR sequences currently characterised is highest in an N-terminal region (Gly_{14} to Glu_{23} in *Arabidopsis* P5CR), two central regions (Arg_{124} to Gly_{134} and Asp_{173} to Tyr_{185} in *Arabidopsis* P5CR) and in a C-terminal region (Pro_{229} to Gly_{246} in *Arabidopsis* P5CR).

The universally conserved modular structure of the twelve P5CRs suggested in Figure 4.12 and moderate but significant homology of these proteins at the amino acid level (Figure 4.13) suggests that they are all derived from the same ancestral protein by a divergent evolutionary process. Convergent evolution of P5CRs from phylogenetically distinct organisms is unlikely as there are probably several possible sequence combinations capable of yielding an equivalent functional structure capable of the reduction of P5C.

Amino acid substitutions do not occur randomly along the sequences of P5CRs (Figure 4.13). As has been demonstrated in other proteins (Doolittle, 1989; Section 2.5.3), a high degree of identity is expected to be found around amino acids likely to be of functional importance in P5CR activity. Such residues may be involved in the binding of the pyridine nucleotide cofactor or the substrate P5C. The nature of evolutionary divergence permitted in the amino acid sequences of P5CRs is thus likely to reflect constraints on the structure and/or function of the enzyme. Identification of blocks of amino acids or even particular residues critical to the functionality of P5CR may be of value in future attempts to modify plant P5CR activity at the molecular level. This may be achieved using techniques such as site-directed mutagenesis. Such manipulation might involve increasing P5CR activity. This may enhance proline accumulation in transgenic crops and thereby increase their environmental tolerance and productivity. Alternatively, an alteration of amino acid residues known to be critical for P5CR activity may reduce or completely destroy activity of the enzyme. Transfer of the corresponding mutated gene into a model plant system such

as *Arabidopsis* may enable assessment of the relative importance of the enzyme in the tolerance of plants to environmental extremes.

Following alignment, 115 residues were found in identical positions in at least six of the twelve P5CR sequences compared (Figure 4.13). Of these conserved residues, 106 occur in the primary structure of *Arabidopsis* P5CR. Three amino acid residues are invariant throughout all of the sequences. In *Arabidopsis* P5CR, these are Asn_{128}, Gly_{241} and Thr_{243} (Figure 4.13).

Inhibition of plant P5CR activity by the carbonyl blocking agent hydroxylamine has been reported in both animal (Peisach and Strecker, 1962) and plant (Splittstoesser and Splittstoeser, 1973; Rena and Splittstoesser, 1975; Miler and Stewart,1976) P5CRs. This suggests the importance of an amino acid residue bearing a side chain containing a carbonyl group in the catalytic mechanism of P5CR. It is possible that inhibition of P5CRs by the carbonyl-blocking reagent hydroxylamine may reflect inactivation of the universally conserved asparagine residue in the central portion of all P5CRs aligned (Asn_{128} in *Arabidopsis* P5CR; Figure 4.13).

Consideration of the reaction catalysed by P5CR suggests that the enzyme must comprise at least two domains : one for binding of the reductant NAD(P)H and one for binding the substrate P5C. All nucleotide cofactor binding enzymes characterised to date fold into at least two distinct domains (Branden and Tooze, 1991). One binds the coenzyme, while the other binds the substrate and provides the amino acids necessary for catalysis. The active site occurs in a cleft between the two domains (Branden and Tooze, 1991). In most of the nucleotide-binding proteins presently characterised at the structural level, these domains are flexible (Branden and Tooze, 1991). This enables the two domains to move closer to each other during catalysis to ensure that reactants are completely shielded from the solvent during hydride transfer (Branden and Tooze, 1991).

With the exception of P5CR from *T. pallidum*, all P5CRs characterised contain an N-terminal NAD(P)H-binding domain (Figure 4.15). However, analysis of the sequences of the N-terminal domains of twelve P5CRs suggests that distinct differences occur in the cofactor preferences of the P5CRs studied. The observation that plant P5CRs appear to display a definite preference for NADPH as a reductant (Figure 4.15) is in keeping with

enzymatic studies in soybean (Kohl *et al.*, 1988; Szoke *et al.*, 1992) and tobacco (LaRosa *et al.*, 1991). Data indicating the affinities of these plant P5CRs for NADPH and NADH are presented in Table 2.4. Furthermore, the K_m of P5CR from tobacco for P5C is lower in the presence of NADPH than in the presence of NADH (LaRosa *et al.*, 1992; Table 2.4). This supports available evidence suggesting that plant P5CRs use NADPH in preference to NADH. Perfect matching to the NADPH-binding consensus in the N-terminal cofactor-binding domain of human P5CR is consistent with the lower K_m of P5CR from erythrocytes for NADPH than NADH (Merrill *et al.*, 1989), but not with the higher affinity of P5CR from human fibroblasts for NADH (Yeh *et al.*, 1981).

Considering that in metabolic terms, NADPH is much more expensive than NADH, preferential use of NADPH by plant P5CRs is consistent with a role for proline biosynthesis in plants in the regulation of the level of reduction of the cellular pool of NADP (Section 2.1.2.4). Human P5CR is the only other enzyme studied which appears to display a definite preference for NADPH as reductant (Figure 4.15). This molecular evidence for a definite cofactor preference in plant P5CRs may be of some significance in elucidating the role of proline in plants under conditions of environmental stress.

Several workers (Argandona and Pahlich, 1991; LaRosa *et al.*, 1991; Szoke *et al.*, 1992; Delauney and Verma, 1993) have concluded that P5CR is unlikely to play a major role in the control of flux through the biosynthetic pathway of proline in plants. LaRosa *et al.* (1991) compared the P5CR activities from salt-adapted and nonadapted tobacco cells. This revealed that although proline levels in the salt-adapted cells were two orders of magnitude in excess of proline levels in nonadapted cells, P5CR activities in the two cell types were not significantly different (LaRosa *et al.*, 1991). Furthermore, Szoke *et al.* (1992) demonstrated that a fifty-fold enhancement of P5CR activity in transgenic tobacco plants expressing soybean P5CR did not result in any significant increase in proline formation. Both LaRosa *et al.* (1991) and Szoke *et al.* (1992) concluded that P5CR activity *in vivo* is limited by the availability of the substrate P5C, and that the enzyme functions at only a fraction of its V_{max}. In the light of these conclusions that P5CR is not important in controlling proline production, it seems feasible to postulate that an important role of the enzyme during acclimation to environmental stress might be stabilisation of the NADP/NADPH couple in the cells of plants exposed to stress (Section 2.1.2.4).

The presence of only a single mismatch with the NADPH-binding site consensus (Hanukoglu and Gutfinger, 1989) in the cofactor binding site of P5CR from *P. aeruginosa* (Figure 4.15) suggests that this enzyme may preferentially use NADPH over NADH. However, this observation is not consistent with enzymatic studies on P5CR activity from *P. aeruginosa* (Krishna *et al.*, 1979). These workers reported that the K_m of P5CR from *P. aeruginosa* for NADPH is four times greater than that for NADH (Table 2.4).

Two mismatches with the NADPH-binding site consensus (Hanukoglu and Gutfinger, 1989) were found in the corresponding regions of the P5CRs from *B. subtilis* and *E. coli* (Figure 4.15). In contrast, analysis of the cofactor-binding domain of the P5CRs from *S. cerevisiae*, *M. smithii*, *T. thermophilus* and *M. leprae* revealed them to possess closer resemblance to the NADH-binding site consensus defined by Branden and Tooze (1991) than to the NADPH-binding site consensus. The evidence that P5CR from *S. cerevisiae* preferentially uses NADH instead of NADPH is in keeping with the higher affinity of the yeast P5CR for NADH in comparison with NADPH and the higher maximum activity of the enzyme with NADH than with NADPH (Matsuzawa and Ishiguro, 1980a; Table 2.4).

Examination of Figure 4.13 indicates that two glycine residues (Gly_{17} and Gly_{19} in *Arabidopsis* P5CR) are found in identical positions in eleven of the twelve sequences following alignment. These glycine residues are likely to play a crucial role in positioning the central part of NAD(P)H in its correct conformation close to the enzyme. As outlined in Section 2.5.1.3, these two glycine residues form a loop between the first β-strand and the succeeding α-helix found in the βαβ motif capable of binding the ADP-moiety of dinucleotide cofactors (Wierenga *et al.*, 1986). The first glycine residue is essential for the tightness of the turn at the end of the first β-strand. The second glycine residue allows the dinucleotide to be bound without obstruction from a bulky amino acid side chain at this position (Scrutton *et al.*, 1990).

The conserved hydrophobic residues in the N-terminal NAD(P)H-binding sites (Figure 4.15) are likely to form a core between the α-helix and β-strands (Branden and Tooze, 1991). The conserved aspartate residues in P5CRs from *M. leprae* (Asp_{33}), *S. cerevisiae* (Asp_{29}) and *M. smithii* (Asp_{23}) as well as Glu_{28} in P5CR from *T. thermophilus* (Figure 4.15) are probably important in binding NADH through hydrogen bonds between the 2'-hydroxyl group of its adenosine ribose (Wierenga *et al.*, 1986). This suggests that P5CRs from these organisms

are unlikely to bind NADPH as this cofactor has a phosphate group attached to the 2'-hydroxyl of the adenosine ribose. This would most likely cause repulsion between the negative charges of the phosphate group in the 2' position and the acidic amino acid residue (Wierenga *et al.*, 1986).

In contrast, it is expected that enzymes binding NADPH using this motif have an amino acid bearing a small uncharged side chain instead of a negatively-charged acidic residue such as aspartate or glutamate (Wierenga *et al.*, 1986; Branden and Tooze, 1991). This provides space for both the phosphate group and a positively charged side chain nearby for binding this phosphate (Branden and Tooze, 1991). This is consistent with the presence of glycine residues in positions of the putative NADPH binding sites of P5CRs from *A. thaliana*, *G. max*, *P. sativum*, *H. sapiens*, *P. aeruginosa* and *E. coli* that correspond with those of the negatively charged acidic residues found in NADH-binding sites (Figure 4.15).

As outlined in Section 2.2.1.2, most P5CRs characterised to date can recognise either NADH or NADPH as cofactors. The ability of oxidoreductases to recognise both NADH and NADPH is not unusual. For example, human glutathione reductase binds both cofactors, although the K_m for NADH is approximately sixty times higher than that for NADPH (Pai *et al.*, 1988). This is because NADH lacks the 2'-phosphate group of NADPH. It is believed that NADH binds to human glutathione reductase in a similar manner to NADPH, except that an inorganic phosphate ion substitutes for the missing 2'-phosphate group on the coenzyme (Pai *et al.*, 1988). It is likely that a similar situation may apply in the case of P5CRs. Despite the kinetic (Table 2.4) and molecular (Figure 4.15) evidence that plant P5CRs appear to preferentially bind NADPH, this does not necessarily imply that NADPH is used by the enzyme *in vivo*. The choice of which of the two possible cofactors is used may depend on the sizes of the two pools of pyridine nucleotide cofactors within the plant cell.

The apparent absence of an NAD(P)H binding site in P5CR from *T. pallidum* may be related to the parasitic lifestyle of this organism. *Treponema pallidum*, the causative agent of syphilis, is generally considered to be a highly fastidious organism with complex growth requirements. To date, it has been impossible to culture this organism *in vitro* on artificial medium (Gherardini *et al.*, 1990). The exacting nutritional requirements of this organism may in some way be related to the fairly unique nature of the N-terminus of its P5CR polypeptide.

Besides the N-terminal NAD(P)H-binding domain, all P5CRs appear to share particularly high homology at their C-terminal ends. In particular, a region located between the amino acid residues Pro_{229} and Gly_{246} in P5CR from *Arabidopsis* is highly conserved in eleven other P5CRs from the same and different biological kingdoms (Figure 4.13). This region includes two invariant amino acid residues. In *Arabidopsis* P5CR, these are the residues Gly_{241} and Thr_{243} (Figure 4.13). Also found within this region, which is eighteen amino acids in length, are three amino acid residues found in eleven of the twelve sequences aligned. In *Arabidopsis* P5CR, these are the residues Pro_{229}, Leu_{232} and Pro_{239} (Figure 4.13).

Another region of high homology is found in the central portion of all P5CR polypeptides which have been characterised to date (Figure 4.13). This region, located between Arg_{124} and Gly_{134}, is centred around an invariant asparagine residue (Asn_{128} in *Arabidopsis* P5CR). It is tempting to speculate that highly conserved amino acids within these central and C-terminal domains may be involved in binding of the substrate P5C or contribute to the catalytic mechanism of the enzyme.

In attempting to identify residues critical to P5CR activity, it may of value to consider the results of chemical-induced mutation studies of the *T. thermophilus* gene encoding P5CR (Hoshino *et al.*, 1994). These workers demonstrated that substitution of Gly_{230} in P5CR from *T. thermophilus* with an aspartate destroys P5CR activity (Hoshino *et al.*, 1994). This lethal substitution occurs in the C-terminal domain defined above (Figure 4.13). This suggests that this glycine residue (Gly_{246} in the *Arabidopsis* P5CR), which is also conserved in P5CRs from soybean, pea, *S. cerevisiae*, *T. pallidum*, *B. subtilis* and *M. smithii*, may also be critical to functionality of P5CR in these species. This supports the proposal that the C-terminal domain defined above may be important in catalytic activity of *Arabidopsis* P5CR. Analysis of another lethal substitution in the *T. thermophilus* gene encoding P5CR involving the replacement of Gly_{148} (Figure 4.13) with glutamate suggests that Gly_{162} in *Arabidopsis* P5CR may also be critical to the functioning of the enzyme. This glycine residue is conserved in the soybean, pea, human, yeast, *E. coli*, *B. subtilis*, *P. aeruginosa* and *M. leprae* P5CR sequences (Figure 4.13). However, it is not within a region of highly conserved residues (Figure 4.13). Alignment of the *T. thermophilus* sequences with these sequences has resulted in the glycine residue being displaced to the right by a single position in this region (Figure 4.13).

Analysis of the positions of introns within the *Arabidopsis* gene encoding P5CR (Verbruggen *et al.*, 1993) reveals that these do not occur within parts of the gene encoding highly conserved stretches of amino acids (Figure 4.13). The shuffling of domains between proteins is currently accepted as one mechanism whereby proteins have evolved (Section 2.5.3). It has been proposed that introns within eukaryotic genes may have resulted from genetic recombination bringing together domains from originally different genes (Gilbert, 1978; Blake, 1979). It would be interesting to compare the position of exon boundaries in P5CRs from plant and other eukaryotic sources in order to investigate whether, as observed for other homologous enzymes or members of large protein families (Gilbert *et al.*, 1986), there is a correspondence between exons and structural modules within P5CRs. Unfortunately, the *Arabidopsis* gene encoding P5CR is the only gene from a plant or animal for which a genomic sequence is available. It is therefore not possible to compare its intron structure with that found in other plants or animals.

In contrast to the conclusion that an N-terminal region of plant P5CRs is involved in binding of NAD(P)H (Verbruggen *et al.*, 1993; this study), Williamson and Slocum (1992) speculated that a C-terminal domain of pea P5CR (spanning the region from Val_{234} to Ala_{267} in the *Arabidopsis* P5CR) may be involved in the binding of NAD(P)H. This was based on their observation that the secondary structure of the sequence fitted the overall consensus conformation characteristic of the βαβ-fold involved in binding of the ADP moiety of NAD-type cofactors (Wierenga *et al.*, 1986; Section 2.5.1.3). Williamson and Slocum (1992) further suggested that the amino acid sequence GTTIAG (found in *Arabidopsis* and human P5CRs as well as in pea P5CR; Figure 4.13) may be an NAD(P)H binding site. This sequence has homology to the GTGIAP motif of the site capable of binding the ribose-moiety of NAD(P)H in several prokaryotic and eukaryotic enzymes (Bredt *et al.*, 1991).

Nevertheless, although the sequence GTTIAG does contain two of the three universally conserved residues in all of the P5CRs studied (Gly_{241} and Thr_{243} in *Arabidopsis* P5CR), it is unlikely that the argument of Williamson and Slocum is tenable. Firstly, analysis of the most likely secondary structures of the C-terminal regions in P5CRs from *Arabidopsis* (Table 4.5), soybean (Table 4.6), pea (Table 4.7), human (Table 4.8) and *E. coli* (Table 4.9) is not consistent with the βαβ folding motif for NAD(P)H-binding sites defined by several workers (Wierenga *et al.*;1986; Hanukoglu and Gutfinger, 1989; Branden and Tooze, 1991; Section 2.5.1.3). Secondly, the comparatively low homology of the relevant regions of the

P5CRs with the I--G-GTGIAPF--F motif for binding the ribose moiety of NADPH (Bredt *et al.*, 1991) is not likely to be significant. The greatest number of matches to this consensus sequence was observed in the C-terminal region of P5CR from *S. cerevisiae* (27% identity).

Furthermore, the site with the core consensus sequence GTGIAP is only capable of binding the ribose moiety of NADPH (Bredt *et al.*, 1991). To date, all enzymes shown to possess the site capable of binding the ribose moiety also possess a region capable of binding the adenine moiety of NADPH. This stretch is found between 90 and 100 residues towards the C-terminal portion of the ribose-binding region (Bredt *et al.*, 1991). All P5CRs characterised to date, with the exception of human P5CR, end within 42 residues on the C-terminal portion of the start of the sequence GTTIAG. The C-terminal extension of human P5CR bears no homology to the consensus sequence of the binding site of the adenine moiety of NADPH defined by Bredt *et al.* (1991). Although it is possible that folding of the polypeptide chain might bring such a domain capable of binding the adenine moiety within close proximity to the ribose moiety-binding domain proposed by Williamson and Slocum (1992), analysis of other regions of P5CRs failed to suggest any homology to the consensus sequence for the adenine moiety-binding domain defined by Bredt *et al.* (1991).

These flaws in the argument presented by Williamson and Slocum (1992) contrast with the high consensus sequences for NADH- (Branden and Tooze, 1991) and NADPH- (Hanukoglu and Gutfinger, 1989) binding sites found at the N-terminal end of all of the P5CRs studied with the exception of P5CR from *T. pallidum* (Figure 4.15). In their search for a putative NADPH-binding site in human P5CR, Dougherty *et al.* (1992) also used the key for NADPH-binding sites defined by Bredt *et al.* (1991). Although these workers did not find any significant homology with the ribose- and adenine-moiety binding sites within human P5CR, they attributed this to the ability of human P5CR to use either NADH or NADPH (Dougherty *et al.*, 1992).

As discussed above, a number of similarities exist in the primary structures of P5CRs from phylogenetically-distinct organisms. This introduces the possibility that analysis of the sequences of P5CRs from organisms belonging to different biological kingdoms may be a useful criterion for determining phylogenetic relationships over the entire spectrum of living organisms. In order to achieve such a molecular-based classification, one needs a molecule

of appropriately broad distribution. Ribosomal RNA (rRNA) sequences have been used extensively in molecular systematic studies (Pace *et al.*, 1986). However, Stackebrandt (1988) has presented evidence that 16S rRNA sequences tend to overestimate phylogenetic relationships between bacteria. It seems possible that P5CR, an enzyme found in all protists, animals and plants capable of proline synthesis (Adams and Frank, 1980), may be a good choice as a molecule to estimate the evolutionary relatedness of phylogenetically-distinct organisms.

The phenograms presented in Figure 4.11 are not in complete agreement with the widely-accepted universal phylogenetic tree determined from analysis of rRNA sequences (Pace *et al.*, 1986). In this arrangement, plants are more closely related to fungi (such as *S. cerevisiae*) than to vertebrates (such as human). The Eubacteria form a grouping distinct from eukaryotes (Pace *et al.*, 1986). Therefore, use of P5CR does not appear to be a good choice in assessing evolutionary relatedness between organisms. It is likely that constraints on the structure and/or function of this enzyme have prevented random alteration of the sequence with time. However, considering that the product of only a small portion of the genome of each of the organisms has been examined and that no choices were made other than those dictated by the statistical analyses, the results are very promising.

Owing to the arbitrary nature of the rule by which the P5CRs have been clustered, it is not possible to decide which phenogram represents the best use of the information. However, both clustering methods provided phenograms with several common features. In particular, the three plant P5CRs form a distinct cluster in both of the phenograms shown in Figure 4.11. Both of the algorithms used suggest that on the basis of amino acid identities, plant P5CRs bear the greatest resemblance to P5CRs from human, *E. coli* and *T. thermophilus*. Both methods assigned the P5CRs from *M. leprae* and *P. aeruginosa* to a discrete cluster.

The high degree of similarity of P5CRs from soybean and pea is consistent with the short genetic distance between these two legumes. The distinct cluster formed by the three plant P5CRs (Figure 4.11) is consistent with the high degree of similarity shown between these enzymes (Figure 4.7). In this context it is interesting to note that several molecular-based phylogenetic studies have suggested that the Fabaceae (of which soybean and pea are members) and Brassicaceae (of which *Arabidopsis* is a member) are closely related. This was originally inferred by Ramshaw *et al.* (1972) using amino acid sequence data of

cytochrome *c* from different flowering plants. In a series of studies (Martin *et al.*, 1983; Martin and Dowd, 1984a, 1984b, 1984c), Martin and co-workers used sequence data from the small subunit of ribulose-1,5-bisphosphate carboxylase, cytochrome *c* and plastocyanin proteins of representative species from several families to construct consensus phylogenetic trees for the Angiospermae. In all of these, the Fabaceae and Brassicaceae occur together. Combination of the data from these three proteins with data from ferredoxin and 5S rRNA sequences of eleven species of angiosperms also indicated close relatedness of the Brassicaceae and Fabaceae (Martin *et al.*, 1985). The high degree of similarity of P5CR from *Arabidopsis* with the homologous proteins from soybean and pea (Figure 4.7, Table 4.4, Figure 4.11) is therefore consistent with other molecular systematic studies of the angiosperms. However, sequence data for P5CR from plant species belonging to families other than Fabaceae and Brassicaceae is currently not available. It is therefore not possible to comment on how strongly conservation of plant P5CR sequences has been selected for during the course of evolution.

Also noticeable in Figure 4.11 is the closer identity of the plant P5CRs to P5CRs from Gram-negative bacteria than to Gram-positive bacteria. In general, Gram-negative bacteria achieve high intracellular concentrations of proline only by enhanced transport (Csonka, 1989). Osmotic stress has no effect on the rates of proline synthesis or degradation in these organisms (Section 2.2.1). In contrast, data is available to suggest that in certain Gram-positive bacteria such as *Brevibacterium lactofermentum* (Kawahara *et al.*, 1989; Kawahara *et al.*, 1990) and *B. subtilis* (Whatmore *et al.*, 1990), proline is synthesised during osmotic stress when proline is not present in the environment. This capacity for osmotically-induced proline synthesis is not reflected in the similarity of P5CR primary sequences from Gram-positive bacteria to plant P5CRs (Figure 4.11). This suggests that the greater relatedness of plant P5CRs to P5CRs from Gram-negative bacteria is unlikely to be related to stress-induced proline biosynthesis by P5CR. This suggests that in plants, the increase in P5CR expression accompanying hyperosmotic stress (Treichel, 1986; Argandona and Pahlich, 1989; Laliberte and Hellebust, 1989b; Section 2.1.2.2) is most probably mediated at the genetic level.

All of the P5CR sequences studied possess low similarity to the P5CR sequence from the archaebacterium *Methanobrevibacter smithii* (Table 4.4, Figure 4.11). This is consistent with the separate evolutionary heritage of the archaebacteria and their classification as a coherent

group distinct from both eubacteria and eukaryotes (Woese and Fox, 1977). The apparent antiquity of these organisms is supported by the fact that the methanogenic phenotype is well suited to the environment presumed to exist on earth three to four billion years ago. The low, but nevertheless significant similarity of eukaryotic and eubacterial P5CRs to P5CR from an archaebacterium (Figure 4.13) supports the proposal already presented above that P5CRs have diverged from a common ancestral protein.

The poor homology of the *Bacillus subtilis* sequence with the other P5CR sequences studied (Table 4.4, Figure 4.11) may be related to the possibility that this protein may not be a functional P5CR (Ahn and Wake, 1991). Despite significant similarity to other P5CR sequences (Figure 4.13) and no significant similarity to any other sequences currently in international sequence databases, a deletion in the *B. subtilis* 168 strain SU153 spanning this ORF does not cause proline auxotrophy (Ahn and Wake, 1991). Furthermore, addition of proline does not stimulate growth of this mutant on a minimal medium. This finding has yet to be accounted for. Southern hybridisation experiments have failed to indicate the presence of any other *proC*-like sequences in the *B. subtilis* genome (Ahn and Wake, 1991). It is possible that this bacterium may possess some means of overcoming proline auxotrophy in the absence of P5CR activity.

It is important to note that the approach used to trace the divergence of P5CRs assumes that the rates of divergence are roughly equal in all branches of the phenogram. In other words, this presumes that P5CR has undergone a fairly constant rate of substitution through time in different organisms. Since the rate of change depends on the mutation rate, this assumes that many of the amino acid substitutions are largely neutral. Furthermore, use of P5CR sequences in phylogenetic classification is based on the assumption that the enzymes used in the analysis are encoded by genes that possess a common origin and are related by descent. The highly conserved modular structure of P5CRs indicated in Figure 4.12 and Figure 4.13 suggests that this assumption is valid.

The data presented in Figure 4.11 cannot strictly be considered as phylogenies of the organisms studied, but rather merely as phylogenetic hypotheses of the P5CRs found in these organisms. Phylogenies deduced from a single molecule cannot be divorced from phylogenies of the organisms generated from other data (Crawford, 1990). It is not possible to conclusively infer phylogenetic relationships based exclusively on analysis of a single molecule. Therefore, this data needs to be combined with sequences from other proteins if

more stable phylogenies are to be established.

The phenograms presented in Figure 4.11 were constructed on the basis of percentage identity and not percentage similarity. Therefore, conservative substitutions in the P5CRs studied have not been taken into account. Furthermore, the global alignment strategy used to generate the values presented in Table 4.4 is likely to underestimate the similarity of proteins, especially if they are distantly related (Schuler *et al.*, 1991). It is likely that a more accurate reflection of the divergence could be generated if the folded conformations of the P5CRs had been considered in the construction of the phenograms. Homologous proteins, even from distantly related organisms, often share a common folding pattern and similar topology (Lesk and Chothia, 1986 cited by Pardo and Serrano, 1989). This suggests that the comparatively low overall sequence identity of P5CRs from distantly related organisms (Figure 4.13, Table 4.4) may be underestimate their relatedness.

Despite the poor success record of protein structure predictive methods (Section 2.5.2), analysis of the most likely features of the secondary structures (Figure 4.16; Tables 4.5 to 4.9) and of the hydrophobicity profiles (Figure 4.17) of five representative P5CRs suggests several similarities in the structures of these five enzymes. The P5CRs for which structural predictions were made were those from *Arabidopsis* (Verbruggen *et al.*, 1993), soybean (Delauney and Verma, 1990), pea (Williamson and Slocum, 1990), human (Dougherty *et al.*, 1992) and *E. coli* (Deutch *et al.*, 1982). The similarity in secondary structure is particularly evident in a graphical representation of the probabilities of individual residues to occur in regions of α-helix or β-sheet (Figure 4.16).

However, owing to the uncertainties of present methods of structure prediction (Fasman, 1989, Section 2.5.2.1), the results of these analyses cannot be accepted unquestioningly. Although many of the predictions made using the methods of Chou and Fasman (1978) and Garnier *et al.* (1978) are identical, discrepancies between the two predictive strategies do occur. For example, the two predictive strategies differed in their prediction of regions comprising α-helix and β-sheet in regions of the *Arabidopsis* P5CR between residues 10-16, 151-165 and 184-188 (Table 4.5). Nonetheless, the pattern emerging from Figure 4.16 is that, making allowance for a range of error in the predictive method used, there is good conservation of predicted secondary structure in all five of the P5CRs for which secondary structure was predicted.

The most conserved regions of P5CRs are predicted to be surface loops or turns (Figure 4.13; Tables 4.5 to 4.9). For example, the stretches of high homology amongst P5CRs centred around Pro_{127}, Gly_{183}, Gly_{201} and Pro_{239} in *Arabidopsis* P5CR (Figure 4.13) are predicted to occur in turns or coil structures in P5CRs from *Arabidopsis* (Table 4.5), soybean (Table 4.6), pea (Table 4.7) and human (Table 4.8).

Analysis of the numerical data obtained using the predictive methods of Chou and Fasman (1978a, 1978b) and Garnier *et al.* (1978) suggests that both proline residues found in the C-terminal putative P5C-binding domain (Pro_{229} and Pro_{235} in *Arabidopsis* P5CR) cause turns in the folding of the polypeptide chain (Table 4.5). These two proline residues are also conserved in all of the other P5CRs for which secondary structural predictions were made. As in the *Arabidopsis* P5CR, they occur in turns or surface coil structures in P5CRs from soybean (Table 4.6), pea (Table 4.7), human (Table 4.8) and *E. coli* (Table 4.9). Given the structural constraints imposed by proline residues in protein structure, it is likely that the high degree of conservation of these residues may be related to spatial requirements within the local environment of the folded chain.

Examination of the structural features (Table 4.5) of regions in *Arabidopsis* P5CR where introns occur (Figure 4.13) reveals that splice sites for the third, fourth and fifth introns (located adjacent to codons encoding Gln_{114}, Ser_{138} and Ser_{179} respectively) correspond with regions predicted to occur as turn or coil structures (Table 4.5). As outlined in Section 2.5.1.1, coil structures usually occur on the surfaces of globular proteins. Hydrophobicity data presented in Figure 4.17 are consistent with the suggestion that Gln_{114} and Ser_{179} found in *Arabidopsis* P5CR are surface residues. Craik *et al.* (1983) have demonstrated that in soluble proteins, intron-exon splice junctions frequently map to surface loops which separate elements of secondary structure.

As already discussed, alternative splicing of intron-exon junctions is one mechanism whereby slightly different proteins may arise from transcription of a single gene (Section 2.4.2.2). Such a mechanism has been proposed to account for the existence of kinetically distinguishable isoforms of P5CR in different animal tissues despite the existence of only a single gene encoding P5CR in human (Dougherty *et al.*, 1992). The occurrence of intron-exon junctions at protein surfaces suggests that alternative splicing of introns can be effected without disrupting the stability of the protein core (Craik *et al.*, 1983). Although

as already discussed, no evidence has been obtained for alternative splicing of pre-mRNAs encoding P5CR in *Arabidopsis* (Verbruggen *et al.*, 1993; this study), it is probable that if such a regulatory mechanism does exist in plants, then the third, fourth and fifth introns are the most likely candidates for alternative splicing. Alternative splicing of the first, second and sixth introns (located adjacent to codons encoding Glu_{57}, Val_{84} and Thr_{214} is less likely as these occur within regions of organised secondary structure (Table 4.5). Alteration of the splicing of these introns would therefore most probably disrupt the stability of the hydrophobic core of the enzyme and destroy its capacity for catalysis.

Analysis of the primary structure of a polypeptide provides no indication of the final three-dimensional conformation assumed by the polypeptide *in vivo*. Despite the fact that the N-terminal NAD(P)H-binding domain and C-terminal domain likely to be capable of binding P5C are located at the opposite extremes of the polypeptide chain, it is likely that these may be brought together by folding of the polypeptide chain into its tertiary structure. The active site of the enzyme might be expected to occur at the interface between the two domains. Folding of the polypeptide chain may also bring the region of homology surrounding the universally conserved asparagine residue in the centre of the polypeptide chain (Asn_{129} in *Arabidopsis* P5CR) into close proximity with the other domains. Furthermore, the possibility of movement of the domains relative to one another as part of the catalytic mechanism cannot be disregarded. However, these proposals are at best highly speculative. The three-dimensional structure of P5CR may only be conclusively established using techniques such as X-ray crystallography or nuclear magnetic resonance.

Using four different predictive strategies for secondary structure including that of Garnier *et al.* (1978), Verbruggen *et al.* (1993) also concluded that the secondary structures of P5CRs from *Arabidopsis*, soybean, human, *S. cerevisiae*, *E.coli*, *P.aeruginosa* and *M. smithii* closely resembled each other. Interpretation of these predictions suggested that P5CR comprises one or more β-sheets surrounded by α-helices to provide a thermodynamically stable structure (Verbruggen *et al.*, 1993).

Using the predictive method of Garnier *et al.* (1978) the overall secondary structural content in *Arabidopsis* P5CR is 54.3 % α-helical and 16.3 % β-sheet structure. This suggests that the enzyme belongs to one of either the α+β or α/β classes of protein (Levitt and Chothia, 1976; Section 2.5.1.4). These two classes of protein are characterised by an α-helical content

greater than 15% and more than 10% of the chain in β-sheet conformation (Nakashima *et al.*, 1986 cited by Fasman, 1989). However, since secondary structural predictions (Tables 4.5 to 4.9) suggest that P5CRs are comprised of approximately alternating α-helices and β-strands, P5CR appears to belong to the α/β class of proteins.

Analysis of amino acid usage in *Arabidopsis* P5CR (Table 4.2) is consistent with its classification as an α/β protein. Not only do the sizes of P5CRs suggest that they belong to the α/β class of proteins (Chou, 1989; Section 2.5.2.2), but the high hydrophobic content of *Arabidopsis* P5CR and low content of cysteine are also consistent with this classification (Chou, 1989; Section 2.5.2.2). Furthermore, the comparatively low contents of asparagine and tyrosine residues contrast with the high abundance of valine in *Arabidopsis* P5CR (Table 4.2). These features are characteristic of α/β proteins (Chou, 1989).

In α/β domains, substrate binding crevices are formed by loop regions (Branden and Tooze, 1991). As already discussed, all of the regions displaying high homology amongst the twelve P5CRs aligned in Figure 4.13 are predicted to occur at the surfaces of the enzymes for which secondary determinations have been made (Tables 4.5 to 4.9). Loop regions in α/β proteins generally do not contribute to structural stability but participate in binding and catalytic action (Branden and Tooze, 1991).

On the basis of this classification of *Arabidopsis* P5CR as an α/β protein, it may be possible to use conformational parameters derived from known proteins of this class to predict the three-dimensional conformation assumed by *Arabidopsis* P5CR. Future investigations might involve comparison of P5CR with α/β proteins of known three-dimensional structure. For example, all of the glycolytic enzymes are α/β structures as are many proteins involved in the binding and transporting of metabolites (Branden and Tooze, 1991).

There are two main classes of α/β proteins (Branden and Tooze, 1991). In the first class, a core of eight twisted parallel β-strands are arranged close together to form a barrel. α-helices and loops that connect the parallel β-strands are all on the outside of this barrel (Branden and Tooze, 1991). In 1989, sixteen enzymes of diverse origin but belonging to the α/β class of proteins were known to possess such an eight-stranded β-barrel structure (Chothia, 1988; Nagano, 1989).

It is possible that P5CR may have a structure resembling this common folding motif. In this regard, it is particularly encouraging to note that both of the methods used in the prediction of secondary structure of *Arabidopsis* P5CR (Table 4.5) predict eight regions of β-strand likely to participate in sheet formation. Analysis of the secondary structural predictions of the four other P5CRs investigated (Tables 4.6 to 4.9) are also consistent with the presence of at least eight regions of β-strand within the polypeptide chains.

All P5CRs for which sequences are currently available satisfy another requirement for a barrel structure. This is that the protein should comprise at least 200 residues (Branden and Tooze, 1991). The high content of hydrophobic residues in *Arabidopsis* P5CR (Table 4.2) is also consistent with observation that hydrophobic side chains dominate the core of α/β barrel structures (Branden and Tooze, 1991). In particular, valine, isoleucine and leucine comprise approximately 40% of the residues in the β-strands that make up the hydrophobic core of α/β barrels (Branden and Tooze, 1991). While the proposal is merely speculation, the abundance of barrel structures in proteins belonging to the α/β class suggests that this structure has arisen independently several times throughout evolution. It provides a conformation in which stability of protein structure can be combined with the functional requirements of many enzymes belonging to the α/β class of proteins (Branden and Tooze, 1991).

Another conformation common to many proteins of the α/β class is that resembling the structure of lactate dehydrogenase. In this structure, a central,open, predominantly parallel twisted β-sheet is surrounded by an array of α-helices or loops (Darby and Creighton, 1993). This $(\beta\text{-}\alpha\text{-}\beta\text{-}\alpha\text{-}\beta)_2$ unit, or Rossman fold has been found in many enzymes capable of binding dinucleotide cofactors (Darby and Creighton, 1993). Almost all binding sites in this class of α/β proteins are located in loops that connect the carboxy ends of the β-strands with the amino end of the α-helices (Branden and Tooze, 1991). However, in contrast to the highly conserved arrangement of eight α-helices and eight β-strands found in α/β barrels, considerable variation occurs in the structures of open α/β structures (Branden and Tooze, 1991). It might therefore be more difficult to prove that P5CR assumes an open α/β structure. However, the occurrence of this structure amongst many other enzymes that also bind dinucleotide cofactors suggests that this may be possible.

Analysis of the hydrophobicity profile of *Arabidopsis* P5CR (Figure 4.17) indicates that despite the comparatively high percentage composition of hydrophobic amino acids (Table 4.2), these are not sufficiently clustered in any particular region of the polypeptide to constitute a region likely to be capable of spanning or being anchored within a membrane (Engelman *et al.*, 1982; Kyte and Doolittle, 1982; Argos *et al.*, 1982; Section 2.5.2.3). The P5CRs of the organisms for which hydrophobicity data is presented in Figure 4.17 all appear to be soluble proteins. This conclusion is consistent with enzymatic studies conducted in plants (Krueger *et al.*, 1986; LaRosa *et al.*, 1991; Chilson *et al.*, 1992), animals (Merrill *et al.*, 1989) and bacteria (Costilow and Cooper, 1978; Meile and Leisinger, 1982) which have failed to indicate that P5CR may be associated with a particulate cellular fraction. Localised regions of hydrophobicity within P5CRs most probably represent stretches of the polypeptide chain concealed within the hydrophobic core of the enzyme.

The hydrophobicity profiles presented in Figure 4.17 indicate that P5CRs from *Arabidopsis*, soybean, pea, human and *E. coli* appear to possess a predominantly hydrophobic N-terminus and a hydrophilic C-terminus. The presence of a localised region of hydrophobicity within the N-terminal regions corresponding to the putative NAD(P)H-binding pockets of P5CRs (Figure 4.15; Figure 4.17) is consistent with the need for this region to partially enclose the hydrophobic nicotinamide ring of pyridine dinucleotide cofactors (Darby and Creighton, 1993). Comparison of hydrophobicity within the polypeptide chains of the five P5CRs studied (Figure 4.17) suggests conservation in the tendency of certain stretches of the polypeptide chain to be concealed within the hydrophobic core.

Analysis of the hydrophobicity profile of pea P5CR (Williamson and Slocum, 1992) is not consistent with the suggestion of Rayapati *et al.* (1989) that a chloroplastic isoform of P5CR in *P. sativum* may be associated with thylakoid membranes. However, this possibility cannot be completely disregarded. Overall hydrophobicity alone does not always provide an unambiguous means of identifying membrane-associated regions in sequences (Finer-Moore *et al.*, 1989). Eisenberg *et al.* (1982) have demonstrated that if the hydrophobic periodicity of an α-helical region is extreme, the helix may bind to membrane surfaces despite the absence of a range of residues with sufficient hydrophobicity to span a lipid bilayer. In such amphipathic helices (Section 2.5.1.1), charged groups cluster on one side of the helix and hydrophobic residues on the other.

In this context, it is interesting to note that although cytosolic, human P5CR migrates with plasma membranes on sucrose density gradients (Yeh and Phang, 1981). Mixson and Phang (1988) have speculated that a membrane-associating domain on this enzyme may allow physical association of P5CR with a putative membrane-bound P5C carrier in humans. The only stretch of amino acids found in human P5CR likely to possess such an ability to associate with a membrane is found between the amino acid residues His_{99} and Ser_{103} (Figure 4.17). This region of hydrophobicity is also conserved in the plant and *E. coli* P5CRs (Figure 4.17).

In the light of this observation, it is important to note that the application of predictive methods derived on data from globular proteins cannot be used unquestioningly in prediction of membrane-embedded proteins (Fasman, 1990). By 1990, only two membrane protein structures had been determined at high resolution by X-ray diffraction methodologies (Fasman, 1990). Both are photosynthetic reaction centres. Therefore, Fasman (1990) warns that one cannot yet evaluate the accuracy of the majority of predictions of membrane proteins. The same caution in the use of current predictive strategies to identify whether or not a protein spans a membrane is suggested by Popot and Engelman (1990). Therefore, it is not possible to conclusively dismiss the proposal of Rayapati *et al.* (1989) that plant P5CR may be a membrane-associated protein.

Analysis of the N-terminal sequence of *Arabidopsis* P5CR (Verbruggen *et al.*, 1993) reveals no structural similarity to the consensus chloroplast or mitochondrial target sequences such as an abundance of hydrophobic residues and a net positive charge (Karlin-Neumann and Tobin, 1986; White and Scandalios, 1988; Von Heijne *et al.*, 1989). More recently, Von Heijne and Nishikawa (1991) suggested that the most important determinant for chloroplast targeting is not so much a particular sequence, but rather the overall secondary structure of the N-terminal region of a polypeptide. In the case of chloroplast targeting peptides, there is an absence of ordered secondary structure in the N-terminal region (Von Heijne and Nishikawa, 1991). Therefore in general, N-terminal targeting peptides assume a random coil conformation. This contrasts with the ordered secondary structure of the N-terminal domains of plant P5CRs characterised to date (Tables 4.5 to 4.7). The regions of coiled structure at the immediate N-termini of P5CRs from *Arabidopsis* (Table 4.5) and soybean (Table 4.6) are of insufficient length to constitute a targeting peptide (Von Heijne *et al.*, 1989; Von Heijne and Nishikawa, 1991).

Furthermore, expression of the soybean cDNA isolated by Delauney and Verma (1990) in transgenic tobacco resulted in the accumulation of the enzyme exclusively in the cytosol (Szoke *et al.*, 1992). The high degree of similarity of *Arabidopsis* P5CR to soybean P5CR at the N-terminal regions of both enzymes (Figure 4.7) suggests that *Arabidopsis* P5CR also lacks a transit peptide capable of targeting the enzyme to the plastid.

Reports of chloroplastic P5CR in tobacco (Noguchi *et al.*, 1966) and pea (Rayapati *et al.*, 1989) cannot be explained using the molecular data on plant P5CRs which is currently available. To date all nuclear-encoded chloroplast proteins analysed in detail are synthesised as precursors with N-terminal transit peptides 35-80 amino acids in length that are cleaved off during or immediately after import into the plastid (White and Scandalios, 1988). An alternative possibility to explain the chloroplastic location of P5CR reported in tobacco (Noguchi *et al.*, 1966) and pea (Rayapati *et al.*, 1989) is that P5CR may be encoded by the chloroplastic genome. Although Delauney and Verma (1990) reported preliminary data from Southern analysis implicating the existence of a P5CR gene in soybean chloroplastic DNA, these results were not presented.

It is unlikely that the *Arabidopsis* P5CR gene corresponding to the cDNA YAP057 is chloroplastic since it contains six introns. Owing to the prokaryotic origin of chloroplasts, chloroplastic genes seldom possess introns (Sugiura, 1992). Furthermore, most of the introns of genes within the plastome possess boundary sequences that differ from those found in nuclear genes of eukaryotes (Sugiura, 1992). Many introns within genes of the plastome are likely to be self-splicing (Sugiura, 1992). The GT and AG dinucleotides at the 5' and 3' ends of introns of the P5CR gene from *Arabidopsis* (Verbruggen *et al.*, 1993; Figure 4.5) suggest that the gene from which the YAP057 cDNA is derived is located in the nucleus. These boundary sequences are characteristic of all known introns within eukaryotic nuclear genes (Csank *et al.*, 1990). In addition, the average chloroplastic genome contains only approximately 120 kb of unique sequence (Sugiura, 1992). Assuming a size of 1.2 kb for an average gene, this suggests that it is likely to comprise no more than 100 genes. Taking into account rRNA and tRNA genes, this leaves no more than 50 genes encoding proteins (Sugiura, 1992). Many of these, such as the large subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase, are involved in photosynthesis. Furthermore, in an extensive review of plant amino acid biosynthetic genes and enzymes, Coruzzi (1991) reported that although many amino acid biosynthetic enzymes are chloroplastic, none have been reported to be encoded by the plastome.

The conclusion that P5CR activity in *Arabidopsis* is restricted exclusively to the cytoplasm is supported by the results of Southern genomic analysis conducted during this study. Genomic copy number determination (Figure 4.18) suggests that the gene encoding P5CR is present in a single copy in the *Arabidopsis* genome. A much more complex pattern of genomic fragments would be expected if the gene was present in multiple copies. It is possible that the faint hybridisation to a 7.1 kb fragment in the lane containing *Eco*RI-digested genomic DNA represents the presence of a P5CR pseudogene in *Arabidopsis*. This may have arisen by duplication of a P5CR gene and subsequent divergence from the functional P5CR gene. With the exception of *Xba*I, none of the enzymes used in the copy number determination (Figure 4.18) cut within the genomic clone encoding P5CR (Verbruggen *et al.*, 1993). The presence of two bands of equal intensity following electrophoresis of the *Xba*I digest is consistent with the existence of an *Xba*I site within the gene encoding P5CR in *Arabidopsis* (Verbruggen *et al.*, 1993). This site occurs within the third intron at nucleotide position 1690 in the genomic clone gP5CR23 (Verbruggen *et al.*, 1993; Figure 4.5).

Comparison of the Southern blot shown in Figure 4.18 with the results of genomic DNA analysis performed by Verbruggen *et al.* (1993) reveals that these workers obtained an identical banding pattern with *Eco*RI-digested genomic DNA from the *Arabidopsis* ecotype Landsberg *erecta*. However, although these workers also observed hybridisation to two fragments generated by digestion of genomic DNA with *Xba*I, these fragments (approximately 1.6 kb and 0.95 kb) were much smaller than the sizes obtained in this study (approximately 4.2 kb and 2.9 kb). This difference most probably arises from polymorphisms in *Xba*I sites upstream and downstream of the genes encoding P5CRs in the two different *Arabidopsis* ecotypes. Whereas Verbruggen *et al.* (1993) used the ecotype Landsberg *erecta*, all work in this study employed the ecotype Columbia. In this context it is interesting to note that *Xba*I is particularly suited to finding random fragment length polymorphisms between the *Arabidopsis* ecotypes Columbia and Landsberg *erecta* (Anderson, 1994). The differences in the hybridisation patterns of *Xba*I-digested genomic DNA are unlikely to have arisen from incomplete digestion in this study as the bands are distinct and there is no evidence of smaller digestion products in the lane containing *Xba*I-digested genomic DNA (Figure 4.18).

The YAP057 probe hybridised to single fragments of sizes approximately 4.2 and 4.3 kb generated by digestion of *Arabidopsis* genomic DNA with *Xho*I and *Kpn*I respectively. Hybridisation patterns of the gene encoding P5CR to *Kpn*I- and *Xho*I-digested genomic DNA were not investigated by Verbruggen *et al.* (1993).

Establishing whether or not isozymes of P5CR exist in plants is of considerable importance and remains a contentious issue. Identification of multiple copies of genes encoding P5CR in a single genome might have important implications for the regulation of proline biosynthesis under normal conditions and during stress. In much the same way that it has been proposed that different forms of P5CR serve different metabolic roles in different animal tissues (Phang, 1985), so it seems possible that different P5CRs may exist in plants. One isozyme (possibly inhibited by proline) might be involved in basic housekeeping functions of the cell such as providing proline needed for protein and cell wall biosynthesis. Another isozyme, for which proline inhibition is considerably reduced or absent, could function in proline biosynthesis under conditions of stress. Therefore, the indication that reduction of P5C to proline is catalysed by a single enzyme in *Arabidopsis* is an important contribution towards molecular characterisation of this step in proline biosynthesis.

This finding in *Arabidopsis* contrasts with results of genomic copy number determinations for genes encoding P5CR in soybean (Delauney and Verma, 1990) and pea (Williamson and Slocum, 1992). Results of these workers suggest that in legumes, there may be two to three copies of the gene encoding P5CR. Nevertheless, extensive restriction analysis and sequencing of a number of cDNA clones from both species failed to indicate that different P5CR genes may be expressed in soybean (Delauney and Verma, 1990) and pea (Williamson and Slocum, 1992). The apparent presence of multiple copies of genes encoding P5CRs in these two legumes may arise from gene duplication events that have given rise to pseudogenes or multiple copies of functional genes inserted at different chromosomal locations. Alternatively it may be attributed to the presence of unrelated sequences possessing chance identity to the P5CR gene over short stretches of nucleotides. The larger size of the nuclear genomes of both soybean and pea (Table 2.5) suggests that the possibility of such cross-hybridisation occurring in analysis of legume genomic DNA is greater than that expected during analysis of the *Arabidopsis* genome.

Nevertheless, this interpretation is speculative. Northern analysis of total mRNA from mature pea leaves indicated the presence of two mRNA transcripts encoding P5CR (Williamson and Slocum, 1992). The sizes of these transcripts led these workers to suggest that this was most likely due to the recognition of different polyadenylation signals in pre-mRNA transcripts arising from a single gene encoding P5CR. However, it is possible that these transcripts may arise from transcription of two different genes encoding P5CR in pea. The existence of only a single copy of the gene encoding P5CR in *Arabidopsis* does not eliminate the possibility of there being P5CR isozymes in other plant species. It has been demonstrated that in *Arabidopsis*, many genes which belong to large multigene families in other plant groups, are present as single copies or in very small gene families (Meyerowitz, 1987). Nevertheless, many amino acid biosynthetic enzymes have isoforms even in *Arabidopsis* (Coruzzi, 1991).

The conclusion that a single gene encodes P5CR in *Arabidopsis* adds significance to the results obtained from *in situ* hybridisation studies (Figures 4.19 to 4.22). The existence of only a single copy of the gene encoding P5CR in *Arabidopsis* eliminates the possibility of there being isoforms of P5CR in this species. This therefore eliminates the possibility that different genes encoding P5CR may exhibit different tissue-specific patterns of transcript accumulation to those observed in Figures 4.19 and 4.20. It can thus be concluded that the results of these *in situ* hybridisations represent hybridisation to transcript encoding the only functional P5CR gene in *Arabidopsis*.

As outlined in Section 2.4.1, current understanding of plant gene regulation implicates the involvement of both *cis*-acting DNA elements and *trans*-acting factors. As more data accumulates from animal systems, which are presently better characterised than the equivalent plant systems, the emerging picture suggests that common sets of cell- or tissue-specific *cis*- or *trans*-acting factors provide the basis for tissue-specific expression in plants. Although this study did not involve analysis of possible mechanisms whereby tissue-specific expression of *Arabidopsis* P5CR may be regulated at the molecular level, the results of *in situ* Northern hybridisation studies (Figures 4.19 and 4.20) suggest a complex mechanism of regulation of P5CR expression. Furthermore, proline levels are much higher in the flowers, siliques and seeds of *Arabidopsis* than in the leaves, stems and roots (Figure 4.2). This suggests differential expression of genes involved in proline metabolism in these structures. Verbruggen *et al.* (1993) have demonstrated that the gene encoding P5CR in

Arabidopsis is differentially expressed in organs of mature plants. Levels of mRNA transcript encoding P5CR are higher in roots than in flowers. Stems and leaves possess approximately equal levels of P5CR transcript, which is less abundant in these organs than in flowers. Ripening seeds possess greater levels of P5CR transcript than any of these organs (Verbruggen *et al.*, 1993).

All previous work on transcriptional regulation of genes encoding proline biosynthetic enzymes from plants has involved studying the effect of salinity stress (Delauney and Verma, 1990; Hu *et al.*, 1992; Williamson and Slocum, 1992; Verbruggen *et al.*, 1993). Although this provides an absolute and highly reproducible method of inducing stress, salinisation introduces a number of complications into the induction of hyperosmotic stress. The effects of salinity on plant processes are at least three-fold. Salinity not only induces water stress as a result of the osmotic effect of high NaCl concentrations, but ionic effects of NaCl include mineral toxicity and interruptions to the mineral nutrition of the plant (Chapin, 1991b). The multipartite nature of salinity stress confers difficulty in establishing the level at which the stress is operating. This is particularly critical in the case of proline accumulation as it appears to be a general response to stress (Section 2.1.3). The effects elicited by salinisation may therefore be the result of either osmotic stress or the signalling of a mineral deficiency.

For studies involving the *in situ* hybridisation of YAP057 probe to mRNA transcript encoding P5CR, plants were water-stressed by drying of the rooting medium in preference to other methods, since this reflects more accurately the natural imposition of water deprivation stress. The results presented in Figure 4.19 and Figure 4.20 represent the first report of an accumulation of P5CR mRNA transcript in plants induced by water deprivation by drying rather than by salinisation.

The water status of stem sections from stressed and non-stressed *Arabidopsis* plants was assessed by determination of their relative water contents (RWCs) in preference to determination of water potentials. As with many plant metabolic processes (Sinclair and Ludlow, 1985), proline accumulation is poorly related to plant water potential (Handa *et al.*, 1986; Naidu *et al.*, 1992). These and other workers (Argandona and Pahlich, 1991) have found that the RWC of tissue is a far better parameter for examining proline metabolism in relation to the water status of plants.

The accumulation of the transcript encoding P5CR in response to water deprivation is consistent with the observed increases in P5CR transcripts induced in salt-stressed soybean (Delauney and Verma, 1990), pea (Williamson and Slocum, 1992) and *Arabidopsis* (Verbruggen *et al.*, 1993). All of these workers concluded that transcription of the gene encoding P5CR in plants is osmoregulated. However, as pointed out by Csonka and Hanson (1991), the observation that osmotic shifts change the rate of transcription of a particular gene is not sufficient evidence to conclude that the gene itself is subject to osmotic control. Transcription of P5CR may depend on some other aspect of cell physiology affected indirectly by the osmolarity. Furthermore, induction of expression of the gene encoding P5CR in response to hyperosmotic stress is not in itself sufficient evidence to conclude that the gene product confers tolerance to the stress imposed.

The results shown in Figure 4.19 indicate that under non-stressed conditions, mRNA transcript encoding P5CR is present in the cortex, vascular cambium and pith parenchyma abutting the protoxylem. Transcript encoding P5CR is present at a low level in the phloem (Figure 4.19). An overall increase in the intensity of the signal generated in all of these tissues is evident in stem sections from water-stressed plants (Figure 4.20). Although the presence of transcript encoding P5CR is not evident in the epidermis of non-stressed stems (Figure 4.19A), a small increase in levels of transcript encoding P5CR is evident in the epidermis of plants subjected to eight days of water deprivation (Figure 4.20A). The slightly more dramatic increases in overall signal generated during the first experiment (Figures 4.19A and 4.20A) in comparison with those observed during the replication of the experiment (Figures 4.19B and 4.20B) may be related to the fact that the differences in the RWCs between stem sections from stressed and non-stressed plants were greater in the first experiment than in the subsequent replication of the *in situ* hybridisation (Section 4.12).

The results presented in Figure 4.19 and Figure 4.20 do not demonstrate changes in transcription of the gene encoding P5CR *per se*, but only differences in abundance of the P5CR transcript in these tissues. No indication of the stability of transcripts encoding P5CR has been obtained.

Like most amino acid biosynthetic genes in plants (Matthews *et al.*, 1988), the *Arabidopsis* gene encoding P5CR is unlikely to be highly expressed. As has already been discussed, the absence of strong codon usage bias in the P5CR gene also suggests only a low to moderate

level of transcription of the gene *in vivo*. Nevertheless, as with most hybridisation techniques, under optimal conditions *in situ* hybridisation is exquisitely sensitive. Detection of RNAs present at only 0.25 - 0.5 $kb.\mu m^{-3}$ (5-10 molecules per 20 μm^3 cell) has been reported (Angerer and Angerer, 1992). This corresponds with a concentration of mRNA approximately 0.02% of that of total message uniformly distributed in the tissue.

The absence of any detectable signal following hybridisation to RNase-treated stem sections (Figure 4.21) is a valid indication that the tissue-specific signal generated in the stressed and non-stressed specimens (Figure 4.19 and Figure 4.20) represents hybridisation to mRNA transcripts. Endogenous peroxidases did not make a substantial contribution to the signal generated in both stressed and non-stressed sections as pre-treatment with hydrogen peroxide in methanol eliminated most of the endogenous peroxidase activity within the tissue (Figure 4.22). Furthermore, since the blossoms and leaves of *Digitalis* plants are the only natural source of digoxigenin (Boehringer Mannheim Nonradioactive *In Situ* Hybridisation Manual, 1992), this eliminates the possibility that non-specific binding of anti-DIG antibody to *Arabidopsis* stem sections may have occurred prior to the detection step (Section 3.15.4).

Gene activity may be regulated at several different levels. Some possibilities include control of initiation, elongation and termination of transcription, DNA methylation, regulation of mRNA stability, post-transcriptional modification such as transcript splicing or alteration of mRNA secondary structure as well as control of translational efficiency (Sullivan and Green, 1993). Further possibilities at the post-translational level are regulation of protein stability and protein modification, such as phosphorylation (Ranjeva and Boudet, 1987) or glycosylation (Hubbard and Ivatt, 1981). Therefore, it would be premature to conclude on the basis of these results alone that P5CR activity is necessarily higher in the tissues mentioned above.

However, searches for potential phosphorylation and N-glycosylation sites in the sequences of plant P5CRs sequenced to date (Delauney and Verma, 1990; Williamson and Slocum, 1992; Verbruggen *et al.*, 1993) failed to implicate regulation of P5CR via glycosylation or phosphorylation. Post-translational regulation of P5CR activity via these mechanisms is therefore unlikely. The possibility of regulation of P5CR activity at the transcriptional level could be investigated by observing whether water stress induces any change in transcription rate. This may be measured using *in vitro* nuclear run-on assays and assessment of transcript

accumulation using Northern analysis. Nuclear run-on experiments might indicate whether or not regulation of P5CR expression by water deprivation operates at the level of altering stability of the corresponding mRNA transcript.

Nevertheless, results of these *in situ* hybridisation studies are consistent with much of the data available in literature concerning proline accumulation in plants. For example, the high level of mRNA transcript encoding P5CR in the cortical parenchyma is consistent with the photosynthetic role of this tissue in the stem. Dependence of proline accumulation on light has been reported by several workers and a role for photosynthesis in the process has been suggested (Joyce *et al.*, 1992).

The high level of P5CR transcript in phloem tissue in response to water deprivation (Figure 4.20) may be related to the findings of Heineke *et al.* (1992). While studying the consequence of the expression of a chimeric yeast acid invertase gene in the apoplast of potatoes, these workers demonstrated that when phloem loading of sucrose is blocked, proline accumulates to very high levels. Cress and Johnson (1987) and more recently Larher *et al.* (1993) have demonstrated a strong correlation between levels of non-structural carbohydrates such as sucrose, and induction of proline accumulation. High sucrose levels in phloem may be responsible for constitutively elevated levels of proline synthesis in this tissue. The well-documented increase in sucrose synthesis in response to osmotic stress (Zrenner and Stitt, 1991) provides further evidence in favour of a role for the accumulation of non-structural carbohydrates in the stimulation of proline biosynthesis. Whether this effect is mediated through the supply of carbon skeletons for proline biosynthesis or via an unspecified indirect effect is currently unknown. However, if the enhanced level of transcription of the P5CR gene in phloem tissue in response to water deprivation (Figure 4.20) is related to sucrose accumulation, it raises the possibility that sucrose accumulation in phloem may represent an osmotic stress at the tissue-level to which transcription of the P5CR gene is responsive.

Alternatively, P5CR activity in the phloem may be important in translocation of proline during water stress. As outlined in Section 2.1.2.4, an adaptive role for proline acting as a sink for reduced nitrogen and carbon has been suggested by several workers (Tully *et al.*, 1979; Dashek and Erickson, 1981; Fukutaku and Yamada, 1984; Ahmad and Hellebust,1988). Tully *et al.* (1979) demonstrated that water stressed barley seedlings

removed approximately 25% of the total reduced leaf blade nitrogen by translocation through the phloem. This nitrogen removal coincided with an active translocation of [^{14}C]-labelled photosynthate. In a subsequent paper (Hanson and Tully, 1979), the same workers reported that proline carried approximately 10% of the nitrogen exported per day from wilted barley leaves. This value was obtained through calculation of mass transfer rates of exported [^{14}C]-glutamate, [^{14}C]-glutamine and [^{14}C]-proline following the metabolism of [^{14}C]-glutamate which had been administered to wilted barley leaves. Both [^{14}C]-glutamate and [^{14}C]-proline entered the phloem following their administration to either turgid or wilted blades. Both tracers were translocated at velocities similar to those for carbon dioxide assimilates. These findings suggest that translocation of free proline through phloem elements may be an important aspect of the adaptive role played by proline during hyperosmotic stress.

The observation of high levels of P5CR transcript in the phloem of water-stressed stems suggests the possibility that a promoter element conferring expression in the phloem may exist in the upstream region of the gene encoding P5CR in *Arabidopsis*. This possibility could be investigated by fusion of the 5' end of the gene to a suitable reporter gene such as the *E. coli uidA* gene encoding β-glucuronidase (GUS; Jefferson *et al.*, 1986) and expression of the construct in transgenic plants. Using this approach, Edwards *et al.* (1990) have demonstrated that the promoter of the gene encoding a cytosolic isoform of glutamine synthase in pea is active exclusively in the phloem elements of all organs in mature transgenic tobacco plants. Like proline, glutamine is also exported from wilted barley leaves via the phloem (Tully *et al.*, 1979).

High levels of transcript encoding P5CR in the vascular cambium are consistent with the abundance of soluble proline in meristematic tissues (Durzan and Steward, 1963; Dashek and Erickson, 1981). This is also in keeping with the speculation (Section 2.1.2.4) that proline biosynthesis might be important in stimulating nucleotide biosynthesis needed for rapid proliferation of cells. Meristematic tissues are characterised by high rates of cell division.

No signal was generated in the epidermis of non-stressed plants (Figure 4.19A). A slight increase in levels of P5CR transcript is apparent in the epidermis of plants deprived of water (Figure 4.20A). However, the results do not suggest that a significant level of proline

biosynthesis occurs in this tissue. Klein and Itai (1989) found no proline accumulation in the epidermis of stressed *Commelina communis*. In contrast, accumulation of proline in the epidermis of water stressed barley leaves has been reported by Argandona and Pahlich (1991). However, it is not known whether or not this increase is a consequence of translocation of the imino acid from photosynthetic tissue (Argandona and Pahlich, 1991).

The dramatic increase in P5CR transcript levels in the vicinity of the protoxylem in response to water deprivation (Figure 4.20) may be part of an adaptive mechanism to localised water deficit. Under conditions of reduced water availability, water may be withdrawn from adjacent living tissues as a result of the reduction in xylem pressure potential (Jarvis, 1981). As suggested by Diamantoglou and Rhizopoulou (1992), proline accumulation in neighbouring tissues may lower their solute potential and thereby diminish their requirement for water. Hanson *et al.* (1977) demonstrated that the increase in free proline at any point in a wilted leaf tracks the decline in water status at that point.

Alternatively, synthesis of proline in the vicinity of the protoxylem may serve a developmental role. A stimulatory effect of auxin-induced xylogenesis by proline in stem explants of *Coleus* was reported by Roberts and Baba (1968). Furthermore, a high level of proline biosynthesis in *Nicotiana* callus grown on media supplemented with 15% PEG has been correlated with stimulated xylogenesis compared with unstressed controls (Bornman and Huber, 1979).

Although a number of workers (Delauney and Verma, 1990; Hu *et al.*, 1992; Williamson and Slocum, 1992; Verbruggen *et al.*, 1993) have examined relative levels of mRNA transcripts encoding proline biosynthetic enzymes in different plant organs, the tissue-specificity of the response has never been investigated. These results therefore represent the first report of a tissue-specific pattern of accumulation of a proline biosynthetic enzyme.

The observation that hyperosmotic stress can induce proline accumulation in callus cultures of sorghum (Bhaskaran *et al.*, 1985), *Brassica napus* (Chandler and Thorpe, 1987a, 1987b), rice (Kishor, 1988), alfalfa (Shah *et al.*, 1990), pearl millet (Das *et al.*, 1990), sugar beet (Le Dily *et al.*, 1991) and tobacco (Eberhardt and Wegmann, 1979; Binzel *et al.*, 1987) as well as in suspension culture cells of tomato (Handa *et al.*, 1983, 1986; Rhodes *et al.*, 1986), *Distichlis spicata* (Ketchum *et al.*, 1991) and *Mesembryanthemum crystallinum*

(Adams *et al.*, 1992) suggests that proline accumulation is strictly a cellular response to stress. In contrast, a requirement for tissue differentiation has been suggested for proline accumulation in response to water-deprivation stress in maize (Ibarra-Caballero *et al.*, 1988). These workers failed to correlate an increase in proline levels in drought-stressed maize seedlings with no increase in proline levels of callus cultures subjected to a similar loss of water. This led Ibarra-Cabellero *et al.* (1988) to conclude that proline accumulation in response to drought stress is dependent on systemic organisation of the plant. In support of this suggestion that communication between different tissues is required to elicit proline accumulation, Voetberg and Sharp (1991) have reported a differential accumulation of proline towards the apex of the maize primary root.

The results of the *in situ* hybridisation studies presented in Figures 4.19 and 4.20 provide evidence in favour of both of the arguments that proline accumulation is either strictly a cellular response (Adams *et al.*, 1992) or that it is dependent on tissue-organisation (Ibarra-Cabellero *et al.*, 1988). The striking tissue-specificity of the hybridisation pattern presented in Figure 4.20 supports the latter hypothesis. However, the possibility that cellular water status might directly affect transcript accumulation in individual cells cannot be excluded. In particular, as has been discussed already, the high levels of P5CR transcript in phloem and in the vicinity of protoxylem elements may be responses to osmotic stress at the tissue level.

Presently little is understood concerning the differences in genetic regulation between plant cells in culture and those present in intact plants. Stress responses characteristic of the facultative halophyte *Mesembryanthemum crystallinum*, such as induction of Crassulacean acid metabolism and pinitol accumulation, are not elicited when suspension cells of this species are subjected to salinity stress (Adams *et al.*, 1992). This suggests that cells in culture mimic only partly the stress response mechanisms of intact plants.

Tissue-specificity of proline accumulation implies that *in vitro* mechanisms to select for salt tolerant lines on the basis of their ability to accumulate proline are not feasible as cell cultures may not reflect the whole plant situation. In a field situation, the magnitude of water stress does not remain constant as progressive dehydration of the soil occurs. The metabolic and adaptive responses of different organs and the tissues within these organs may vary greatly. Although the use of cultured tissues may circumvent such problems, the

behaviour of excised tissues or cell cultures may not be consistent with the behaviour of the whole plants from which they were derived. In the light of this observation, it is interesting to note that many workers (Tal *et al.*, 1978; Bhaskaran *et al.*, 1985; Chandler and Thorpe, 1987a, 1987b; Hassan and Wilkins, 1988; Rodriguez and Heyser, 1988) who have discredited the value of proline accumulation as an adaptive mechanism to hyperosmotic stress have used cell cultures as experimental systems. Examination of the tissue-specific accumulation of proline in plant parts other than stems (this work) and root tips (Voetberg and Sharp, 1991; Ober and Sharp, 1994) may provide further evidence that proline accumulation might be dependent on the systemic organisation of the plant.

At present it is not possible to unequivocally correlate the elevated levels of P5CR transcript that accompany water stress directly with elevated P5CR activity and proline content. Furthermore, although changes at the mRNA level are often thought to reflect changes in the level of transcription, this may not be the case. As already mentioned, expression of P5CR may be regulated at either a post-transcriptional or post-translational level.

Some evidence exists for post-translational regulation of P5CR activity. Argandona and Pahlich (1991) failed to correlate P5CR activity with the amount of enzyme present in barley leaves. These workers used immunoblotting to demonstrate that a five- to eight-fold increase in levels of P5CR activity accompanying water deprivation stress did not result from increased protein content of P5CR (Argandona and Pahlich, 1991). The mechanism behind this activation remains to be clarified. In *Arabidopsis*, a greater abundance of P5CR transcript in roots than in leaves (Verbruggen *et al.*, 1993) does not correlate with higher free proline levels in leaves in comparison to roots (Verbruggen *et al.*, 1993; Figure 4.2). Furthermore, Szoke *et al.* (1992) demonstrated an effect of NaCl on the activity profile of soybean P5CR with changes in pH (Section 2.2.1.2). Whether this finding is of any significance *in vivo* is currently not known. LaRosa *et al.* (1991) failed to demonstrate any differences in the activities or kinetic properties of P5CR isolated from salt-adapted and non-adapted tobacco cells.

Other workers have presented evidence implicating post-transcriptional regulation of P5CR activity in plants. Although Williamson and Slocum (1992) did not measure P5CR activities prior to and during salinisation of pea seedlings, immunoblot analysis failed to indicate that P5CR transcript levels paralleled levels of the enzyme. However, proline levels did increase

together with transcript accumulation (Williamson and Slocum, 1992). This was interpreted to indicate that possibly salt stress simply induces synthesis and/or stabilisation of mRNA transcript encoding P5CR without a corresponding increase in the level of the gene product (Williamson and Slocum, 1992). Furthermore, Verbruggen *et al.* (1993) demonstrated that in ten-day-old *Arabidopsis* seedlings, addition of the translational inhibitor cordycepin inhibited proline accumulation when added 12 h after imposition of salt stress, whereas addition of the transcriptional inhibitor cycloheximide did not.

There is some evidence that other stress-related proteins in plants share a pattern of regulation analogous to that observed with P5CR. For example, expression of the mRNA encoding osmotin in tobacco is strongly induced by ABA, although there is no effect of ABA treatment on osmotin protein (LaRosa *et al.*, 1992). This suggests that as with water stress and P5CR expression, ABA exerts at least two different effects on osmotin gene expression; one operating on mRNA levels, and the other affecting either translation or stability of osmotin.

The exponential increase in our understanding of plant gene regulation is likely to lead to the identification of both specific regulatory DNA sequences responsible for stress-inducible and tissue-specific expression and the *trans*-acting protein factors that bind to these elements. To identify these regions and the proteins that bind them, the techniques of DNaseI footprinting (Green *et al.*, 1989) and gel mobility shift assays (Garner and Revzin, 1981; Fried and Crothers, 1981; Foster *et al.*, 1990) will be important.

It is likely that in order to fully understand the sequence of events responsible for the transcriptional activation of the gene encoding P5CR, it will be necessary to establish whether *trans*-acting factors interact with specific DNA sequences upstream of the open reading frame. In the event of such transcriptional factors being identified, investigation of the regulation of their expression is likely to be necessary in order to fully understand the osmotically-sensitive induction of P5CR gene expression and the tissue-specific localisation of P5CR transcript. Thus, the genes encoding such putative *trans*-acting factors will need to be isolated, and their expression and functional regulation examined. It is hoped that use of such data or the fusion of a reporter gene such as the GUS gene (Jefferson *et al.*, 1986) to upstream regions of the P5CR gene may confirm the findings of this study concerning the tissue-specific and osmoregulated expression of the gene encoding P5CR.

An important goal of future studies is likely to be overproduction of proline in agriculturally-significant crops via recombinant DNA technology. This may increase their environmental tolerance and productivity. The spatial distribution of the proline accumulated in such transgenic plants may be an important consideration in ensuring their efficacy under field conditions. The identification of *cis*-acting DNA elements and *trans*-acting transcriptional factors controlling the osmoregulation and tissue-specific expression of P5CR may also be critical in facilitating such an endeavour.

6. Conclusion

Presently, there is no unified model available for the role of proline accumulation in plants exposed to environmental stress (Section 2.1.2). With regard to hyperosmotic stress, proline accumulation most likely represents a mechanism of dehydration avoidance. Accumulated proline most probably enables the plant to maintain sufficient tissue hydration for appropriate functioning of the metabolic processes involved in growth and development under conditions of water stress.

Sufficient data is currently not available to conclude that the correlation between proline accumulation and development of stress is proof that the imino acid has some adaptive value in postponing stress or increasing stress tolerance. Many critics argue that any beneficial effects from this stress-induced disturbance of nitrogen metabolism are coincidental. Nevertheless, this viewpoint seems unlikely since if proline accumulation served no function, natural selection would undoubtably have operated against conservation of such an energy-intensive process.

The use of recombinant DNA techniques in studies on proline metabolism in plants is likely to provide some long overdue answers to a number of questions concerning the value of proline accumulation as a stress response. Already, the demonstration that in plants a bifunctional enzyme catalyses the rate-limiting step in proline biosynthesis (Hu *et al.*, 1992) represents a major step in the understanding of the system. However, the elucidation of the complete nucleotide sequence of a genomic clone of plant Δ^1-pyrroline-5-carboxylate synthetase (P5CS) remains a requirement for the functional analysis of elements involved in expression of this gene.

The potential of transgenic plant systems in studying the physiological effects of proline overproduction or underproduction cannot be overemphasised. In this respect, it is particularly encouraging to note that elevated levels of another organic osmolyte mannitol in transgenic tobacco plants expressing the bacterial gene encoding mannitol-1-phosphate dehydrogenase, enhances osmotolerance (Tarczynski *et al.*, 1993).

If the engineering of proline biosynthesis in a model system such as *Arabidopsis* is

successful in increasing osmotolerance, the same approach may be applicable to important crops such as maize, soybean, wheat and rice. It might also be feasible to transform a proline-overproducer with additional genes implicated in the determination of osmotolerance in order to augment osmoregulation. For example, Hanson *et al.* (1994) have recently shown that members of the Plumbaginaceae subjected to particularly harsh osmotic stress have evolved the ability to convert their chronically large pool of proline into proline betaine and hydroxyproline betaine. In bacterial osmoprotection bioassays, these two osmolytes are more effective than proline (Hanson *et al.*, 1994). This raises the possibility of enhancing the effectiveness of proline even further by engineering its conversion to betaines during severe stress. The enzymes and genes involved in biosynthesis of glycine betaine have already been isolated (Brouquisse *et al.*, 1989; Weretilnyk and Hanson, 1990). The list of potential candidates for transfer into plants is likely to increase. A novel methyl transferase involved in pinitol production in *Mesembryanthemum crystallinum* has been identified by Vernon and Bohnert (1992). Manipulation of the expression of the corresponding gene in transgenic plants is likely to contribute to our overall appreciation of the value of osmolytes. Overexpression of the yeast HAL1 protein, which has homologues in plants, has been shown to overcome Na^+ toxicity in yeast (Gaxiola *et al.*, 1992). This may well be another candidate for transfer into plants in the near future.

Since agricultural inputs are becoming more costly and scarce, plants possessing genetic adaptations for improved performance in the field are likely to be readily accepted by the public. Furthermore, the accumulation of proline in response to a wide spectrum of stresses suggests that modification of proline metabolism is an ideal target for increasing the overall stress tolerance of plants. Studies in prokaryotes suggest that in bacteria, osmotic tolerance is not necessarily dependent on the interaction of a large array of gene products, but may simply be reduced to a simple phenomenon such as accumulation of a compatible solute (Csonka, 1989; Csonka and Hanson, 1991).

However, it would be rash to ignore the complexity of the response of the plant genome to the full range of environmental stresses plants may experience. Transfer or manipulation of a single gene or even upregulation of an entire biosynthetic pathway is unlikely to be sufficient to convert a stress-susceptible genotype into one tolerant of the same stress(es). Nevertheless, manipulation of the proline biosynthetic pathways using the techniques of molecular biology may provide valuable insights into the important roles these pathways

may play in acclimation to a range of environmental stresses. For example, inhibiting proline biosynthesis by partially blocking activity of proline biosynthetic enzymes with antisense constructs (Bird and Ray, 1991) under the control of stress-inducible promoters (Nelson *et al.*, 1992) may furnish conclusive evidence of the adaptive significance of proline accumulation in increasing resistance to stress. A reduction in the level of proline degradation by proline oxidase using an antisense approach may cause elevated levels of proline accumulation in stressed plants. Elevated synthesis of P5CS in transgenic plants is unlikely to cause proline overproduction since the enzyme from *Vigna aconitifolia* exhibits almost complete inhibition *in vitro* in the presence of 10 mM proline (Hu *et al.*, 1992). Proline accumulates up to 129 mM in salt-adapted tobacco cells (Binzel *et al.*, 1987). Modification of this enzyme, possibly by site-directed mutagenesis may cause a loss in its feedback inhibition by proline (Verma *et al.*, 1992) and may therefore be a prerequisite for increased expression of P5CS activity. Alternatively, the mechanisms mediating the loss in feedback inhibition of P5C synthesis during stress will need to be elucidated.

Unlike pest or herbicide resistance, the tolerance trait of a plant to drought or salinity is determined by multiple genes located in a large number of loci and usually on different chromosomes. Drought tolerance should therefore not be considered as a unique heritable trait, but perhaps as a complex of often unrelated plant properties (Visser, 1994). To date, the complexity of tolerance mechanisms has limited the application of genetic engineering in developing transgenic stress-tolerant crops. Furthermore, a stress is seldom experienced in isolation. Therefore ideally, transgenic plants should exhibit tolerance to several of the abiotic stresses they encounter in the field. The induction of proline accumulation in response to many stresses commonly encountered in natural environments (Section 2.1.3) suggests that overproduction of proline in transgenic plants may increase their overall tolerance to biotic and abiotic stresses.

In spite of the explosive development of biotechnology over the past two decades, its use for obtaining new abiotic stress-tolerant varieties has been very limited. The lack of sufficient information on the molecular biology of plants is certainly the greatest constraint to the effective use of biotechnology for increasing stress tolerance. For example, even though transgenic tobacco capable of accumulating mannitol exhibited increased salt tolerance, growth under stress was severely retarded (Tarczynski *et al.*, 1993).

Abiotic stress tolerance is a complex physiological response resulting from expression of many genes either simultaneously or in succession. This multigenic character of stress resistance imposes limits to genetic manipulation. Most importantly, very little is known about the molecular basis of important physiological and biochemical responses and characteristics of plants. In this respect, further characterisation of proline biosynthesis and degradation in *Arabidopsis* is likely to contribute substantially to the establishment of this plant species as a research tool in elucidating stress mechanisms in all plants.

One prerequisite for the manipulation of proline biosynthetic rates alone is complete characterisation of the relative contributions of the pathways from glutamate and ornithine under different physiological conditions and during different developmental stages.

Clearly, cloning and characterisation of the genes involved in proline metabolism is unlikely to answer many of the questions about the functionality of proline accumulation under stress. However, comparison of the amino acid sequences of a range of P5CR enzymes in this study has indicated conserved features of all P5CRs that are likely to be of importance in functioning of the *Arabidopsis* enzyme. These may be of value in future efforts to manipulate P5CR activity at the molecular level. Integration of this information with the results of studies of the physiology and anatomical location of proline accumulation is likely to go some way to answering an enigma that has intrigued plant physiologists over the past forty years.

LITERATURE CITED

Adams, E. and Goldstone, A. (1960) Hydroxyproline metabolism. III. Enzymatic synthesis of hydroxyproline from Δ^1-pyrroline-3-hydroxy-5-carboxylate. *J. Biol. Chem. 235* : 3499-3503

Adams, E. and Frank, L. (1980) Metabolism of proline and the hydroxyprolines. *Annu. Rev. Biochem. 49* : 1005-1061

Adams, M.D., Kelley, J.M., Gocayne, J.D., Dubnick, M., Polymeropoulous, M.H., Xiao, H., Merrill, C.R., Wu, A., Olde, B., Moreno, R.F., Kerlavage, A.R., McCombie, W.R. and Venter, J.C. (1991) Complementary DNA sequening : Expressed sequence tags and human genome project. *Science 252* : 1651-1656

Adams, M.D., Dubnick, M., Kerlavage, A.R., Moreno, R., Kelley, J.M., Utterback, T.R., Nagle, J.W., Fields, C. and Venter, J.C. (1992) Sequence identification of 2375 human brain genes. *Nature 355* : 632-634

Adams, M.D., Kerlavage, A.R., Fields, C. and Venter, J.C. (1993) 3400 new expressed sequence tags identify diversity of transcripts in human brain. *Nature Genet. 4* : 256-267

Adams, P., Thomas, J.C., Vernon, D.M., Bohnert, H.J. and Jensen, R.G. (1992) Distinct cellular and organismic responses to salt stress. *Plant Cell Physiol. 33* : 1215-1223

Ahmad, I and Hellebust, J.A. (1984) Osmoregulation in the extremely euryhaline marine microalga *Chlorella autotrophica. Plant Physiol. 74* : 1010-1015

Ahmad, I. and Hellebust, J.A. (1988) The relationship between inorganic nitrogen metabolism and proline accumulation in osmoregulatory responses of two euryhaline microalgae. *Plant Physiol. 88* : 348-355

Ahmad, I., Larher, F. and Stewart, C.R. (1979) Sorbitol, a compatible osmotic solute in *Plantago maritima. New Phytol. 82* : 671-678

Ahn, K.S. and Wake, R.G. (1991) Variations and coding features of the sequence spanning the replication terminus of *Bacillus subtilis* 168 and W23 chromosomes. *Gene 98* : 107-112

Albrecht, A.M. and Vogel, H.J. (1964) Acetylornithine δ-transaminase : Partial purification and repression behaviour. *J.Biol. Chem. 239* : 1872-1875

Alia, Saradhi, P.P and Mohanty, P. (1991) Proline enhances primary photochemical activities in isolated thylakoid membranes of *Brassica juncea* by arresting photoinhibitory damage. *Biochem. Biophys. Res. Comm. 181* : 1238-1244

Allen, S. and Raven, J.A. (1987) Intracellular pH regulation in *Ricinus communis* grown with ammonium or nitrate as N source : The role of long distance transport. *J. Exp. Bot. 38* : 580-596

Aloni, B. and Rosenshtein, G. (1982) Effect of flooding on tomato cultivars : The relationship between proline accumulation and other morphological and physiological changes. *Physiol. Plant. 56* : 513-517

Al-Sulaiti, A., Lea, P.J. and Davies, W.J. (1990) Effects of soil drying on morphogenetic and physiological responses of barley photorespiratory mutants. *Br. Soc. Plant Growth Reg. Monograph 21* : 347-350

Altschul, S.F., Boguski, M.S., Gish, W. and Wootton, J.C. (1994) Issues in searching molecular sequence databases. *Nature Genet. 6* : 119-129

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) Basic local alignment search tool. *J. Mol. Biol. 215* : 403-410

An, G., Mitra, A., Choi, H.K., Costa, M.A., An, K., Thornburg, R.W. and Ryan, C.A. (1989) Functional analysis of the 3' control region of the potato wound-inducible proteinase inhibitor II gene. *Plant Cell 1* : 115-122

Anbazhagan, M., Krishnamurthy, R. and Bhagwat, K.A. (1988) Proline : An enigmatic indicator of air pollution tolerance in rice cultivars. *J. Plant Physiol. 133* : 122-123

Anderson, M. (1994) The Nottingham *Arabidopsis* Stock Centre Seed List, June 1994. Nottingham *Arabidopsis* Stock Centre, Nottingham, pp. 2-23, 55-83, 85

Anderson, C.B. and Witter, L.D. (1982) Glutamine and proline accumulation by *Staphylopcoccus aureus* with reduction in water activity. *Appl. Environ. Microbiol. 43* : 1501-1503

Andreadis, A., Gallego, M.E. and Nadal-Ginard, B. (1987) Generation of protein isoform diversity by alternative splicing : Mechanistic and biological implications. *Annu. Rev. Cell Biol. 3* : 207-242

Anfinsen, C.B. (1973) Principles that govern the folding of protein chains. *Science 181* : 223-230

Anfinsen, C.B., Haber, E., Sela, M. and White, F.H. (1961) The kinetics of formation of the native ribonuclease during oxidation of the reduced polypeptide chain. *Proc. Natl. Acad. Sci. USA* **47** : 1309-1314

Angerer, L.M. and Angerer, R.C. (1992) *In situ* hybridisation to cellular RNA with radiolabelled RNA probes. *In* D.G. Wilkinson (ed) *In situ* Hybridisation : A Practical approach. IRL Press, Oxford, pp. 15-32

Apostol, I., Heinstein, P.F. and Low, P.S. (1989) Rapid stimulation of an oxidative burst during elicitation of cultured plant cells : role in defense and signal transduction. *Plant Physiol.* **90** : 109-116

Arakawa, T. and Timasheff, S.N. (1983) Preferential interactions of proteins with solvent components in aqueous amino acid solutions. *Arch. Biochem. Biophys.* **244** : 169-177

Arakawa, T. and Timasheff, S.N. (1985) The stabilisation of proteins by osmolytes. *Biophys. J.* **47** : 411-414

Argandona, V. and Pahlich, E. (1991) Water stress on proline content and enzyme activities in barley seedlings. *Phytochemistry* **30** : 1093-1094

Argos, P. Rao, J.K.M. and Hargrave, P.A. (1982) Structural prediction of membrane-bound proteins. *Eur. J. Biochem.* **128** : 565-572

Armstrong, C.L. and Green, C.E. (1985) Establishment and maintenance of friable, embryogenic maize callus and the involvement of L-proline. *Planta* **164** : 207-214

Arndt, K. and Fink, G.R. (1986) GCN4 protein, a positive transcription factor in yeast, binds general control promoters at all 5' TGACTC 3' sequences. *Proc. Natl. Acad. Sci. USA* **83** : 8516-8520

Ashraf, M. (1989) The effect of NaCl on water relations, chlorophyll and proline contents of two cultivars of blackgram (*Vigna mungo* L.) *Plant Soil* **119** : 205

Aspinall, D., Singh, T.N. and Paleg, L.G. (1973) Stress metabolism V. Abscisic acid and nitrogen metabolism in barley and *Lolium temulentum* L. *Aust. J. Biol. Sci.* **26** : 319-327

Atkinson, D.E. (1977) Cellular energy metabolism and its regulation. Academic Press, New York, p. 127

Aufsatz, W. and Grimm, C. (1994) A new, pathogen-inducible gene of *Arabidopsis* is expressed in an ecotype-specific manner. *Plant Mol. Biol.* **25** : 229-239

Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A. and Struhl, K. (1987) Current Protocols in Molecular Biology. John Wiley and Sons, New York. pp. 14.3.1 - 14.3.14

Bachmann, B.J. (1990) Linkage Map of *Escherichia coli* K-12, Edition 8. *Microbiol. Rev. 54* : 130-197

Badzinski Buhl, M. and Stewart, C.R. (1983) Effects of NaCl on proline synthesis and utilisation in excised barley leaves. *Plant Physiol. 72* : 664-667

Baich, A. (1969) Proline synthesis in *Escherichia coli* : A proline-inhibitable glutamic acid kinase. *Biochim. Biophys Acta. 192* : 462-467

Baich, A. (1971) The biosynthesis of proline in *Escherichia coli* : Phosphate-dependent glutamate γ-semialdehyde dehydrogenase (NADP), the second enzyme in the pathway. *Biochim. Biophys. Acta. 244* : 129-134

Baich, A. and Pierson, D.J. (1965) Control of proline biosynthesis in *Escherichia coli. Biochim. Biophys. Acta. 104* : 397-404

Baker, S.S., Wilhelm, K.S. and Thomashow, M.F. (1994) The 5'-region of *Arabidopsis thaliana cor15a* has *cis*-acting elements that confer cold-, drought- and ABA-regulated gene expression. *Plant Mol. Biol. 24* : 701-713

Balboni, E. (1978) A proline shuttle in insect flight muscle. *Biochem. Biophys. Res. Commun. 85* : 1090-1096

Barnett, N.M. and Naylor, A.W. (1966) Amino acid and protein metabolism in bermuda grass during water stress. *Plant Physiol. 41* : 1222-1230

Bar-Nun, N. and Poljakoff-Mayber, A. (1977) Salinity stress and the content of proline in roots of *Pisum sativum* and *Tamarix tetragyna. Ann. Bot. 41* : 173-179

Barta, A., Sommergruber, K., Thompson, D., Hartmuth, K., Matzke, M.A. and Matzke, A.J.M. (1986) The expression of a nopaline synthase-human growth hormone chimaeric gene in transformed tobacco and sunflower callus tissue. *Plant Mol. Biol. 6* : 347-357

Bartels, D., Singh, M. and Salamini, F. (1988) Onset of dessication tolerance during development of the barley embryo. *Planta 175* : 485-492

Bascaran, V., Hardisson, C. and Brana, A.F. (1989) Regulation of nitrogen catabolic enzymes in *Streptomyces clavuligerus*. *J. Gen. Microbiol.* ***135*** : 2465-2474

Bassi, R. and Sharma, S.S. (1993) Changes in proline content accompanying the uptake of zinc and copper by *Lemna minor*. *Ann. Bot.* ***72*** : 151-154

Bates, L.S., Waldren, R.P and Teare, I.D. (1973) Rapid determination of free proline for water-stress studies. *Plant and Soil.* ***39*** : 205-207

Beffagna, N., Pesci, P. and Radice, M. (1986) Thioproline-induced accumulation of proline in barley leaf sections. *Plant Cell Environ.* ***9*** : 141-144

Bellinger, Y. and Larher, F. (1987) Proline accumulation in higher plants : A redox buffer? *Plant Physiol. (Life Sci. Adv.)* ***6*** : 23-27

Benfey, P.N. and Chua, N.-H. (1989) Regulated genes in transgenic plants. *Science* ***244*** : 174-181

Benfey, P.N. and Chua, N.-H. (1990) The CaMV 35S promoter and combinatorial regulation of transcription in plants. *Science* ***250*** : 959-966

Benfey, P.N., Ren, L. and Chua, N.-H. (1989) The CaMV 35S enhancer contains at least two domains which can confer different developmental and tissue-specific expression patterns. *EMBO J.* ***8*** : 2195-2202

Benfey, P.N., Ren, L. and Chua, N.-H. (1990a) Tissue-specific expression from CaMV 35S enhancer subdomains in early stages of plant development. *EMBO J.* ***9*** : 1677-1684

Benfey, P.N., Ren, L. and Chua, N.-H. (1990b) Combinatorial and synergistic properties of CaMV 35S enhancer domains. *EMBO J.* ***9*** : 1685-1696

Bengston, C., Klockare, B., Klockare, R, Larsson, S. and Sundqvist, C. (1978) The after-effect of water stress on chlorophyll formation during greening and the levels of abscisic acid and proline in dark grown wheat seedlings. *Physiol. Plant.* ***43*** : 205-212

Bennetzen, J.L. and Hall, B.D. (1982) Codon selection in yeast. *J. Biol. Chem.* ***257*** : 3026-3031

Bhaskaran, S., Smith, R.H. and Newton, R.J. (1985) Physiological changes in cultured sorghum cells in response to induced water stress. I. Free proline. *Plant Physiol.* ***79*** : 266-269

Biggen, M.D., Gibson, T.J. and Hong, G.F. (1983) Buffer gradient gels and ^{35}S label as an aid to rapid DNA sequence determination. *Proc. Natl. Acad. Sci. USA 80* : 3963-3965

Binzel, M.L., Hess, F.D., Bressan, R.A. and Hasegawa, P.M. (1988) Intracellular compartmentation of ions in salt-adapted tobacco cells. *Plant Physiol. 86* : 607-614

Binzel, M.L., Hasegawa, P.M., Rhodes, D., Handa, S., Handa, A.V. and Bressan, R.A. (1987) Solute accumulation in tobacco cells adapted to NaCl. *Plant Physiol. 84* : 1408-1415

Bird, C.R. and Ray, J.A. (1991) Manipulation of plant gene expression by antisense RNA. *Biotech. Genet. Eng. Rev. 9* : 207-227

Birnstiel, M.L., Busslinger, M. and Strub, K. (1985) Transcription termination and 3' processing : The end is in site! *Cell 41* : 349-359

Blake, C.C.F. (1978) Do genes-in-pieces imply proteins-in-pieces? *Nature 273* : 267

Blake, C.C.F. (1979) Exons encode protein functional units. *Nature 277* : 598

Blake, C.C.F. (1985) Exons and the evolution of proteins. *Int. Rev. Cytol. 93* : 149-185

Blake, R.L. (1972) Animal model for hypoprolinaemia : Deficiency of mouse proline oxidase activity. *Biochem. J. 129* : 987-989

Blanchard, M., Green, D.E., Nocito, V. and Ratner, S. (1944) L-amino acid oxidase of animal tissue. *J. Biol. Chem. 155* : 421-440

Bloom, F., Smith, C.J., Jessee, J., Villeux, B. and Deutch, A.H. (1983) The use of genetically engineered strains of *Escherichia coli* for the overproduction of free amino acids : Proline as a model system. *In* Downey, K., Voellmy, R.W., Ahmad, F. and Schultz, J. (eds) Advances in Gene Technology : Molecular Genetics of Plants and Animals. Academic Press, Orlando, Florida, pp. 383-394

Blum, A. and Ebercon, A. (1976) Genotypic responses in soghum to drought stress. III. Free proline accumulation and drought resistance. *Crop Sci. 16* : 428-431

Boggess, S.F. and Stewart, C.R. (1976) Contribution of arginine to proline accumulation in water-stressed barley leaves. *Plant Physiol. 58* : 796-797

Boggess, S.F. and Stewart, C.R. (1980) The realtionship between water stress induced proline accumulation and inhibition of protein synthesis in tobacco leaves. *Plant Sci. Lett. 17* : 245-252

Boggess, S.F., Paleg, L.G. and Aspinall, D. (1975) Δ^1-pyrroline-5-carboxylic acid dehydrogenase in barley, a proline-accumulating species. *Plant Physiol. 56* : 259-262

Boggess, S.F., Aspinall, D. and Paleg, L.G. (1976a) Stress metabolism. IX. The significance of end-product inhibition of proline biosynthesis and of compartmentation in relation to stress-induced proline accumulation. *Aust. J. Plant Physiol. 3* : 513-525

Boggess, S.F., Koeppe, D.E. and Stewart, C.R. (1978) Oxidation of proline by plant mitochondria. *Plant Physiol. 62* : 22-25

Boggess, S.F., Stewart, D. Aspinall, D. and Paleg, L.G. (1976b) Effect of water stress on proline biosynthesis from radioactive precursors. *Plant Physiol. 58* : 398-401

Bogusz, D., Appelby, C.A., Landsmann, J., Dennis, E.S., Trinnick, M.J. and Peacock, W.J. (1988) Functioning haemoglobin genes in non-nodulating plants. *Nature 331* : 178-180

Bone, D.H. (1959) Metabolism of citrulline and ornithine in mung bean mitochondria. *Plant Physiol. 34* : 171-175

Bonham-Smith, P.C., Kapoor, M. and Bewley, J.D. (1988) Exogenous application of abscisic acid or triadimefon affects the recovery of *Zea mays* seedlings from heat shock. *Physiol. Plant. 73* : 27-30

Bornman, C.H. and Huber, W. (1979) *Nicotiana tabacum* callus studies. IX. Development in stressed explants. *Biochem. Physiol. Pflanz. 174* : 345-356

Bornman, C.H. and Jansson, E. (1980) *Nicotiana tabacum* callus studies. X. ABA increases resistance to cold damage. *Physiol. Plant. 48* : 491-493

Boswell, D.R. and Lesk, A.M. (1988) Sequence comparison and alignment : The measurement and interpretation of sequence similarity. *In* Lesk, A.M. (ed) Computational Molecular Biology. Oxford University Press, New York, pp. 166-167

Bowler, C., Van Montagu, M. and Inzé, D. (1992) Superoxide dismutase and stress tolerance. *Annu. Rev. Plant Physiol. Plant Mol. Biol. 43* : 83-116

Bowman, W.D. (1988) Ionic and water relation responses of two populations of a non-halophyte to salinity. *J. Exp. Bot.* ***39*** : 97-105

Boyer, J.S. (1982) Plant productivity and the environment. *Science* ***218*** : 443-448

Brady, R.A. and Csonka, L.N. (1988) Transcriptional regulation of the *proC* gene of *Salmonella typhimurium*. *J. Bacteriol.* ***170*** : 2379-2382

Brady, C.J., Gibson, T.S., Barlow, E.W.R., Speirs, J. and Wyn Jones, R.G. (1984) Salt tolerance in plants. I. Ions, compatible organic solutes and the stability of plant ribosomes. *Plant Cell Environ.* ***7*** : 571-578

Branden, C. and Tooze, J. (1991) Introduction to Protein Structure. Garland, New York, pp.1-10, 11-30, 43-56, 141-158, 201-214, 247-267

Brandriss, M.C. (1979) Isolation and preliminary characterisation of *Saccharomyces cerevisiae* proline auxotrophs. *J. Bacteriol.* ***138*** : 816-822

Brandriss, M.C. (1983) Proline utilisation in *Saccharomyces cerevisiae* : Analysis of the cloned *PUT2* gene. *Mol. Cell. Biol.* ***3*** : 1846-1856

Brandriss, M.C. (1987) Evidence for the positive regulation of the proline utilisation pathway in *Saccharomyces cerevisiae*. *Genetics* ***117*** : 429-435

Brandriss, M.C. and Falvey, D.A. (1992) Proline biosynthesis in *Saccharomyces cerevisiae* : Analysis of the *PRO3* gene encoding Δ^1-pyrroline-5-carboxylate reductase. *J. Bacteriol.* ***174*** : 3782-3788

Brandriss, M.C. and Krwywicki, K.A. (1986) Amino-terminal fragments of Δ^1-pyrroline-5-carboxylate dehydrogenase direct β-galactosidase to the mitochondrial matrix in *Saccharomyces cerevisiae*. *Mol. Cell Biol.* ***6*** : 3502-3512

Brandriss, M.C. and Magasanik, B. (1979a) Genetics and physiology of proline utilisation in *Saccharomyces cerevisiae* : Enzyme induction by proline. *J. Bacteriol.* ***140*** : 498-503

Brandriss, M.C. and Magasanik, B. (1979b) Genetics and physiology of proline utilisation in *Saccharomyces cerevisiae* : Mutation causing constitutive enzyme expression. *J. Bacteriol.* ***140*** : 504-507

Brandriss, M.C. and Magasanik, B. (1980) Proline : An essential intermediate in arginine degradation in *Saccharomyces cerevisiae*. *J. Bacteriol.* ***143*** : 1403-1410

Brandriss, M.C. and Magasanik, B. (1981) Subcellular compartmentation in control of converging pathways for proline and arginine metabolism in *Saccharomyces cerevisiae*. *J. Bacteriol.* ***145*** : 1359-1364

Bray, E.A. (1991) Regulation of gene expression by endogenous ABA during drought stress. *In* Davies, W.J. and Jones, H.G. (eds) Abscisic Acid : Physiology and Biochemistry. Bios Scientific Publishers, Oxford, pp. 81-98

Bray, E.A. (1994) Molecular responses to water deficit. *Plant Physiol.* ***103*** : 1035-1040

Bredt, D.S., Hwang, P.M., Glatt, C.E., Lowenstein, C., Reed, R.R. and Snyder, S.H. (1991) Cloned and expressed nitric oxide synthase structurally resembles cytochrome P-450 reductase. *Nature* ***351*** : 714-718

Breitbart, R.E., Andreadis, A. and Nadal-Ginard, B. (1987) Alternative splicing : A ubiquitous mechanism for the generation of multiple protein isoforms from single genes. *Annu. Rev. Biochem.* ***56*** : 467-495

Briggs, G.E. and Robertson, R.N. (1957) Apparent free space. *Annu. Rev. Plant Physiol.* ***8*** : 11-30

Brouquisse, R., Weigel, P., Rhodes, D., Yocum, C.F. and Hanson, A.D. (1989) Evidence for a ferredoxin-dependent choline monooxygenase from spinach chloroplast stroma. *Plant Physiol.* ***90*** : 322-329

Brown, J. (1986) A catalogue of splice junction and putative branch point sequences from plant introns. *Nucl. Acids Res.* ***14*** : 9549-9559

Brown, A.D. and Simpson, J.R. (1972) Water relations of sugar tolerant yeasts : The role of intracellular polyols. *J. Gen. Microbiol.* ***72*** : 589-591

Brown, L.M. and Hellebust, J.A. (1978) Sorbitol and proline as intracellular osmotic solutes in the green alga *Stichococcus bacillaris*. *Can. J. Bot.* ***56*** : 676-679

Brown, L.M. and Hellebust, J.A. (1980) The contribution of organic solutes to osmotic balance in some green and eustigmatophyte algae. *J. Phycol.* ***16*** : 265-270

Brown, J.W.S., Felix, G. and Frendewey, D. (1986) Accurate *in vitro* splicing of two pre-mRNA plant introns in a HeLa cell nuclear extract. *EMBO J. 5* : 2749-2758

Brugarolas, A. and Gosalvez, M. (1980) Treatment of cancer by an inducer of reverse transformation. *Lancet I* : 65-70

Brunner, G. and Neupert, W. (1969) Localisation of proline oxidase and Δ^1-pyrroline-5-carboxylic acid dehydrogenase in rat liver. *FEBS Lett. 3* : 283-286

Burrows, W.J. and Carr, D.J. (1969) Effects of flooding the root system of sunflowers on the cytokinin content of the root sap. *Physiol. Plant. 22* : 1105-1112

Burton, R.S. (1991) Regulation of proline biosynthesis in osmotic response : Effects of protein synthesis inhibitors. *J. Exp. Zool. 259* : 272-277

Cairney, J., Booth, I.R. and Higgins, C.F. (1985a) *Salmonella typhimurium proP* gene encodes a transport system for the osmoprotectant betaine. *J. Bacteriol. 164* : 1218-1223

Cairney, J., Booth, I.R. and Higgins, C.F. (1985b) Osmoregulation of gene expression in *Salmonella typhimurium* : *proU* encodes an osmotically induced betaine transport system. *J. Bacteriol. 164* : 1224-1232

Cakmak, I and Marschner, H. (1988) Enhanced superoxide radical production in roots of zinc-deficient plants. *J. Exp. Bot. 39* : 1449-1460

Campbell, W.H. and Gowri, G. (1990) Codon usage in higher plants, green algae, and cyanobacteria. *Plant Physiol. 92* : 1-11

Cármenes, R.S., Freije, J.P., Molina, M.M. and Martín, J.M. (1989) Predict7, a program for protein structure prediction. *Biochem. Biophys. Res. Commun. 159* : 687-693

Cassab, G.I. and Varner, J.E. (1988) Cell wall proteins. *Annu. Rev. Plant Physiol. Plant Mol. Biol. 39* : 321-353

Cavalieri, A.J. and Huang, A.H.C. (1980) Carrier protein-mediated transport of neutral amino acids into mung bean mitochondria. *Plant Physiol. 66* : 588-591

Chandler, S.F. and Thorpe, T.A. (1987a) Proline accumulation and sodium sulphate tolerance in callus cultures of *Brassica napus* L. cv. Westar. *Plant Cell Rep. 6* : 176-179

Chandler, S.F. and Thorpe, T.A. (1987b) Characterisation of growth, water relations and proline accumulation in sodium sulphate tolerant callus of *Brassica napus* L. cv. Westar (Canola) *Plant Physiol.* ***84*** : 106-111

Chandler, P.M. and Robertson, M. (1994) Gene expression regulated by abscisic acid and its relation to stress tolerance. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* ***45*** : 113-141

Chang, C. and Meyerowitz, E.M. (1988) Molecular cloning and DNA sequence of the *Arabidopsis thaliana* alcohol dehydrogenase gene. *Proc. Natl. Acad. Sci. USA* ***83*** : 1408-1412

Chapin, F.S. III (1991a) Integrated responses of plants to stress : A centralised system of physiological responses. *Bioscience* ***41*** : 29-36

Chapin, F.S., III (1991b) Effects of multiple environmental stresses on nutrient availability and use. *In* Mooney, H.A., Winner, W.E. and Pell, E.J. (eds) Responses of Plants to Multiple Stresses. Academic Press, San Diego, pp. 67-88

Chaves, M.M. (1991) Effects of water deficits on carbon assimilation. *J. Exp. Bot.* ***42*** : 1-16

Chen, H.-H., Li, P.H. and Brenner, M.L. (1983) Involvement of abscisic acid in potato cold acclimation. *Plant Physiol.* ***71*** : 362-365

Chen, Z., Silva, H. and Klessig, D.F. (1993) Active oxygen species in the induction of systemic aquired resistance by salicylic acid. *Science* ***262*** : 1883-1886

Chiang, H.-H. and Dandekar, A.M. (199) The regulation of proline accumulation at low water potentials in *Arabidopsis thaliana*. *Plant Physiol. Suppl.* ***96*** : 108

Chilson, O.P., Kelly-Chilson, A.E. and Siegel, N.R. (1991) Pyrroline-5-carboxylate reductase in soybean nodules : Isolation/partial primary structure/evidence for isozymes. *Arch. Biochem. Biophys.* ***288*** : 350-357

Chothia, C. (1988) The 14th barrel rolls out. *Nature* ***333*** : 598-599

Chou, P.Y. (1989) Prediction of protein structural classes from amino acid compositions. *In* Fasman, G.D. (ed) Prediction of Protein Structure and the Principles of Protein Conformation. Plenum Press, New York, pp. 549-586

Chou, P.Y. and Fasman, G.D. (1974a) Conformational parameters for amino acids in helical, β-sheet, and random coil regions calculated from proteins. *Biochemistry 13* : 211-221

Chou, P.Y. and Fasman, G.D. (1974b) Prediction of protein conformation. *Biochemistry 13* . 222-245

Chou, P. and Fasman, G. (1977) β-turns in proteins. *J. Mol Biol. 115* : 135-175

Chou, P.Y. and Fasman, G.D. (1978a) Empirical prediction of protein conformation. *Annu. Rev. Biochem. 47* : 251-276

Chou, P.Y. and Fasman, G.D. (1978b) Prediction of the secondary structure of proteins from their amino acid sequence. *Adv. Enzymol. 47* : 45-148

Chou, P.Y. and Fasman, G.D. (1979) Prediction of β-turns. *Biophys. J. 26* : 367-384

Chou, I.T., Chen, C.T. and Kao, C.H. (1991) Characteristics of the induction of the accumulation of proline by abscisic acid and isobutyric acid in detached rice leaves. *Plant Cell Physiol. 32* : 269-272

Christian, J.H.B. (1955) The influence of nutrition on the watewr relations of *Salmonella oranienburg*. *Aust. J. Biol. Sci. 8* : 75-82

Chu, T.M., Aspinall, D. and Paleg, L.G. (1974) Stress metabolism. VI. Temperature stress and the accumulation of proline in barley and radish. *Aust. J. Plant Physiol. 1* : 87-97

Chu, T.M., Jusaitis, M., Aspinall, D. and Paleg, L.G. (1978) Accumulation of free proline at low temperatures. *Physiol. Plant. 43* : 254-260

Cohen, F.E. and Sternberg, M.J.E. (1980) On the prediction of protein structure : The significance of the root-mean-square deviation. *J. Mol. Biol. 138* : 321-333

Collins, F. and Galas, D. (1993) A new five-year plan for the U.S. human genome project. *Science 262* : 43-46

Corcuera, L.J., Hintz, M. and Pahlich, E. (1989) Proline metabolism in *Solanum tuberosum* cell suspension cultures under water stress. *J. Plant Physiol. 134* : 290-293

Coruzzi, G.M. (1991) Molecular approaches to the study of amino acid biosynthesis in plants. *Plant Sci. 74* : 145-155

Costilow, R.N. and Cooper, D. (1978) Identity of proline dehydrogenase and Δ^1-pyrroline-5-carboxylic acid reductase in *Clostridium sporogenes*. *J. Bacteriol.* ***134*** : 139-146

Coughlan, S.J. and Wyn Jones, R.G. (1980) Some responses of *Spinacea oleracea* to salt stress. *J. Exp. Bot.* ***31*** : 883-893

Craik, C.S., Rutter, W.J. and Fletterick, R. (1983) Splice junctions : Association with variation in protein structure. *Science* ***220*** : 1125-1129

Cramer, G.R., Abel-Basset, R. and Seeman, J.R. (1990) Salinity-calcium interactions on root growth and osmotic adjustment of two corn cultivars differing in salt tolerance. *J. Plant Nutr.* ***13*** : 1453-1462

Crawford, D.J. (1990) Plant Molecular Systematics : Macromolecular Approaches. John Wiley and Sons, New York, p.189

Creelman, R.A. and Zeevaart, J.A.D. (1985) Abscisic acid accumulation in spinach leaf slices in the presence of penetrating and nonpenetrating solutes. *Plant Physiol.* 77 : 25-28

Cress, W.A. and Johnson, G.V. (1987) The effect of three osmotic agents on free proline and amino acid pools in *Atriplex canescens* and *Hilaria jamesii*. *Can. J. Bot.* ***65*** : 799-801

Csank, C., Taylor, F.M. and Martindale, D.W. (1990) Nuclear pre-mRNA introns : Analysis and comparison of intron sequences from *Tetrahymena thermophila* and other eukaryotes. *Nucl. Acids Res.* ***18*** : 5133-5141

Csonka, L.N. (1981) Proline overproduction results in enhanced osmotolerance in *Salmonella typhimurium*. *Mol. Gen. Genet.* ***182*** : 82-86

Csonka, L.N. (1982) A third L-proline permease in *Salmonella typhimurium* which functions in media of elevated osmotic strength. *J. Bacteriol.* ***151*** : 1433-1443

Csonka, L.N. (1989) Physiological and genetic responses of bacteria to osmotic stress. *Microbiol. Rev.* ***53*** : 121-147

Csonka, L.N. and Hanson, A.D. (1991) Prokaryotic osmoregulation : Genetics and physiology. *Annu. Rev. Microbiol.* ***45*** : 569-606

Csonka, L.N., Gelvin, S.B., Goodner, B.W., Orser, C.S., Siemieniak, D. and Slighton, J.L. (1988) Nucleotide sequence of a mutation in the *proB* gene of *Escherichia coli* that confers proline overproduction and enhanced tolerance to osmotic stress. *Gene 64* : 199-205

Cushman, J.C., Vernon, D.M. and Bohnert, H.J. (1992) ABA and the transcriptional control of CAM induction during salt stress in the common ice plant. *In* Verma, D.P.S. (ed) Control of Plant Gene Expression, CRC Press, Boca Raton, Florida, pp. 287-300

Dale, J and Campbell, W.F. (1981) Response of tomato plants to stressful temperatures. *Plant Physiol. 67* : 26-29

Dallmier, K.A. and Stewart, C.R. (1992) Effect of exogenous abscisic acid on proline dehydrogenase activity in maize (*Zea mays* L.) *Plant Physiol. 99* : 762-764

Dalton, D.A., Langeberg, L. and Robbins, M. (1992) Purification and characterisation of monodehydroascorbate reductase from soybean root nodules. *Arch. Biochem. Biophys. 292* : 281-286

Dandekar, A.M. and Uratsu, S.L. (1988) A single base pair change in proline biosynthesis genes causes osmotic stress tolerance. *J. Bacteriol. 170* : 5943-5945

Darby, N.J. and Creighton, T.E. (1993) Protein Structure. IRL Press, Oxford, pp. 1-21, 23-41, 63-89

Das, N., Misra, M. and Misra, A.N. (1990) Sodium chloride salt stress induced metabolic changes in callus cultures of pearl millet (*Pennisetum americanum* L. Leeke) : Free solute accumulation. *J. Plant Physiol. 137* : 244-246

Dashek, W.V. and Erickson, S.S. (1981) Isolation, assay, biosynthesis, metabolism, uptake and translocation, and function of proline in plant cells and tissues. *Bot. Rev. 47* : 349-385

DasSarma, S., Tischer, E. and Goodman, H.M. (1986) Plant glutamine synthetase complements a *glnA* mutation in *Escherichia coli*. *Science 232* : 1242-1244

Davies, D.D. (1986) The fine control of cytosolic pH. *Physiol. Plant. 67* : 702-706

Dean, C., Tamaki, S., Dunsmir, P., Favreau, M., Katayama, C., Dooner, H. and Bedbrook, J. (1986) mRNA transcripts of several plant genes are polyadenylated at multiple sites *in vivo*. *Nucl. Acids Res. 14* : 2229-2240

Degols, G. (1987) Functional analysis of the regulatory region adjacent to the *cargB* gene of *Saccharomycet cerevisiae* : Nucleotide sequence, gene fusion experiments and *cis*-dominant regulatory mutation analysis. *Eur. J. Biochem.* ***169*** : 193-200

Delauney, A.J and Verma, D.P.S. (1990) A soybean Δ^1-pyrroline-5-carboxylate reductase cDNA was isolated by functional complementation in *Escherichia coli* and was found to be osmoregulated. *Mol. Gen. Genet.* ***221*** : 299-305

Delauney, A.J and Verma, D.P.S. (1993) Proline biosynthesis and osmoregulation. *Plant J.* ***4*** : 215-223

Delauney, A.J, Hu, C-A.A., Kavi Kishor, P.B. and Verma, D.P.S. (1993) Cloning of ornithine-δ-aminotransferase cDNA from *Vigna aconitifolia* by *trans*-complementation in *Escherichia coli* and regulation of proline biosynthesis. *J. Biol. Chem.* ***268*** : 18673-18678

Demmig, B. and Winter, K. (1986) Sodium, potassium, chloride and proline concentrations of chloroplasts isolated from a halophyte *Mesembryanthemum crystallinum* L. *Planta* ***168*** : 421-4426

Deutch, A.H., Smith, C.J., Rushlow, K.E. and Kretschmer, P.J. (1982) *Escherichia coli* Δ^1-pyrroline-5-carboxylate reductase : Gene sequence, protein overproduction and purification. *Nucl. Acids Res.* ***10*** : 7701-7714

Deutch, A.H., Smith, C.J., Rushlow, K.E. and Kretschmer, P.J. (1984) Analysis of the *Escherichia coli proBA* locus by DNA and protein sequencing. *Nucl. Acids Res.* ***12*** : 6337-6355

De Vos, C.H.R., Vonk, M.J., Vooijs, R. and Schat, H. (1992) Glutathione depletion due to copper-induced phytochelatin synthesis causes oxidative stress in *Silene cucubalus*. *Plant Physiol.* ***98*** : 853-858

Diamantoglou, S. and Rhizopoulou, S. (1992) Free proline accumulation in sapwood, bark and leaves of three evergreen sclerophylls and a comparison with an evergreen conifer. *J. Plant Physiol.* ***140*** : 361-365

Dijkshoorn, W., Lathwell, D.J. and De Wit, C.T. (1968) Temporal changes in carboxylate content of ryegrass with stepwise change in nutrition. *Plant Soil* ***29*** : 369-390

Ding, J.P. and Pickard, B.G. (1993) Mechanosensory calcium-selective cation channels in epidermal cells. *Plant J.* ***3*** : 83-110

Doddema, H., Eddin, R.S. and Mahasneh, A. (1986) Effects of seasonal changes of soil salinity and soil nitrogen on the N-metabolism of the halophyte *Arthrocnemum fruticosum* (L.) Moq. *Plant Soil* **92** : 279

Dolferus, R., de Bruxelles, G., Dennis, E.S. and Peacock, W.J. (1994) Regulation of the *Arabidopsis Adh* gene by anaerobic and other environmental stresses. *Ann. Bot.* **74** : 301-308

Doolittle, R.F. (1985) The genealogy of some recently evolved vertebrate proteins. *Trends Biochem. Sci.* **10** : 233-237

Doolittle, R.F. (1989) Redundancies in protein sequences. *In* Fasman, G.D. (ed) Prediction of Protein Structure and the Principles of Protein Conformation. Plenum Press, New York, pp. 599-623

Doolittle, R.F. (1990) Similar amino acid sequences revisited. *In* Bradshaw, R.A. and Purton, M. (eds) Proteins : Form and Function. Elsevier, Cambridge, pp. 71-74

Douce, R. and Neuburger, M. (1990) Metabolite exchange between the mitochondrion and the cytosol. *In* Dennis, D.T. and Turpin, D.H. (eds) Plant Physiology, Biochemistry and Molecular Biology. Longmans, London, pp. 173-190

Dougherty, K.M., Brandriss, M.C. and Valle, D. (1992) Cloning human pyrroline-5-carboxylate reductase cDNA by complementation in *Saccharomyces cerevisiae*. *J. Biol. Chem.* **267** : 871-875

Draper, S.R. (1972) Amino acid changes associated with low temperature treatment of *Lolium perenne*. *Phytochemistry* **11** : 639-641

Dungey, N.O. and Davies, D.D. (1982a) Protein turnover in isolated barley leaf segments and the effects of stress. *J. Exp. Bot.* **33** : 12-20

Dungey, N.O. and Davies, D.D. (1982b) Protein turnover in the attached leaves of non-stressed and stressed barley seedlings. *Planta* **154** : 435-440

Dure, L., III (1993) The lea proteins of higher plants. *In* Verma, D.P.S. (ed) Control of Plant Gene Expression. CRC Press, Boca Raton, Florida, pp. 325-335

Durzan, D.J. and Steward, F.C. (1963) The nitrogen metabolism and seasonal changes in shoots of *Picea glauca* (Moench) Voss. *Plant Physiol. Suppl.* **38** : vi

Eagle, H., Washington, C.L. and Levy, M. (1965) End product control of amino acid synthesis by cultured human cells. *J. Biol. Chem. 240* : 3944-3950

Eberhardt, H.J. and Wegmann, K. (1979) Effects of abscisic acid and proline on adaptation of tobacco callus cultures to salinity and osmotic shock. *Physiol. Plant. 76* : 283-288

Eder, A. and Huber, W. (1977) About the effects of abscisic acid and kinetin on biochemical changes in *Pennisetum typhoides* during stress conditions. Z. *Pflanzenphysiol. 84* : 301-311

Edwards, J.W., Walker, E.L. and Coruzzi, G.M. (1990) Cell-specific expression in transgenic plants reveals nonoverlapping roles for chloroplast and cytosolic glutamine synthetase. *Proc. Natl. Acad. Sci. USA 87* : 3459-3463

Eisenberg, D., Weiss, R. and Terwilliger, T. (1982) The helical hydrophobic moment : A measure of the amphilicity of a helix. *Nature 299* : 371-374

Elthon, T.E. and Stewart, C.R. (1981) Submitochondrial location and electron transport characteristics of enzymes involved in proline oxidation. *Plant Physiol. 67* : 780-784

Elthon, T.E. and Stewart, C.R. (1982) Proline oxidation in corn mitochondria : Involvement of NAD, relationship to ornithine metabolism, and sidedness of the inner membrane. *Plant Physiol. 70* : 567-572

Elthon, T.E. and Stewart, C.R. (1984) Effects of the proline analog L-thiazoline-4-carboxylic acid on proline metabolism. *Plant Physiol. 74* : 213-218

Emes, M.J. and Fowler, M.W. (1979) Intracellular interactions between the pathways of carbohydrate oxidation and nitrate assimilation in plant roots. *Planta 145* : 287-292

Engelman, D.M., Goldman, A. and Steitz, T.A. (1982) The identification of helical segments in the polypeptide chain of bacteriorhodopsin. *Methods Enzymol. 88* : 81-88

Ericson, M.C. and Alfinito, S. (1984) Proteins produced during salt stress in tobacco cell culture. *Plant Physiol. 74* : 506-509

Farr, S.B. and Kogoma, T. (1991) Oxidative stress responses in *Escherichia coli* and *Salmonella typhimurium*. *Microbiol. Rev. 55* : 561-585

Fasman, G.D. (1985) A critique of the utility of the prediction of protein secondary structure. *J. Biosci. 8* : 15-23

Fasman, G.D. (1989) The development of the prediction of protein structure. *In* Fasman, G.D. (ed) Prediction of Protein Structure and the Principles of Protein Conformation. Plenum Press, New York, pp. 193-316

Fasman, G.D. (1990) Protein conformational prediction. *In* Bradshaw, R.A. and Purton, M. (eds) Proteins : Form and Function. Elsevier, Cambridge, pp. 135-145

Feinberg, A.P. and Vogelstein, B. (1983) A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. *Anal. Biochem. 132* : 6-13

Feldman, K.A. and Marks, M.D. (1987) *Agrobacterium*-mediated transformation of germinating seeds of *Arabidopsis thaliana* : A non-tissue culture approach. *Mol. Gen. Genet. 208* : 1-9

Feliciello, I. and Chinali, G. (1993) A modified alkaline lysis method for the preparation of highly purified plasmid DNA from *Escherichia coli*. *Anal. Biochem. 212* : 394-401

Felsenstein, J. (1988) Phylogenies from molecular sequences : Inference and reliability. *Annu. Rev. Genet. 22* : 521-553

Ferreira, L.G.R., DeSouza, J.G. and Prisco, J.T. (1979) Effects of water deficit on proline accumulation and growth of two cotton genotypes of different drought resistance. *Z. Pflanzenphysiol. 93* : 189-199

Finer-Moore, J., Fernando Baxan, J., Rubin, J. and Stroud, R.M. (1989) Identification of membrane proteins and soluble protein secondary structure elements, domain structure, and packing arrangements by Fourier-transform amphipathic analysis. *In* Fasman, G. (ed) Prediction of Protein Structure and the Principles of Protein Conformation. Plenum Press, New York, pp. 719-759

Finkelstein, R.R. and Somerville, C.R. (1990) Three classes of abscisic acid (ABA) mutations of *Arabidopsis* define genes that control overlapping subsets of ABA responses. *Plant Physiol. 94* : 1172-1179

Fitch, W.M. and Margoliash, E. (1967) Construction of phylogenetic trees. *Science 155* : 279-284

Flowers, T.J. and Yeo, A.R. (1986) Ion relations of plants under drought and salinity stress. *Aust. J. Plant*

Floyd, R.A. and Nagy, Z.S. (1984) Formation of long-lived hydroxyl free radical adducts of proline and hydroxyproline in a Fenton reaction. *Biochem. Biophys. Acta 790* : 94-97

Fluhr, R. and Chua, N.-H. (1986) Developmental regulation of two genes encoding ribulose-bisphosphate carboxylase subunit in pea and transgenic petunia plants : Phytochrome response and blue-light induction. *Proc. Natl. Acad. Sci. USA 83* : 2358-2362

Fons, M., Cami, B. and Chippaux, M. (1991) Possible involvement of a L-Δ^1-pyrroline-5-carboxylate (P5C) reductase in the synthesis of proline in *Desulphovibrio desulphuricans* Norway. *Biochem. Biophys. Res. Commun. 179* : 1088-1094

Foster, R., Gasch, A., Kay, S. and Chua, N.-H. (1990) Analysis of protein/DNA interactions. *In* Koncz, C., Chua, N.-H. and Schell, J. (eds) Methods in *Arabdopsis* Research. World Scientific, Singapore, pp. 378-392

Fou'gere, F., Rudelier, D.L. and Streeter, J.G. (1991) Effects of salt stress on amino acid, organic acid, and carbohydrate composition of roots, bacterioids and cytosol of alfalfa (*Medicago sativa* L.) *Plant Physiol. 96* : 1228-1236

Frank, L. and Ranhand, B. (1964) Proline metabolism in *E. coli*. III. The proline catabolic pathway. *Arch. Biochem. Biophys. 107* : 325-331

Frehner, M., Pozueta-Romero, J. and Akazawa, T. (1990) Enzyme sets of glycolysis, gluconeogenesis, and oxidative pentose phosphate pathway are not complete in non-green highly purified amyloplasts of sycamore (*Acer pseudoplatanus* L.) cell suspension cultures. *Plant Physiol. 94* : 538-544

Freney, J.R., Delwiche, C.C. and Johnson, C.M. (1959) The effect of chloride on the free amino acids of cabbage and cauliflower plants. *Aust. J. Biol. Sci. 12* : 160-166

Fried, M. and Crothers, D.M. (1981) Equilibria and kinetics of the lac repressor-operator interactions by polyacrylamide gel electrophoresis. *Nucl. Acids Res. 9* : 6505-6524

Fukusawa, H. (1954) On the free amino acids in anthers of male-sterile wheat and maize. *Jap. J. Genet. 29* : 135-138

Fukutaku, Y. and Yamada, Y. (1984) Sources of proline-nitrogen in water-stressed soybean (*Glycine max*). II. Fate of ^{15}N-labelled protein. *Physiol. Plant. 61* : 622-628

Gallie, D.R. (1993) Posttranscriptional regulation of gene expression in plants. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* ***44*** : 77-105

Gallie, D.R., Lucas, W.J. and Walbot, V. (1989) Visualising mRNA expression in plant protoplasts : Factors influencing efficient mRNA uptake and translation. *Plant Cell* ***1*** : 301-311

Garcia-Rios, M.G., LaRosa, P.C., Bressan, R.A., Csonka, L.N. and Hanquier, J.M. (1991) Third International Congress, International Society for Plant Molecular Biology, 6-11 October 1991, Tucson, Arizona : Abstract No. 1507

Garner, M.M. and Revzin, A. (1981) A gel electrophoresis method for quantifying the binding of proteins to specific DNA regions : Applications to components of the *Escherichia coli* lactose operon regulatory system. *Nucl. Acids Res.* ***9*** : 3047-3060

Garnier, J., Osguthorpe, D.J. and Robson, B. (1978) Analysis of the accuracy and implications of a simple method for predicting the secondary structure of globular proteins. *J. Mol. Biol.* ***120*** : 97-120

Gaxiola, R., de Larrinoa, I.F., Villalba, J.M. and Serrano, R. (1992) A novel and conserved salt-induced protein is an important determinant of salt tolerance in yeast. *EMBO J.* ***11*** : 3157-3164

Geoffrion, Y. and Larochelle, J. (1984) The free amino acid contribution to osmotic stress in *Acanthamoeba castellanii*. *Can. J. Zool.* ***62*** : 1954-1959

Gherardini, F.C., Hobbs, M.M., Stamm, L.V. and Bassford, Jr., P.J. (1990) Complementation of an *Escherichia coli proC* mutation by a gene cloned from *Treponema pallidum*. *J. Bacteriol.* ***172*** : 2996-3002

Ghislain, M., Frankard, V., Vendenbossche, D., Matthews, B.F. and Jacobs, M. (1994) Molecular analysis of the aspartate kinase-homoserine dehydrogenase gene from *Arabidopsis thaliana*. *Plant Mol. Biol.* ***24*** : 835-851

Gilbert, W. (1978) Why genes in pieces? *Nature* ***271*** : 501

Gilbert, W., Marchionni, M. and McKnight, G. (1986) On the antiquity of introns. *Cell* ***46*** : 151-154

Gilles, R. (1979) Intracellular organic osmotic effectors. *In* Gilles, R. (ed) Mechanisms of osmoregulation in animals. Maintenance of cell volume. John Wiley and Sons, New York, pp. 111-154

Giraudat, J., Hauge, B.M., Valon, C., Smalle, J., Parcy, F. and Goodman, H.M. (1992) Isolation of the *Arabidopsis* ABI3 gene by positional cloning. *Plant Cell 4* : 1251-1261

Gish, W. and States, D.J. (1993) Identification of protein coding regions by database similarity search. *Nature Genet. 3* : 266-272

Gloux, K. and Le Rudulier, D. (1989) Transport and catabolism of proline betaine in salt stressed *Rhizobium meliloti*. *Arch. Microbiol. 151* : 143

Goas, G., Goas, M. and Larher, F. (1982) Accumulation of free proline and glycine betaine in *Aster tripolium* subjected to a saline shock : a kinetic study related to light period. *Physiol. Plant. 55* : 383-388

Golan-Goldhirsh, A., Hankamer, B. and Lips, S.H. (1990) Hydroxyproline and proline content of cell walls of sunflower, peanut and cotton grown under salt stress. *Plant Sci. 69* : 27-32

Goodall, G.J. and Filipowicz, W. (1989) The AU-rich sequences present in the introns of plant nuclear pre-mRNAs are required for splicing. *Cell 58* : 473-483

Göring, H. and Plescher, F. (1986) Proline accumulation induced by weak acids and IAA in coleoptiles of wheat seedlings. *Biol. Plant. 28* : 401-406

Gotoh, O. (1986) Alignment of three biological sequences with an efficient traceback procedure. *J. Theor. Biol. 121* : 327-337

Graham, I.A., Smith, L.M., Brown, J.W.S., Leaver, C.L. and Smith, S.M. (1989) The malate synthase gene of cucumber. *Plant Mol. Biol. 13* : 673-684

Grantham, R., Gautier, C. and Gouy, M. (1980) Codon frequencies in 119 individual genes confirm consistent choices of degenerate bases according to genome type. *Nucl. Acids Res. 8* : 1893-1912

Grantham, R., Gautier, C., Gouy, M., Jacobzone, M. and Mercier, R. (1981) Codon catalogue usage is a genome strategy modulated for gene expressivity. *Nucl. Acids Res. 9* : r43-r74

Green, M.R. (1986) Pre-mRNA splicing. *Annu. Rev. Genet. 20* : 671-708

Green, P.J., Kay, S.A., Lam, E. and Chua, N.-H. (1989) *In vitro* DNA footprinting. *In* Gelvin, S.B., Schilperoort, R.A. and Verma, D.P.S. (eds) Plant Molecular Biology Manual. Kluwer Academic, Dordrecht, pp. B11/1 - B11/22

Greenberg, D.M. (1962) Pyrroline-5-carboxylate reductase. *Methods Enzymol. 5* : 959-964

Greenberg, J.T. and Demple, B. (1989) A global response induced in *E.coli* by redox-cycling agents overlaps with that induced by peroxide stress. *J. Bacteriol. 171* : 3933-3939

Greenway, H. and Setter, T.L. (1979) Accumulation of proline and sucrose during the first hours after transfer of *Chlorella emersonii* to high NaCl. *Aust. J. Plant Physiol. 6* : 69-79

Greenway, H. and Munns, R. (1980) Mechanisms of salt tolerance in nonhalophytes. *Annu. Rev. Plant Physiol. 31* : 149-190

Greer, S. and Perham, R.N. (1986) Glutathione reductase from *Escherichia coli* : Cloning and sequence analysis of the gene and relationship to other flavoprotein disulphide oxidoreductases. *Biochemistry 25* : 2736-2742

Grier, R.L., Merkley, D.F. and Roth, J.A. (1983) Pilot study on the treatment of thioproline of 14 small animals with tumours. *Am. J. Vet. Res. 4* : 2162-2166

Grierson, D. and Covey, S.N. (1988) Plant Molecular Biology (Second Edition). Blackie, New York, pp. 7, 31

Guiltinan, M.J., Marcotte, W.R. and Quatrano, R.S. (1990) A plant leucine zipper protein that recognises an abscisic acid response element. *Science 250* : 267-271

Hagedorn, C.H. and Phang, J.M. (1983) Transfer of reducing equivalents into mitochondria by the interconversions of proline and Δ^1-pyrroline-5-carboxylate. *Arch. Biochem. Biophys. 225* : 95-101

Hagler, A.T. and Honig, B. (1978) On the formation of protein tertiary structure on a computer. *Proc. Natl. Acad. Sci. USA 75* : 554-558

Hall, J.L., Harvey, D.M.R. and Flowers, T.J. (1978) Evidence for the cytoplasmic localisation of betaine in leaf cells of *Suaeda maritima*. *Planta 140* : 59-62

Hamilton, P.T. and Reeve, J.N. (1985) Structure of genes and an insertion element in the methane producing archaebacterium *Methanobrevibacter smithii*. *Mol. Gen. Genet. 200* : 47-59

Handa, S., Handa, A.K., Hasegawa, P.M. and Bressan, R.A. (1986) Proline accumulation and the adaptation of cultured plant cells to salinity stress. *Plant Physiol. 80* : 938-945

Handa, S., Bressan, R.A., Handa, A.K., Carpita, N.C. and Hasegawa, P.M. (1983) Solutes contributing to osmotic adjustment in cultured plant cells adapted to water stress. *Plant Physiol. 73* : 834-843

Hanks, R.W. and Feldman, A.W. (1963) Comparison of free amino acids and amides in roots of healthy and *Radopholus similis* infected grapefruit seedlings. *Phytopathology 53* : 419-422

Hanson, A.D. and Nelsen, C.E. (1978) Betaine accumulation and [^{14}C]formate metabolism in water-stressed barley leaves. *Plant Physiol. 62* : 305-312

Hanson, A.D. and Tully, R.E. (1979) Amino acids translocated from turgid and water-stressed barley leaves. II. Studies with ^{15}N and ^{14}C. *Plant Physiol. 64* : 467-471

Hanson, A.D. and Hitz, W.D. (1982) Metabolic responses of mesophytes to plant water deficits. *Annu. Rev. Plant Physiol. 33* : 163-203

Hanson, A.D., Nelsen, C.E. and Everson, E.H. (1977) Evaluation of free proline accumulation as an index of drought resistance using two contrasting barley cultivars. *Crop Sci. 17* : 720-726

Hanson, A.D., Nelsen, C.E., Pedersen, A.R. and Everson, E.H. (1979) Capacity for proline accumulation during water stress in barley and its implication for breeding for drought resistance. *Crop Sci. 19* : 489-493

Hanson, A.D., Rathinasabapathi, B., Rivoal, J., Burnet, M., Dillon, M.O. and Gage, D.A. (1994) Osmoprotective compounds in the Plumbaginaceae : A natural experiment in metabolic engineering of stress tolerance. *Proc. Natl. Acad. Sci. USA 91* : 306-310

Hanukoglu, I. and Gutfinger, T. (1989) cDNA sequence of adrenodoxin reductase : Identification of NADP-binding sites in oxidoreductases. *Eur. J. Biochem. 180* : 479-484

Hartmuth, K. and Barta, A. (1986) *In vitro* processing of a plant pre-mRNA in HeLa cell nuclear extracts. *Nucl. Acids Res. 14* : 7513-7528

Hassan, N.S. and Wilkins, D.A. (1988) *In vitro* selection for salt-tolerant lines in *Lycopersicon peruvianum*. *Plant Cell Rep. 7* : 463-466

Hasson, E. and Poljakoff-Mayber, A. (1983) Changes in osmolarity and solute content of pea plants exposed to salinity and abscisic acid. *Aust. J. Plant Physiol. 10* : 573-583

Hayward, D.C., Delaney, S.J., Campbell, H.D., Ghysen, A., Benzer, S., Kasprzak, A.B., Cotsell, J.N., Young, I.G. and Gabor Miklos, G.L. (1993) The sluggish-A gene of *Drosophila melanogaster* is expressed in the nervous system and encodes proline oxidase, a mitochondrial enzyme involved in glutamate biosynthesis. *Proc. Natl. Acad. Sci. USA* ***90*** : 2979-2983

Hayzer, D.J. (1983) Sub-cloning of the wild-type *proAB* region of the *Escherichia coli* genome. *J. Gen. Microbiol.* ***129*** : 3215-3225

Hayzer, D.J. and Moses, V. (1978) The enzymes of proline biosynthesis in *Escherichia coli* : Their molecular weights and the problem of enzyme aggregation. *Biochem. J.* ***173*** : 219-228

Hayzer, D.J. and Leisinger, T. (1980) The gene-enzyme relationships of proline biosynthesis in *Escherichia coli*. *J. Gen. Microbiol.* ***118*** : 287-293

Hayzer, D.J. and Leisinger, T. (1981) Proline biosynthesis in *Escherichia coli*. Stoichiometry and end-product identification of the reaction catalysed by glutamate semialdehyde dehydrogenase. *Biochem. J.* ***197*** : 269-274

Hayzer, D.J. and Leisinger, T. (1982) Proline biosynthesis in *Escherichia coli*. Purification and characterisation of glutamate semialdehyde dehydrogenase. *Eur. J. Biochem.* ***121*** : 561-565

Hayzer, D.J. and Leisinger, T. (1983) Proline biosynthesis in *Escherichia coli*. Kinetic and mechanistic properties of glutamate semialdehyde dehydrogenase. *Biochem. Biophys. Acta* ***742*** : 391-398

Heber, U. (1967) Freezing injury and uncoupling of photophosphorylation from electron transport in chloroplasts. *Plant Physiol.* ***42*** : 1343-1350

Heber, U., Tyankova, L. and Santarius, K.A. (1973) Effects of freezing on biological membranes *in vivo* and *in vitro*. *Biochem. Biophys. Acta* ***291*** : 23-37

Heidecker, G. and Messing, J. (1986) Structural analysis of plant genes. *Annu. Rev. Plant Physiol.* ***37*** : 439-466

Heineke, D., Sonnewald, U., Bussis. D., Gunter, G., Leidreiter, K., Wilke, I., Rashke, K., Willmitzer, L. and Heldt, H.W. (1992) Apoplastic expression of yeast-derived invertase in potato. *Plant Physiol.* ***100*** : 301-308

Heino, P., Sandman, D., Lang, V., Nordin, K. and Palva, E.T. (1990) Abscisic acid deficiency prevents development of freezing tolerance in *Arabidopsis thaliana* (L.) Heynh. *Theor. Appl. Genet. 79* : 801-806

Hejala, R.K., Horvath, D.P., Gilmour, S.J. and Thomashow, M.F. (1990) Molecular cloning and expression of two cold regulated genes in *Arabidopsis thaliana*. *Plant Physiol. 93* : 1246-1252

Hellebust, J.A. and Larochelle, J. (1988) Compartmentation and activities of enzymes involved in the metabolism of amino acids implicated in osmoregulatory mechanisms in *Acanthamoeba castellanii*. *J. Protozool. 35* : 498-502

Hellergren, J. and Li, P.H. (1981) Survival of *Solanum tuberosum* suspension cultures to -14°C : The mode of action of proline. *Physiol. Plant. 52* : 449-453

Henikoff, S. and Henikoff, J.G. (1992) Amino acid substitution matrices from protein blocks. *Proc. Natl. Acad. Sci. USA* **89** : 10915-10919

Henslee, J.G., Wakabayashi, Y., Small, W.C. and Jones, M.E. (1983) Factors influencing pyrroline-5-carboxylate synthesis from glutamate by rat intestinal mucosa mitochondria. *Arch. Biochem. Biophys. 226* : 693-703

Henson, I.E. (1985) Solute accumulation and growth in plants of pearl millet (*Pennisetum americanum* L. Leeke) exposed to abscisic acid or water stress. *J. Exp. Bot. 36* : 1889-1899

Hernandez-Lucas, C., Royo, J., Paz-Ares, J., Ponz, F., Garcia-Olmeds, F. and Carbonero, P. (1986) Polyadenylation site heterogeneity in mRNA encoding the precursor of the barley toxin β-hordothionin. *FEBS Lett. 200* : 103-106

Herzfeld, A. and Raper, S.M. (1976) Enzymes of ornithine metabolism in adult and developing rat intestine. *Biochem. Biophys. Acta 428* : 600-610

Herzfeld, A., Mezl, V.A. and Knox, W.E. (1977) Enzymes metabolising Δ^1-pyrroline-5-carboxylate in rat tissues. *Biochem. J. 166* : 95-103

Hille, J., Koornneef, M., Ramanna, M.S. and Zabel, P. (1989) Tomato : A crop species amenable to improvement by cellular and molecular methods. *Euphytica 42* : 1-24

Hinnebusch, A.G. (1988) Mechanisms of gene regulation in the general control of amino acid biosynthesis in *Saccharomyces cerevisiae*. *Microbiol. Rev. 52* : 248-273

Höfte, H., Desprez, T., Amselem, J., Chiapello, H., Caboche, M., Moisan, A., Jourjon, M.-F., Charpenteau, L., Berthomieu, P., Guerrier, D., Giraudat, J., Quigley, F., Thomas, F., Yu, D.-Y., Mache, R., Raynal, M., Cooke, R., Grellet, F., Delseny, M., Parmentier, Y., de Marcillac, G., Gigot, G., Fleck, J., Phillips, G., Axelos, M., Bardet, C., Tremousaygue, D. and Lescure, B. (1993) An inventory of 1152 expressed sequence tags obtained by partial sequencing of cDNAs from *Arabidopsis thaliana*. *Plant J. 4* : 1051-1061

Hong-qi, Z., Croes, A.F. and Linskens, H.F. (1982) Protein synthesis in germinating pollen of *Petunia* : Role of proline. *Planta 154* : 199-203

Hõõg, C. (1991) Isolation of a large number of novel mammalian genes by a differential cDNA library screening strategy. *Nucl. Acids Res. 19* : 6123-6127

Horsch, R.B. (1993) Commercialisation of genetically engineered crops. *Phil. Trans. R. Soc. Lond. B 342* : 287-291

Hoshino, T., Kosuge, T., Hidaka, Y., Tabata, K. and Nakahara, T. (1994) Molecular cloning and sequence analysis of the *proC* gene encoding Δ^1-pyrroline-5-carboxylate reductase from an extremely thermophilic eubacterium *Thermus thermophilus*. *Biochem. Biophys. Res. Commun. 199* : 410-417

Hsiao, T.C. (1970) Rapid changes in levels of polyribosomes in *Zea mays* in response to water stress. *Plant Physiol. 46* : 281-285

Hsiao, T.C. (1973) Plant responses to water stress. *Annu. Rev. Plant Physiol. 24* : 519-570

Hsiao, T.C., Acevedo, F., Ferreres, E. and Henderson, D. (1976) Stress metabolism, water stress, growth and osmotic adjustment. *Philos. Trans. R. Soc. Lond. B 273* : 479-500

Hu, C.-A.A., Delauney, A.J. and Verma, D.P.S. (1992) A bifunctional enzyme (Δ^1-pyrroline-5-carboxylate synthetase) catalyses the first two steps in proline biosynthesis in plants. *Proc. Natl. Acad. Sci. USA 89* : 9354-9358

Huang, A.H.C. and Cavalieri, A.J. (1979) Proline oxidase and water stress-induced proline accumulation in spinach leaves. *Plant Physiol. 63* : 531-535

Hubac, C., Guerrier, D. and Bousquet, U. (1986) Effect of far-red light on malate and potassium contents in cotton leaves : relation to drought resistance. *Physiol. Plant. 66* : 37-40

Hubbard, S.C. and Ivatt, R.J. (1981) Synthesis and processing of asparagine-linked oligosaccharides. *Annu. Rev. Biochem. 50* : 555-583

Hunkapiller, M.W., Strickler, J.E. and Wilson, K.J. (1974) Contemporary methodology for protein structure determination. *Science 226* : 304

Hunt, A.G. (1988) Identification and characterisation of cryptic polyadenylation sites in the 3' region of a pea ribulose-1,5-bisphosphate carboxylase small subunit gene. *DNA 7* : 329-336

Hunt, A.G. (1994) Messenger RNA 3' end formation in plants. *Annu. Rev. Plant Mol. Biol. 45* : 47-60

Hunt, A.G. and MacDonald, M.H. (1989) Deletion analysis of the polyadenylation signal of a pea ribulose-1,5-bisphosphate carboxylase small-subunit gene. *Plant Mol. Biol. 13* : 125-138

Hunt, A.G., Chu, N.M., Odell, J.T., Nagy, F. and Chua, N.-H. (1987) Plant cells do not properly recognise animal gene polyadenylation signals. *Plant Mol. Biol. 8* : 23-35

Ibarra-Caballero, J., Villanueva-Verduzco, C., Molina-Galán, J., Sánchez-de-Jiménez, E. (1988) Proline accumulation as a symptom of drought stress in maize : A tissue differentiation requirement *J. Exp. Bot. 39* : 889-897

Ikemura, T. (1982) Correlation between the abundance of yeast transfer RNAs and the occurrence of the respective codons in protein genes. *J. Mol. Biol. 158* : 573-597

Inana, G., Totsuka, S., Redmond, M., Dougherty, T., Nagle, J., Shiono, T., Ohura, T., Kominami, E. and Katunuma, N. (1986) Molecular cloning of human ornithine aminotransferase mRNA. *Proc. Natl Acad. Sci. USA 83* : 1203-1207

Ingelbrecht, I.L.W., Herman, L.M.F., Dekeyser, R.A., Van Montagu, M.C. and Depicker, A.G. (1989) Different 3' end regions strongly influence the level of gene expression in plant cells. *Plant Cell 1* : 671-680

Itai, C. and Vaadia, Y. (1971) Cytokinin activity in water-stressed shoots. *Plant Physiol. 47* : 87-90

Itai, C. and Paleg, L.G. (1982) Responses of water-stressed *Hordeum distichum* L. and *Cucumis sativus* to proline and betaine. *Plant Sci. Lett. 25* : 329-335

Jähnig, F. (1989) Structure prediction for membrane proteins. *In* Fasman, G. D. (ed) Prediction of Protein Structure and the Principles of Protein Conformation. Plenum Press, New York, pp. 707-717

Jakowec, M.W., Smith, L.T. and Dandekar, A.M. (1985) Recombinant plasmid conferring proline overproduction and osmotic tolerance. *Appl. Environ. Microbiol.* ***50*** : 441-446

Janin, J. (1979) Surface and inside volumes in globular proteins. *Nature* ***277*** : 491-492

Jarvis, P.G. (1981) Plant water relations in models of tree growth. *Stud. Fores. Suecic.* ***160*** : 51-60

Jefferson, R.A., Burgess, S.M. and Hirsh, D. (1986) β-Glucuronidase from *Escherichia coli* as a gene-fusion marker. *Proc. Natl. Acad. Sci. USA* ***83*** : 8447-8451

Jenkins, D.E., Schultz, J.E. and Matin, A. (1988) Starvation-induced cross-protection against heat or H_2O_2 challenge in *Escherichia coli*. *J. Bacteriol.* ***170*** : 3910-3914

Jensen, E.O., Paludan, K., Hyldig-Nielsen, J.Land Jorgensen, P. and Marcker, K.A. (1981) The structure of a chromosomal leghaemoglobin gene from soybean. *Nature* ***291*** : 677-679

Johnson, A.B. and Strecker, H.J. (1962) The interconversion of glutamic acid and proline. IV. The oxidation of proline by rat liver mitochondria. *J. Biol. Chem.* ***237*** : 1876-1882

Jolivet, Y., Pireaux, J.C. and Dizengremel, P. (1990) Changes in properties of barley leaf mitochondria isolated from NaCl-treated plants. *Plant Physiol.* ***94*** : 641-646

Jones, M.E. (1985) Conversion of glutamate to ornithine and proline : Pyrroline-5-carboxylate, a possible modulator of arginine requirements. *J. Nutr.* ***115*** : 509-515

Joshi, C.P. (1987) Putative polyadenylation signals in nuclear genes of higher plants : A compilation and analysis. *Nucl. Acids Res.* ***15*** : 9627-9640

Joshi, S.S. (1984) Effect of salinity stress on organic and mineral constituents in the leaves of pigeonpea (*Cajanus cajan* L. var. C-11) *Plant Soil* ***82*** : 69-75

Jost, A., Perlman, S., Valentino, O., Castanier, M., Scholler, R. and Marge, S. (1988) Experimental control of the differentiation of Leydig cells in the rat fetal testis. *Proc. Natl. Acad. Sci. USA* ***85*** : 8094-8097

Journet, E.P. and Douce, R. (1985) Enzymic capacities of purified cauliflower bud plastids for lipid synthesis and carbohydrate metabolism. *Plant Physiol.* ***79*** : 458-467

Joyce, P.A., Aspinall, D. and Paleg, L.B. (1992) Photosynthesis and the accumulation of proline in response to water deficit. *Aust. J. Plant Physiol.* ***19*** : 249-261

Kagawa, Y., Nojima, H., Nukiwa, N., Ishizuka, M., Nakajima, T., Yasuhara, T., Tanaka, T. and Oshima, T. (1984) High guanine plus cytosine content in the third letter of codons of an extreme thermophile: DNA sequence of the isopropylmalate dehydrogenase of *Thermus thermophilus*. *J. Biol. Chem.* ***259*** : 2956-2960

Kaiser, W.M. (1987) Effects of water deficit on photosynthetic capacity. *Physiol. Plant.* ***71*** : 141-149

Kandpal, R.P. and Rao, N.A. (1985) Alterations in the biosynthesis of proteins and nucleic acids in finger millet (*Eleucine coracana)* seedlings during water stress and the effect of proline on protein biosynthesis. *Plant Sci.* ***40*** : 73-79

Kao, S.M. and Hassan, H.M. (1985) Biochemical characterisation of a paraquat-tolerant mutant of *Escherichia coli*. *J. Biol. Chem.* ***260*** : 10478-10481

Karlin-Neumann, G.A. and Tobin, E.M. (1986) Transit peptides of nuclear-encoded chloroplast proteins share a common amino acid framework. *EMBO J.* ***5*** : 9-13

Kawahara, Y., Oshumi, T., Yoshigara, Y. and Ikeda, S. (1989) Proline in the osmoregulation of *Brevibacterium lactofermentum*. *Agric. Biol. Chem.* ***53*** : 2475-2479

Kawahara, Y., Yoshigara, Y., Ikeda, S. and Hirose, Y. (1990) Effect of glycine betaine, an osmoprotective compound on the growth of *Brevibacterium lactofermentum*. *Appl. Microbiol. Biotechnol.* ***33*** : 574-577

Kay, R., Chan, A., Daly, M. and McPherson, J. (1987) Duplication of CaMV 35S promoter sequences creates a strong enhancer for plant genes. *Science* ***236*** : 1299-1302

Keith, C.S., Hoang, D.O., Barret, B.M., Feigelman, B., Nelson, M.C., Thai, H., Baysdorfer, C. (1993) Partial sequence analysis of 130 randomly selected maize cDNA clones. *Plant Physiol.* ***101*** : 329-332

Kemble, A.R. and MacPherson, H.T. (1954) Liberation of amino acids in perennial rye grass during wilting. *Biochem. J.* ***58*** : 46-59

Ketchum, R.E.B., Warren, R.S., Klima, L.J. , Gutierrez, F.L. and Nabors, M.W. (1991) The mechanism and regulation of proline accumulation in suspension cell cultures of the halophytic grass *Distichlis spicata* L. *J. Plant Physiol.* ***137*** : 368-374

Key, J.L., Lin, C.Y. and Chen, Y.M. (1981) Heat shock proteins of higher plants. *Proc. Natl. Acad. Sci. USA* ***78*** : 3526-3530

Khan, R.A. and Lehle, F.R. (1991) Inheritence of azetidine-2-carboxylic acid resistance in *Arabidopsis thaliana*. *Plant Physiol. Suppl.* ***96*** : 135

Khoo, U. and Stinson, N.T. (1957) Free amino acid differences between cytoplasmic male sterile and normal fertile anthers. *Proc. Natl. Acad. Sci. USA* ***43*** : 603-607

Kieft, T.L. and Spence, S.D. (1988) Osmoregulation in *Thiobacillus ferrooxidans* : stimulation of iron oxidation by proline and betaine under salt stress. *Curr. Microbiol.* ***17*** : 255-273

Kim, P.S. and Baldwin, R.L. (1982) Specific intermediates in the folding reactions of small proteins and the mechanism of protein folding. *Annu. Rev. Biochem.* ***51*** : 459-489

Kirst, G.O. (1977) Coordination of ionic relations and mannitol concentrations in the euryhaline alga *Platymonas subcordiformis* (Hazen) after osmotic shocks. *Planta* ***135*** : 69-75

Kishor, P.B.K. (1988) Effect of salt stress on callus cultures of *Oryza sativa* L. *J. Exp. Bot.* ***39*** : 235-240

Kishor, P.B.K. (1989) Salt stress in cultured rice cells : Effects of proline and abscisic acid. *Plant Cell Environ.* ***12*** : 629-633

Kleczkowski, K. and Kretovich, V.L. (1960) Transamination of amino acids of the ornithine cycle in pea and wheat seedlings. *Biochemistry.* ***25*** : 118-120

Klein, A. and Itai, C. (1989) Is proline involved in stomata regulation of *Commelina communis* plants recovering from salinity stress? *Physiol. Plant.* ***75*** : 399-404

Klein, H. and Jäger, H.J. (1978) Some aspects of relation between nitrogen metabolism in *Pisum sativum* and mineral nutrition. *Plant Soil* ***50*** : 25-35

Kohl, D.H., Schubert, K.R., Carter, M.B., Hagedorn, C.H. and Shearer, G. (1988) Proline metabolism in N_2-fixing root nodules : Energy transfer and regulation of purine synthesis. *Proc. Natl. Acad. Sci. USA* ***85*** : 2036-2040

Kohl, D. H., Kennelly, E.J., Zhu, Y., Schubert, K.R. and Shearer, G. (1991) Proline accumulation, nitrogenase (C_2H_2 reducing) activity and activities of enzymes related to proline metabolism in drought-stressed soybean nodules. *J. Exp. Bot.* **42** : 831-837

Koornneef, M. (1991) Isolation of higher plant developmental mutants. *In* Jenkins, G.I. and Schuch, W. (eds) Molecular Biology of Plant Development. The Company of Biologists Ltd., Cambridge, pp. 1-19

Kowaloff, E.M., Phang, J.M., Granger, A.S. and Downing, S.J. (1977) Regulation of proline oxidase activity by lactate. *Proc. Natl. Acad. Sci. USA* **74** : 5368-5371

Krall, J.P., Edwards, G.E. and Andreo, C.S. (1989) Protection of pyruvate Pi dikinase from maize against cold lability by compatible solutes. *Plant Physiol.* **89** : 280-285

Kramer, J.J., Henslee, J.G., Wakabayashi, Y. and Jones, M.E. (1985) Δ^1-pyrroline-5-carboxylate synthase from rat intestinal mucosa. *Methods Enzymol.* **113** : 113-120

Kregenow, F.M. (1981) Osmoregulatory salt transporting mechanisms : Control of cell volume in anisotonic media. *Annu. Rev. Physiol.* **43** : 493-505

Kreis, M., Williamson, M.S., Forde, J., Schmutz, D., Clark, J., Buxton, B., Pywell, J., Marris, C., Hendersen, J., Harris, N., Shewry, P.R., Forde, B.G. and Miflin, B.J. (1986) Differential gene expression in the developing barley endosperm. *Phil. Trans. R. Soc. Lond. B* **314** : 355-365

Krishna, R.V. and Leisinger, T. (1979) Biosynthesis of proline in *Pseudomonas aeruginosa* : Partial purification and characterisation of γ-glutamyl kinase. *Biochem. J.* **181** : 215-222

Krishna, R.V. Beilstein, P. and Leisinger, T. (1979) Biosynthesis of proline in *Pseudomonas aeruginosa*. Properties of γ-glutamyl phosphate reductase and 1-pyrroline-5-carboxylate reductase. *Biochem. J.* **181** : 223-230

Krueger, R., Jäger, H.J., Hintz, M. and Pahlich, E. (1986) Purification to homogeneity of pyrroline-5-carboxylate reductase of barley. *Plant Physiol.* **80** : 142-144

Krzywicki, K.A. and Brandriss, M.C. (1984) primary structure of the nuclear *PUT2* gene involved in the mitochondrial pathway for proline utilisation in *Saccharomyces cerevisiae*. *Mol. Cell Biol.* **4** : 2837-2842

Kueh, J.S.H. and Bright, S.W.J. (1982) Biochemical and genetical analysis of three proline-accumulating barley mutants. *Plant Sci. Lett.* **27** : 233-241

Kuhlemeier, C. (1992) Transcriptional and post-transcriptional regulation of gene expression in plants. *Plant Mol. Biol.* **19** : 1-14

Kuhlemeier, C., Green, P.J. and Chua, N.-H. (1987) Regulation of gene expression in plants. *Annu. Rev. Plant Physiol.* **38** : 221-257

Kuhlemeier, C., Fluhr, R. and Chua, N.-H. (1988a) Upstream sequences determine the difference in transcript abundance of pea *rbcS* genes. *Mol. Gen. Genet.* **212** : 405-411

Kuhlemeier, C., Cuozzo, M., Green, P.J., Goyvaerts, E., Ward, K. and Chua, N.-H. (1988b) Localisation and conditional redundancy of regulatory elements in *rbcS-3A*, a pea gene encoding the small subunit of ribulose-bisphosphate carboxylase. *Proc. Natl. Acad. Sci. USA* **85** : 4662-4666

Kumar, V. and Sharma, D.R. (1989) Selection and characterisation of an L-thiazolidine-4-carboxylic acid resistant callus culture of *Vigna radiata* (L) Wilczek var. radiat. *Plant Cell Rep.* **7** : 648-651

Kuntz, I.E. (1972) Protein folding. *J. Am. Chem. Soc.* **94** : 4009-4012

Kuo, C.G. and Chen, B.W. (1980) Physiological responses of tomato cultivars to flooding. *J. Amer. Soc. Hort. Sci.* **105** : 751-755

Kuo, C.G., Chen, H.M. and Ma, L.H. (1986) Effect of high temperature on proline content in tomato floral buds and leaves. *J. Amer. Soc. Hort. Sci.* **111** : 746-750

Kurkela, S. and Franck, M. (1990) Cloning and characterisation of a cold- and ABA-inducible *Arabidopsis* gene. *Plant Mol. Biol.* **15** : 137-144

Kyte, J. and Doolittle, R.K. (1982) A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157** : 105-132

Labanauskas, C.K., Stolzy, L.H. and Handy, M.F. (1974) Soil oxygen and *Phytophthora* spp. root infestation effects on protein and free amino acids in lemon and orange leaves. *J. Am. Soc. Hort. Sci.* **99** : 497-500

Laliberte, G. and Hellebust, J.A. (1989a) Regulation of proline content of *Chlorella autotrophica* in response to changes in salinity. *Can. J. Bot.* **67** : 1959-1965

Laliberte, G. and Hellebust, J.A. (1989b) Pyrroline-5-carboxylate reductase in *Chlorella autotrophica* and *Chlorella saccharophila* in relation to osmoregulation. *Plant Physiol.* ***91*** : 917-923

Lam, E. and Chua, N.-H. (1990) GT-1 binding site confers light responsive expression in transgenic tobacco. *Science* ***248*** : 471-474

Lambert, L. and Hoffman, E.K. (1982) Amino acid metabolism and protein turnover under different osmotic conditions in Ehrlich ascites tumour cells. *Mol. Physiol.* ***2*** : 273-286

Lang, A.R.G. (1967) Osmotic coefficients and water potentials of sodium chloride solutions from 0°C to 40°C. *Aust. J. Chem.* ***20*** : 2017-2023

Lang, V., Heino, P. and Palva, E.T. (1989) Low temperature acclimation and treatment with exogenous abscisic acid induce common polypeptides in *Arabidopsis thaliana* (L.) Heynh. *Theor. Appl. Genet.* 77 : 729-734

Larher, F. (1987) Natural abundance of ^{13}C-nuclear magnetic resonance studies on the compatible solutes of halophytic higher plants. *Plant Physiol. Biochem.* ***26*** : 35-45

Larher, F., Leport, L., Petrivalsky, M. and Chappart, M. (1993) Effectors of the osmoinduced proline response in higher plants. *Plant Physiol. Biochem.* ***31*** : 911-922

LaRosa, P.C., Rhodes, D., Rhodes, J.C., Bressan, R.A. and Csonka, L.N. (1991) Elevated accumulation of proline in NaCl-adapted tobacco cells is not due to altered Δ^1-pyrroline-5-carboxylate reductase. *Plant Physiol.* ***96*** : 245-250

LaRosa, P.C., Chen, Z., Nelson, D.E., Singh, N.K.,Hasegawa, P.M. and Bressan, R.A. (1992) Osmotin gene is posttranscriptionally regulated. *Plant Physiol.* ***100*** : 409-415

Layzell, D.B. (1990) N_2 fixation, NO_3^- reduction and NH_4^+ assimilation. *In* Dennis, D.T. and Turpin, D.H. (eds) Plant Physiology, Biochemistry and Molecular Biology. Longmans, London, p. 401

Le Dily, F., Billard, J.B. and Boucaud, J. (1991) Polyamine levels in relation to growth and NaCl concentration in normal and habituated sugarbeet callus cultures. *Plant Cell Environ.* ***14*** : 327-332

Lee, E.C., Walker, R.I. and Guerry, P. (1985) Expression of *Campylobacter* genes for proline biosynthesis in *Escherichia coli*. *Can. J. Microbiol.* ***31*** : 1064-1067

Lefebvre, D.D. (1990) Fundamentals of gene structure and control. *In* Dennis, D.T. and Turpin, D.H. (eds) Plant Physiology, Biochemistry and Molecular Biology. Longman, London, pp. 3-15

Leisinger, T. (1987) Biosynthesis of proline. *In* Neidhart, F.C., Ingram, J.L., Low, K.B., Magasanik, B., Schaecter, M. and Umbarger, H.E. (eds) *Escherichia coli* and *Salmonella typhimurium* : Cellular and Molecular Biology. American Society for Microbiology, Washington D.C., pp. 345-351

Le Rudulier, D., Strom, A.R., Dandekar, A.M., Smith, L.T. and Valentine, R.C. (1984) Molecular biology of osmoregulation. *Science. 224* : 1064-1068

Leuhrsen, K.R. and Walbot, V. (1991) Intron enhancement of gene expression and the splicing efficiency of introns in maize cells. *Mol. Gen. Genet. 225* : 81-93

Leutwiler, L.S., Hough-Evans, B.R. and Meyerowitz, E.M. (1984) The DNA of *Arabidopsis thaliana*. *Mol Gen. Genet. 194* : 15-23

Levitt, J. (1980) Responses of Plants to Environmnetal Stress. Vol. 1 : Chilling, Freezing and High Temperature Stress. Academic Press, New York, pp. 3-18

Levitt, M. and Warshell, A. (1975) Computer simulation of protein folding. *Nature 253* : 694-698

Levitt, M. and Chothier, C. (1976) Structural patterns in globular proteins. *Nature 261* : 552-558

Levy, D., Fogelman, E. and Itzhak, Y. (1988) The effect of water salinity on potatoes (*Solanum tuberosum* L.) : Physiological indices and yielding capacity. *Potato Res. 31* : 601

Lewin, B. (1990) Commitment and activation at PolII promoters : A tale of protein-protein interactions. *Cell 61* : 1161-1164

Lewis, P.J. and Wake, R.G. (1989) DNA and protein sequence conservation at the replication terminus in *Bacillus subtilis* 168 and W23. *J. Bacteriol. 171* : 1402-1408

Lewis, P.N., Momany, F.A. and Scheraga, H.A. (1971) Folding of polypeptide chains in proteins : A proposed mechanism for folding. *Proc. Natl. Acad. Sci. USA 68* : 2293-2297

Li, W. and Brandriss, M.C. (1992) Proline biosynthesis in *Saccharomyces cerevisiae* : Molecular analysis of the *PRO1* gene, which encodes γ-glutamyl kinase. *J. Bacteriol. 174* : 4148-4156

Lim, V.I. (1974a) Structural principles of the globular organisation of protein chains : A stereochemical theory of globular protein secondary structure. *J. Mol. Biol.* ***88*** : 857-872

Lim, V.I. (1974b) Algorithms for prediction of α-helices and β-structural regions in globular proteins *J. Mol. Biol.* ***88*** : 873-894

Lin, C.Y. and Key, J.L. (1967) Dissociation and reassembly of polyribosomes in relation to protein synthesis in the soybean root. *J. Mol. Biol.* ***26*** : 237-247

Ling, C.M. and Hedrick, L.R. (1964) Proline oxidases in *Hansenula subpelliculosa*. *J. Bacteriol.* ***87*** : 1462-1470

Linskens, H.F. and Schrauwen, J. (1969) The release of free amino acids out of the germinating pollen. *Acta Bot. Neerl.* ***18*** : 605-614

Liu, M.S. and Hellebust, J.A. (1976a) Regulation of proline metabolism in the marine centric diatom *Cyclotella cryptica*. *Can. J. Bot.* ***54*** : 949-959

Liu, M.S. and Hellebust, J.A. (1976b) Effects of salinity and osmolarity of the medium on amino acid metabolism in *Cyclotella cryptica*. *Can J. Bot.* ***54*** : 938-948

Lloyd, A.T. and Sharp, P.M. (1992) CODONS : A microcomputer program for codon usage analysis. *J. Hered.* ***83*** : 239-240

Lone, M.I., Kueh, J.S.H., Wyn Jones, R.G. and Bright, S.W.J. (1987) Influence of proline and glycinebetaine on salt tolerance cultured barley embryos. *J. Exp. Bot.* ***38*** : 479-

Lorans, G. and Phang, J.M. (1981) Proline synthesis and redox regulation : Differential functions of pyrroline-5-carboxylate reductase in human lymphoblastoid cell lines. *Biochem. Biophys. Res. Commun.* ***101*** : 1018-1025

Louie, H. and Chan, V.L. (1993) Cloning and characterisation of the gamma-glutamyl phosphate reductase gene of *Campylobacter jejuni*. *Mol. Gen Genet.* ***240*** : 29-35

Lu, T.-S. and Mazelis, M. (1975) L-ornithine : 2-oxoacid amino-transferase from squash (*Cucurbita pepo* L.) cotyledons : Purification and properties. *Plant Physiol.* ***55*** : 502-506

Ludlow, M.M. (1980) Adaptive significance of stomatal responses to water stress. *In* Turner, N.C. and Kramer, P.J. (eds) Adaptation of Plants to Water and High Temperature. Wiley Interscience, New York, pp. 123-138

MacDonald, M.H., Mogen, B.D. and Hunt, A.G. (1991) Characterisation of the polyadenylation signal of the T-DNA-encoded octopine synthase gene. *Nucl. Acids Res. 19* : 5575-5581

Machicado, M. and Boynton, D. (1961) Effects of potassium, magnesium and calcium deficiencies on the intermediary nitrogenous constituents of cocoa (*Theobroma cacao* L.) leaves. *Proc. Am. Soc. Hort. Sci. 77* : 231-235

Madkour, M.A., Smith, L.T. and Smith, G.M. (1990) Preferential osmolyte accumulation : A mechanism of osmotic stress adaptation in diazotrophic bacteria. *Appl. Environ. Microbiol. 56* : 2876-2881

Mahan, M.J. and Csonka, L.N. (1983) Genetic analysis of the *proBA* genes of *Salmonella typhimurium* : Physical and genetic anlyses of the cloned $ProB^{+}A^{+}$ genes of *Escherichia coli* and of a mutant allele that confers proline overproduction and enhanced tolerance. *J. Bacteriol. 156* : 1249-1262

Malik, R.S., Dhankar, J.S. and Turner, N.C. (1979) Influence of soil water deficits on root growth in cotton seedlings. *Plant Soil 53* : 109-115

Manetas, J., Petropoulou, J. and Karabourniotis, G. (1986) Compatible solutes and their effects on PEP carboxylase of C_4 halophytes. *Plant Cell Environ. 9* : 145-151

Maniatis, T. and Reed, R. (1987) The role of small nuclear ribonucleoprotein particles in pre-mRNA splicing. *Nature 325* : 673-678

Maniatis, T., Fritsch, E.F. and Sambrook, J. (1989) Molecular Cloning - A Laboratory Manual. Book 1. (Second Ed.) Cold Spring Harbour Laboratory Press, New York, p. 1.82

Marcotte, W.R. Jr., Russel, R.H. and Quatrano, R.S. (1989) Abscisic acid-response sequences from the Em gene of wheat. *Plant Cell 1* : 969-976

Martin, P.G. and Dowd, J.M. (1984a) The study of plant phylogeny using amino acid sequences of ribulose-1,5-bisphosphate carboxylase. III. Addition of Malvaceae and Ranunculaceae to the phylogenetic tree. *Aust. J. Bot. 32* : 283-292

Martin, P.G. and Dowd, J.M. (1984b) The study of plant phylogeny using amino acid sequences of ribulose-1,5-bisphosphate carboxylase. IV. Proteaceae and Fagaceae and the rate of evolution of the small subunit. *Aust. J. Bot. 32* : 291-299

Martin, P.G. and Dowd, J.M. (1984c) The study of plant phylogeny using amino acid sequences of ribulose-1,5-bisphosphate carboxylase. V. Magnoliaceae, Polygonaceae and the concept of primitiveness. *Aust. J. Bot. 32* : 301-309

Martin, P.G., Dowd, J.M. and Stone, S.J.L. (1983) The study of plant phylogeny using amino acid sequences of ribulose-1,5-bisphosphate carboxylase. II. The analysis of small subunit data to form phylogenetic trees. *Aust. J. Bot. 31* : 411-419

Martin, P.G., Boulter, D. and Penny, D. (1985) Angiosperm phylogeny studied using sequences of five macromolecules. *Taxon 34* : 393-400

Maslenkova, L.T., Miteva, T.S. and Popova, L.P. (1992) Changes in the polypeptide patterns of barley seedlings exposed to jasmonic acid and salinity. *Plant Physiol. 98* : 700-707

Matsuzawa, T. and Ishiguro, I. (1980a) Δ^1-pyrroline-5-carboxylate reductase from baker's yeast. Purification, properties and its application in the assays of L-Δ^1-pyrroline-5-carboxylate and L-ornithine in tissue. *Biochim. Biophys. Acta 613* : 318-323

Matsuzawa, T. and Ishiguro, I. (1980b) Δ^1-pyrroline-5-carboxylate reductase from baker's yeast, Further purification by affinity chromatography with 5' AMP-Sepharose 4B. *Biochem. Biophys. Acta 616* : 381-383

Matthews, B.F., Reardon, E.M., Turano, F.J. and Wilson, B.J. (1988) Amino acid biosynthesis in plants : Approaching an understanding at the molecular level. *Plant Mol. Biol. Rep. 6* : 137-154

Mayer, A.M. and Poljakoff-Mayber, A. (1989) The Germination of Seeds (Fourth Edition). Pergamon Press, Oxford, pp. 161, 165

Mazelis, M. and Creveling, R.K. (1974) L-proline dehydrogenase of *Triticum vulgare* germ : Purification, properties and cofactor interactions. *Phytochemistry 13* : 559-565

Mazelis, M. and Fowden, L. (1969) Conversion of ornithine into proline by enzymes from germinating peanut cotyledons. *Phytochemistry 8* : 801-809

Mazelis, M. and Fowden, L. (1971) The metabolism of proline in higher plants. II. L-proline dehydrogenase from cotyledons of germinating peanut (*Arachis hypogaea* L.) seedlings. *J. Exp. Bot.* **22** : 137-145

McCammon, J.A., Gelin, B.R. and Karplus, M. (1977) Dynamics of folded proteins. *Nature* **267** : 585-590

McCombie, W.R., Adams, M.D., Kelley, J.M., Fitzgerald, M.G., Utterback, T.R., Khan, M., Dubnick, M., Kerlavage, A.R., Venter, J.C. and Fields, C. (1992) *Caenorhabditis elegans* expressed sequence tags identify gene families and potential disease gene homologues. *Nature Genet.* **1** : 124-131

McCue, K.F. and Hanson, A.D. (1990) Drought and salt tolerance : Towards understanding and application. *Trends Biotech.* **8** : 358-362

McDonnell, E.M., Coughlan, S.J. and Wyn Jones, R.G. (1983) Differential effects of abscisic acid on glycinebetaine and proline accumulation in three plant species. *Z. Pflanzenphysiol.* **109** : 207-213

McMichael, B.L. and Elmore, C.D. (1978) Proline accumulation in water stressed cotton leaves. *Crop Sci.* **17** : 905-908

Measures, J.C. (1975) Role of amino acids in osmoregulation of non-halophilic bacteria. *Nature* **257** : 398-400

Mehlhorn, H., Tabner, B.J. and Wellburn, A.R. (1990) Electron spin resonance evidence for the formation of free radicals in plants exposed to ozone. *Physiol. Plant.* **79** : 377-383

Meile, L. and Leisinger, T. (1982) Purification and properties of the bifunctional proline dehydrogenase/Δ^1-pyrroline-5-carboxylate dehydrogenase from *Pseudomonas aeruginosa*. *Eur. J. Biochem.* **129** : 67-75

Meister, A. (1962) Δ^1-piperidine-2-carboxylate and Δ^1-pyrroline-2-carboxylate reductase. *Methods Enzymol.* **5** : 878

Meister, A., Radhakrishnan, A.N. and Buckley, S.D. (1957) Enzymatic synthesis of L-pipecolic acid and L-proline. *J. Biol. Chem.* **229** : 789-800

Menzel, R. and Roth, J. (1981) Purification of the *PutA* gene product : A bifunctional membrane-bound protein from *Salmonella typhimurium* responsible for the two-step oxidation of proline to glutamate. *J. Biol. Chem.* **256** : 9755-9761

Meon, S., Fisher, J.M. and Wallace, H.R. (1978) Changes in free proline following infection of plants with either *Meloidogyne javanica* or *Agrobacterium tumefaciens*. *Physiol. Plant Pathol.* ***12*** : 251-256

Merrill, M.J., Yeh, G.C. and Phang, J.M. (1989) Purified human erythrocyte pyrroline-5-carboxylate reductase : Preferential oxidation of NADPH. *J. Biol. Chem.* ***264*** : 9352-9358

Mestichelli, L.J.J., Gupta, R.N. and Spenser, I.D. (1979) The biosynthetic route from ornithine to proline. *J. Biol. Chem.* ***254*** : 640-647

Meyerowitz, E.M. (1987) *Arabidopsis thaliana*. *Annu. Rev. Genet.* ***21*** : 93-111

Meyerowitz, E. (1992) Introduction to the *Arabidopsis* genome. *In* Koncz, C., Chua, N.-H. and Schell, J. (eds) Methods in *Arabidopsis* Research. World Scientific, Singapore, pp. 100-118

Meyerowitz, E.M. and Pruitt, R.E. (1985) *Arabidopsis thaliana* and plant molecular genetics. *Science* **229** : 1214-1218

Michel, B.E. and Kaufmann, M.R. (1973) The osmotic potential of polyethyleneglycol 6000. *Plant Physiol.* ***51*** : 914-916

Miernyk, J.A. (1990) Glycolysis, the oxidative pentose phosphate pathway and anaerobic respiration. *In* Dennis, D.T. and Turpin, D.H. (eds) Plant Physiology, Biochemistry and Molecular Biology. Longman, London, pp. 77-100

Miler, P.M. and Stewart, C.R. (1976) Pyrroline-5-carboxylic acid reductase from soybean leaves. *Phytochemistry* ***15*** : 1855-1857

Miller, R.J., Bell, D.T. and Koeppe, D.E. (1971) The effect of water stress on some membrane characteristics of corn mitochondria. *Plant Physiol.* ***48*** : 229-231

Mindrinos,, M., Katagiri, F., Yu, G.-L. and Ausubel, F.M. (1994) The *A. thaliana* disease resistance gene *RPS2* encodes a protein containing a nucleotide-binding site and leucine-rich repeats. *Cell* ***78*** : 1089-1099

Mishra, R.K. and Singhal, G.S. (1992) Function of the photosynthetic apparatus of intact wheat leaves under high light and heat stress and its relationship with peroxidation of thylakoid lipids. *Plant Physiol.* ***98*** : 1-6

Mixson, A.J. and Phang, J.M. (1988) The uptake of pyrroline-5-carboxylate : Group translocation mediating the transfer of reducing-oxidising potential. *J. Biol. Chem.* ***263*** : 10720-10724

Moftah, A.E. and Michel, B.E. (1987) The effect of sodium chloride on solute potential and proline accumulation in soybean leaves. *Plant Physiol.* ***83*** : 238-240

Mogen, B., MacDonald, M.H., Graybosch, R. and Hunt, A.G. (1990) Upstream sequences other than AAUAAA are required for efficient messenger RNA 3' end formation in plants. *Plant Cell* ***2*** : 1261-1272

Mogen, B.D., MacDonald, M.H., Leggewie, G. and Hunt, A.G. (1992) Several distinct types of sequence elements are required for efficient mRNA 3' end formation in a pea *rbcS* gene. *Mol. Cell. Biol.* ***12*** : 5406-5414

Mohanty, S.K. and Sridhar, R. (1982) Physiology of rice tungro virus disease : Proline accumulation due to infection. *Physiol. Plant.* ***56*** : 89-93

Montoliu, L., Rigau, J. and Puigdomenech, P. (1990) Multiple polyadenylation sites are active in the α1-tubulin gene from *Zea mays*. *FEBS Lett.* ***277*** : 29-32

Moran, J.F., Becana, M., Iturbe-Ormaetxe, I., Frechilla, S., Klucas, R.V. and Aparicio-Tejo, P. (1994) Drought induces oxidative stress in pea plants. *Planta* ***194*** : 346-352

Morelli, G., Nagy, F., Fraley, R.T., Rogers, S.G. and Chua, N.-H. (1985) A short conserved sequence is involved in light-inducibility of a gene encoding ribulose bisphosphate carboxylase small subunit of pea. *Nature* ***315*** : 200-204

Morgan, R.W., Christman, M.F., Jacobson, F.S., Storz, G. and Ames, B.N. (1986) Hydrogen peroxide-inducible proteins in *Salmonella typhimurium* overlap with heat shock and other stress proteins. *Proc. Natl. Acad. Sci. USA* ***83*** : 8059-8063

Morris, C.J., Thompson, J.F. and Johnson, C.M. (1969) Metabolism of glutamic acid and *N*-acetylglutamic acid in leaf discs and cell-free extracts of higher plants. *Plant Physiol.* ***44*** : 1023-1026

Mundy, J., Yamaguchi-Shinozaki, K. and Chua, N.-H. (1990) Nuclear proteins bind conserved elements in the abscisic acid-responsive promoter of rice *rab* genes. *Proc. Natl. Acad. Sci. USA* ***87*** : 406-410

Munns, R., Brady, C.J. and Barlow, E.W. (1979) Solute accumulation in the apex and leaves of wheat during water stress. *Aust. J. Plant Physiol.* ***6*** : 379-389

Murray, E.E., Lotzer, J. and Eberle, M. (1989) Codon usage in plant genes. *Nucl. Acids Res. 17* : 477-498

Muth, W.L. and Costilow, R.N. (1974) Ornithine cyclase (deaminating). III. Mechanism of conversion of ornithine to proline. *J. Biol. Chem. 249* : 7463-7467

Mutters, R.G., Ferreira, L.G.R. and Hall, A.E. (1989) Proline content in the anthers and pollen of heat-tolerant and heat-sensitive cowpea subjected to different temperatures. *Crop Sci. 29* : 1497-1500

Nagy, F., Morelli, G., Fraley, R.T., Rogers, S.G. and Chua, N.-H. (1985) Photoregulated expression of a pea *rbcS* gene in leaves of transgenic plants. *EMBO J. 4* : 3063-3068

Naidu, B.P., Aspinall, D. and Paleg, L.G. (1992) Variability in proline-accumulating ability of barley (*Hordeum vulgare* L.) cultivars induced by vapor pressure deficit. *Plant Physiol. 98* : 716-722

Naidu, B.P., Paleg, L.G., Aspinall, D., Jennings, A.C. and Jones, G.P. (1990) Rate of imposition of water stress alters the accumulation of nitrogen-containing solutes by wheat seedlings. *Aust. J. Plant Physiol. 17* : 653-664

Naidu, B.P., Paleg, L.G., Aspinall, D., Jennings, A.C., and Jones, G.P. (1991) Amino acid and glycine betaine accumulation in cold-stressed wheat seedlings. *Phytochemistry. 30* : 407-409

Naik, G.P. and Joshi, G.V. (1983) Ineffectual role of proline metabolism in salt-stressed sugarcane leaves. *Proc. Indian Acad. Sci. 92* : 265-272

Nash, D., Paleg, L.G. and Wiskich, J.T. (1982) Effect of proline, betaine and some other solutes on the heat stability of mitochondrial enzymes. *Aust. J. Plant Physiol. 9* : 45-57

Nelson, D.E., Raghathema, K.G., Singh, N.K., Hasegawa, P.M. and Bressan, R.A. (1992) Analysis of structure and transcriptional activation of an osmotin gene. *Plant Mol. Biol. 19* : 577-588

Nemec, S. and Meredith, F.I. (1981) Amino acid content of leaves in mycorrhizal and non-mycorrhizal citrus rootstocks. *Ann. Bot. 47* : 351-358

Neuhaus, G., Neuhaus-Uri, G., Katagiri, F., Seipel, K. and Chua, N.-H. (1994) Tissue-specific expression of *as-1* in transgenic tobacco. *Plant Cell 6* : 827-834

Nikolopolous, D. and Manetas, Y. (1991) Compatible solutes and *in vitro* stability of *Salsola soda* enzymes : Proline incompatibility. *Phytochemistry 30* : 411-413

Nir, I, Poljakoff-Mayber, A. and Klein, S. (1971) The effect of water stress on mitochondria of root cells : A biochemical and cytochemical study. *Plant Physiol. 45* : 173-177

Noguchi, M., Koiwai, A. and Tamaki, A. (1966) Studies on nitrogen metabolism in tobacco plants. VII. Δ^1--pyrroline-5-carboxylate reductase from tobacco leaves. *Agric. Biol. Chem. 30* : 452-456

Nover, L. (1989) General stress response and the interaction of stresses. *In* Nover, L., Neumann, D. and Scharf, K.-D. (eds) Heat Shock and Other Stress Response Systems of Plants. Springer-Verlag, Berlin, pp.78-85

Nuti-Ronchi, V., Caligo, M.A., Nozzolini, M. and Luccarini, G. (1984) Stimulation of carrot somatic embryogenesis by proline and serine. *Plant Cell Rep. 3* : 210-214

Ober, E.S. and Sharp, R.E. (1994) Proline accumulation in maize (*Zea mays* L.) primary roots at low water potentials. *Plant Physiol. 105* : 981-987

Odell, J.T., Nagy, F. and Chua, N.-H. (1985) Identification of DNA sequences required for activity of the cauliflower mosaic virus 35S promoter. *Nature 313* : 810-812

Oeda, K., Salinas, J. and Chua, N.-H. (1991) A tobacco bZip transcription activator (TAF-1) binds to a G-box-like motif conserved in plant genes. *EMBO J. 10* : 1793-1802

Okubo, K., Hori, N., Matoba, R., Niiyama, T., Fukushima, A., Kojima, Y. and Matsubara, K. (1992) Large scale cDNA sequencing for analysis of quantitative and qualitative aspects of gene expression. *Nature Genet. 2* : 173-179

Omori, K., Suzuki, S.-I., Imai, Y. and Komatsubara, S. (1991) Analysis of the *Serratia marcescens proBA* operon and feedback control of proline biosynthesis. *J. Gen. Micro. 137* : 509-517

O'Regan, B.P., Cress, W.A. and Van Staden, J. (1993) Root growth, water relations, abscisic acid and proline levels of drought-resistant and drought-sensitive maize cultivars in response to water stress. *S. Afr. J. Bot. 59* : 98-104

Orser, C.S., Goodner, B.W., Johnston, M., Gelvin, S.B. and Csonka, L.N. (1988) The *Escherichia coli proB* gene corrects the proline auxotrophy of *Saccharomyces cerevisiae pro1* mutants. *Mol. Gen. Genet. 212* : 124-128

Osmond, C.B. (1978) Crassulacean acid metabolism : A curiosity in context. *Annu. Rev. Plant Physiol. 29*: 379-414

Osmond, C.B., Austin, M.D., Berry, J.A., Billings, W.D., Boyer, J.S., Dacey, J.W., Nobel, P.S., Smith, S.D. and Winner, W.E. (1987) Stress physiology and the distribution of plants. *Bioscience 37* : 38-48

Ow, D.A., Jacobs, J.D. and Howell, S.H. (1987) Functional regions of the cauliflower mosaic virus 35S RNA promoter determined by use of the firefly luciferase gene as a reporter of promoter activity. *Proc. Natl. Acad. Sci. USA 84* : 4870-4874

Pace, N.R., Olsen, G.J. and Woese, C.R. (1986) Ribosomal RNA phylogeny and the primary lines of evolutionary descent. *Cell 45* : 325-326

Padgett, R.A., Grabowski, P.J., Konarska, M.M., Seiler, S. and Sharp, P.A. (1986) Splicing of messenger RNA precursors. *Annu. Rev. Biochem. 55* : 1119-1150

Pahlich, E., Kerres, R. and Jäger, H.J. (1983) Influence of water stress on the vacuole/extravacuole distribution of proline in protoplasts of *Nicotiana rustica*. *Plant Physiol. 72* : 590-591

Pai, E.F. and Schulz, G.E. (1983) The catalytic mechanism of glutathione reductase as derived from X-ray diffraction analyses of reaction intermediates. *J. Biol. Chem. 258* : 1751-1757

Pai, E.F., Karplus, P.A. and Schulz, G.E. (1988) Crystallographic analysis of the binding of NADPH, NADPH fragments and NADPH analogues to glutathione reductase. *Biochemistry 27* : 4465-4474

Paleg, L.G. and Aspinall, D. (1981) Proline accumulation : Physiological aspects. *In* Paleg, L.G. and Aspinall, D. (eds) The Physiology and Biochemistry of Drought Resistance in Plants. Academic Press, Sydney, pp.206-241

Paleg, L.G., Stewart, C.R. and Bradbeer, J.W. (1984) Proline and glycine betaine influence protein solvation. *Plant Physiol. 75* : 974-978

Paleg, L.G., Douglas, T.J., Van Daal, A. and Keech, D.B. (1981) Proline, betaine and other organic solutes protect enzymes against heat denaturation. *Aust. J. Plant Physiol. 8* : 107-114

Palfi, G and Palfi, Z. (1982) A rapid method for the determination of fertility of maize pollen with the proline-isatin reaction. *Maydica 27* : 107-111

Pandey, R. and Ganapathy, P.S. (1985) The proline enigma : NaCl tolerant and NaCl-sensitive callus lines of *Cicer arietinum*. *Plant Sci. 40* : 13-17

Pandey, U.K. and Srivastava, R.D.L. (1990) Salinity index in relation to nitrate reductase activity and proline accumulation in paddy genotypes. *Indian J. Plant Physiol. 32* : 175-183

Pardo, J.M. and Serrano, R. (1989) Structure of a plasma membrane H^+-ATPase gene from the plant *Arabidopsis thaliana*. *J. Biol. Chem. 264* : 8557-8562

Park, Y.S., Kwak, J.M., Kwon, O.-Y., Kim, Y.S., Lee, D.S., Cho, M.J., Lee, H.H. and Nam, H.G. (1993) Generation of expressed sequence tags of random root cDNA clones of *Brassica napus* by single-run partial sequencing. *Plant Physiol. 103* : 359-370

Parker, J., Szabo, V., Staskawicz, B., Lister, C., Dean, C., Daniels, M. and Jones, J. (1993) Phenotypic characterisation and molecular mapping of the Arabidopsis tha*liana* locus *RPP5* determining resitance to *Peronospora parasitica*. *Plant J. 4* : 821-831

Parker, R. Siliciano, P.G. and Guthrie, C. (1987) Recognition of the TACTAAC box during mRNA splicing in yeast involves base pairing to the U2-like snRNA. *Cell 49* : 229-239

Parthier, B. (1989) Hormone-induced alterations in plant gene expression. *Biochem. Biophys. Pflanzen 185*: 289-314

Paul, M.J. and Cockburn, W. (1989) Pinitol, a compatible solute in *Mesembryanthemum crystallinum* L.? *J. Exp. Bot. 40* : 1093-1098

Peisach, J. and Strecker, H.J. (1962) The interconversion of glutamate and proline V. The reduction of Δ^1-pyrroline-5-carboxylic acid to proline. *J. Biol. Chem. 237* : 2255-2260

Pelham, H.R.B. and Bienz, M. (1982) A synthetic heat-shock promoter element confers heat-inducibility on the herpes simplex virus thymidine kinase gene. *EMBO J. 1* : 1473-1477

Perino, C. and Côme, D. (1991) Physiological and metabolical study of the germination phases in apple embryo. *Seed Sci. Technol. 19* : 1-14

Pesci, P. (1987) ABA-induced proline accumulation in barley leaf segments : Dependence on protein synthesis. *Physiol. Plant 71* : 287-291

Pesci, P. and Beffagna, N. (1984) Inhibiting effect of fusicoccin on abscisic acid induced proline accumulation in barley leaves. *Plant Sci. Lett. 36* : 7-12

Pesci, P. and Beffagna, N. (1985) Effect of weak acids on proline accumulation in barley leaves : a comparison between abscisic acid and isobutyric acid. *Plant Cell Environ.* **8** : 129-133

Pesci, P., Palumbo, M., Boggini, G. and Soave, C. (1992) Proline accumulation induced by abscisic acid and KCl in barley and *durum* wheat genotypes. *J. Genet. Breed.* **46** : 29-34

Petrie, A.K.H. and Wood, J.G. (1958) Studies on the nitrogen metabolism of plants. III. On the effect of water content on the relationship between proteins and amino acids. *Ann. Bot.* **33** : 887-898

Phang, J.M. (1985) The regulatory functions of proline and pyrroline-5-carboxylic acid. *Curr. Topics Cell. Regul.* **25** : 91-132

Phang, J.M., Downing, S.J., Yeh, G.C., Smith, R.J., Williams, J.A. and Hagedorn, C.H. (1982) Stimulation of hexose monophosphate pathway by pyrroline-5-carboxylate in cultured cells. *J. Cell Physiol.* **110** : 255-261

Platt, T. (1986) Transcription termination and the regulation of gene expression. *Annu. Rev. Biochem.* **55** : 339-372

Pollard, A. and Wyn Jones, R.G. (1979) Enzyme activities in concentrated solutions of glycinebetaine and other solutes. *Planta* **144** : 291-298

Pongor, S., Hátsági, Z., Degtyarenko, K., Fábián, P., Skerl, V., Hegyi, H., Murvai, J. and Bevilacqua, V. (1994) The SBASE protein domain library, release 3.0 : A collection of annotated protein sequence segments. *Nucl. Acids Res.* **22** : 3610-3615

Popot, J.-L. and Engelman, D.M. (1990) Membrane protein models : Possibilities and Probabilities. *In* Bradshaw, R.A. and Purton, M. (eds) Proteins : Form and Function. Elsevier, Cambridge, pp. 147-151

Porath, E. and Poljakoff-Mayber, A. (1964) Effect of salinity on metabolic pathways in pea root tips. *Israel. J. Bot.* **13** : 115-121

Poulin, R., Larochelle, J. and Hellebust, J.A. (1987) The regulation of amino acid metabolism during hyperosmotic stress in *Acanthamoeba castellanii*. *J. Exp. Zool.* **243** : 365-378

Poulsen, C., Fluhr, R., Kaufmann, J.M., Boutry, M. and Chua, N.-H. (1986) Characterisation of an *rbcS* gene from *Nicotiana plumbaginifolia* and expression of an *rbcS*-CAT chimeric gene in homologous and heterologous nuclear backgrounds. *Mol. Gen. Genet.* **205** : 193-200

Prevelige, P. and Fasman, G.D. (1989) Chou-Fasman prediction of the secondary structure of proteins : The Chou-Fasman-Prevelige Algorithm. *In* Fasman, G.D. (ed) Prediction of Protein Structure and the Principles of Protein Conformation. Plenum Press, New York, pp 391-416

Radin, J.W., Parker, L.L. and Guinn, G. (1982) Water relations of cotton plants under nitrogen deficiency. V. Environmental control of abscisic acid accumulation and stomatal sensitivity to abscisic acid. *Plant Physiol. 70* : 1066-1070

Radojevic, L. (1988) Plant regeneration of *Aesculus hippocastanum* L. (Horse chestnut) through somatic embryogenesis. *J. Plant Physiol. 132* : 322-326

Radojevic, L. and Subotic, A. (1992) Plant regeneration of *Iris setosa* Pall. through somatic embryogenesis. *J. Plant Physiol. 139* : 690-696

Rajagopal, V. and Anderson, A.S. (1978) Does abscisic acid influence proline accumulation in stressed leaves? *Planta 143* : 85-88

Ranjeva, R. and Boudet, A.M. (1987) Phosphorylation of proteins in plants : Regulatory effects and potential involvement in stimulus response coupling. *Annu. Rev. Plant Physiol. 38* : 73-93

Ramachandran, G.N. and Sasisekharan, V. (1968) Conformation of polypeptides and proteins. *Adv. Protein Chem. 23* : 283-437

Ramshaw, J.A.M., Richardson, D.L., Mestyard, B.T., Brown, R.H., Richardson, M., Thompson, E.W. and Boulter, D. (1972) The time and origin of the flowering plants by using amino acid sequence data of cytochrome *c*. *New Phytol. 71* : 773-779

Rao, S.T. and Rossman, M.G. (1973) Comparison of supersecondary structures in proteins. *J. Mol. Biol. 76* : 241-250

Ratzkin, B. and Roth, J. (1978) Cluster of genes controlling proline degradation in *Salmonella typhimurium*. *J. Bacteriol. 133* : 744-754

Raven, J.A. (1985) Regulation of pH and generation of osmolarity in vascular plants : A cost benefit analysis in relation to efficiency of use of energy, nitrogen and water. *New Phytol. 101* : 25-77

Rayapati, P.J. and Stewart, C.R. (1991) Solubilization of a proline dehydrogenase from maize (*Zea mays* L.) mitochondria. *Plant Physiol. 95* : 787-791

Rayapati, P.J., Stewart, C.R. and Hack, E. (1989) Pyrroline-5-carboxylate reductase is in pea (*Pisum sativum* L.) leaf chloroplasts. *Plant Physiol. 91* : 581-586

Raz, V. and Fluhr, R. (1993) Ethylene signal is transduced via protein phosphorylation events in plants. *Plant Cell 5* : 523-530

Reaney, M.J.T., Gusta, L.V., Abrams, S.R. and Robertson, A.J. (1989) The effects of abscisic acid, kinetin, and gibberellin on freezing tolerance in smooth bromegrass (*Bromus inermis*) cell suspensions. *Can. J. Bot. 67* : 3640-3646

Rédei, G.P. (1992) A heuristic glance at the past of *Arabidopsis* genetics. *In* Koncz, C., Chua, N-H. and Schell, J. (eds) Methods in *Arabidopsis* Research. World Scientific, Singapore, pp.1-15

Rédei, G.P. amd Koncz, C. (1992) Classical mutagenesis. *In* Koncz, C., Chua, N-H. and Schell, J. (eds) Methods in *Arabidopsis* Research. World Scientific, Singapore, pp.16-83

Reed, R.H., Davison, I.R., Chudek, J.A. and Foster, R. (1985) The osmotic role of mannitol in the *Phaeophyta* : An appraisal. *Phycologist 24* : 35-47

Reinbothe, S., Mollenhauer, B. and Reinbothe, C. (1994) JIPs and RIPs : The regulation of plant gene expression by jasmonates in response to environmental cues and pathogens. *Plant Cell 6* : 1197-1209

Rena, A.B. and Splittstoesser, W.E. (1974) The metabolism of proline in the cotyledons of pumpkin (*Cucurbita moschata*). *Plant Cell Physiol. 15* : 681-686

Rena, A.B. and Spittstoesser, W.E. (1975) Proline dehydrogenase and pyrroline-5-carboxylate reductase from pumpkin cotyledons. *Phytochemistry 14* : 657-661

Rhodes, D. and Hanson, A.D. (1993) Quaternary ammonium and tertiary sulphonium compounds in higher plants. *Annu. Rev. Plant Physiol. Plant Mol. Biol. 44* : 357-384

Rhodes, D., Handa, S. and Bressan, R.A. (1986) Metabolic changes associated with adaptation of plant cells to water stress. *Plant Physiol. 82* : 890-903

Ricard, B., Couée, I., Raymond, P., Saglio, P.H., Saint-Ges, V. and Pradet, A. (1994) Plant metabolism under hypoxia and anoxia. *Plant Physiol. Biochem. 32* : 1-10

Riccardi, G., Cella, R., Camerino, G. and Ciferri, O. (1983) Resistance to azetidine-2-carboxylic acid and sodium chloride tolerance in carrot cell cultures and *Spirulina platensis*. *Plant Cell Physiol. 24* : 1073-1978

Richards, R.A. and Thurling, N. (1979) Genetic analysis of drought stress response in rapeseed (*Brassica campestris* and *B. napus*). III. Physiological characters. *Euphytica 28* : 755-759

Richardson, J. (1981) The anatomy and taxonomy of protein structure. *Adv. Protein Chem. 34* : 167-339

Richardson, J.S. and Richardson, D.C. (1989) Principles and patterns of protein conformation. *In* Fasman, G.D. (ed) Prediction of Protein Structure and the Principles of Protein Conformation. Plenum Press, New York, pp. 1-99

Richardson, J., Getzoff, E. and Richardson, D. (1978) The β-bulge : A common small unit of nonrepetitive protein structure. *Proc. Natl. Acad. Sci. USA 75* : 2574-2578

Roberts, L.W. and Baba, S. (1968) Effect of proline on wound vessel member formation. *Plant Cell Physiol. 9* : 353-360

Rodriguez, M.M. and Heyser, J.W. (1988) Growth inhibition by exogenous proline and its metabolism in saltgrass (*Distichlis spicata*) suspension cultures. *Plant Cell Rep. 7* : 305-311

Rogozinska, J. and Flasinki, S. (1987) The effect of nutrient salt and osmotic stresses on proline accumulation in oilseed rape plants. *Acta Physiol. Plant. 9* : 61-72

Rohlf, F.J. (1992) NTSYSpc User's Manual. Exeter Software, Oxford, pp. 15-4, 16-7

Rose, G.D. (1978) Prediction of chain turns in globular proteins on a hydrophobic basis. *Nature 272* : 586-590

Rose, G.D. and Dworkin, J.E. (1989) The hydrophobicity profile. *In* Fasman, G.D. (ed) Prediction of Protein Structure and the Principles of Protein conformation. Plenum Press, New York, pp. 625-633

Rose, G.D. and Seltzer, J.P. (1977) A new algorithm for finding the peptide chain turns in globular proteins. *J. Mol. Biol. 113* : 153-164

Rose, G.D., Young, W.B., and Gierasch, L.M. (1983) Interior turns in globular proteins. *Nature 304* : 654-657

Rose, G.D., Geselowitz, A., Lesser, G., Lee, R. and Zehfus, M. (1985) Hydrophobicity of amino acid residues in globular proteins. *Science 229* : 834-838

Rossi, J.J., Vender, J., Berg, C.M. and Coleman, W.H. (1977a) Partial purification and some properties of Δ^1-pyrroline-5-carboxylate reductase from *Escherichia coli*. *J. Bacteriol. 129* : 108-114

Rossi, J.J., Vender J. and Berg, C.M. (1977b) proline excretion in *Escherichia coli* : A comparison of an *argD*$^+$ strain and a proline-excreting *argD*$^-$ derivative. *Biochem. Genet. 15* : 287-296

Rossman, M.G. and Argos, P. (1981) Protein folding. *Annu. Rev. Biochem. 50* : 497-533

Rossman, M.G., Moras, D. and Olsen, K.W. (1974) Chemical and biological evolution of a nucleotide-binding domain. *Nature 250* : 194-197

Rudolf, A.S., Crowe, J.H. and Crowe, L.M. (1986) Effects of three stabilising agents - proline, betaine and trehalose - on membrane phospholipids. *Arch. Biochem. Biophys. 245* : 134-143

Rushlow, K.E., Deutch, A.H. and Smith, C.J. (1984) Identification of a mutation that relieves gamma-glutamyl kinase from allosteric feedback inhibition by proline. *Gene 39* : 109-112

Sachs, M.M. and Ho, T.-H. D. (1986) Alteration of gene expression during environmental stress in plants. *Annu. Rev. Plant Physiol. 37* : 363-376

Sachs, M.M., Freeling, M. and Okimoto, R. (1980) The anaerobic proteins of maize. *Cell 20* : 761-767

Sachs, M.M., Dennis, E.S., Gerlach, W.L. and Peacock, W.J. (1986) Two alleles of maize alcohol dehydrogenase 1 have 3' structural and poly(A) addition polymorphisms. *Genetics 113* : 449-467

Sanderson, K.E. and Hartman, P.E. (1978) Linkage map of *Salmonella typhimurium*, Edition V. *Microbiol. Rev. 42* : 471-519

Sanfacon, H., Brodmann, P. and Hohn, T. (1991) A dissection of the cauliflower mosaic virus polyadenylation signal. *Genes Dev. 5* : 141-149

Sanger, F., Nicklen, S. and Coulson, A.R. (1977) DNA sequencing with chain-termination inhibitors. *Proc. Natl. Acad. Sci. USA 74* : 5463-5467

Santarius, K.A. (1973) The protective effect of sugars on chloroplast membranes during temperature stress and water stress and its relationship to frost, dessication and heat resistance. *Planta 113* : 105-114

Saradhi, A. and Saradhi, P.P. (1991) Proline accumulation under heavy metal stress. *J. Plant Physiol. 138* : 554-558

Savioz, A., Jeenes, D.J., Kocher, H.P. and Haas, D. (1990) Comparison of *proC* and other housekeeping genes of *Pseudomonas aeruginosa* with their counterparts in *Escherichia coli*. *Gene 86* : 107-111

Scarpulla, R.C. and Soffer, R.L. (1978) Membrane-bound proline dehydrogenase from *Escherichia coli* : Solubilisation, purification and characterisation. *J. Biol. Chem. 253* : 5997-6001

Schindler, U., Sans, N. and Schröder, J. (1989) Ornithine cyclodeaminase from octopine Ti plasmid Ach5 : identification, DNA sequence, enzyme properties, and comparison with gene and enzyme from nopaline Ti plasmid C58. *J. Bacteriol. 171* : 847-854

Schmitt, N. and Dizengremel, P. (1989) Effect of osmotic stress on mitochondria isolated from etiolated mung bean and sorghum seedlings. *Plant Physiol. Biochem. 27* : 17-26

Schobert, B. (1977a) The influence of water stress on the metabolism of diatoms. II. Proline accumulation under different conditions of stress and light. *Z. Pflanzenphysiol. 85* : 451-465

Schobert, B. (1977b) Is there an osmotic regulatory mechanism in algae and higher plants? *J. Theor. Biol. 68* : 17-26

Schobert, B. and Tschesche, H. (1978) Unusual solution properties of proline and its interactions with proteins. *Biochem. Biophys. Acta. 541* : 270-277

Schoner, S. and Krause, G.H. (1990) Protective systems against active oxygen species in spinach : Response to cold acclimation in excess light. *Planta 180* : 383-389

Schuler, G.D., Altschul, S.F. and Lipman, D.J. (1991) A workbench for multiple alignment construction and analysis. *Proteins : Struct. Func. Genet. 9* : 180-190

Schwab, K.B. and Gaff, D.F. (1990) Influence of compatible solutes on soluble enzymes from dessication-tolerant *Sporobolus pyramidalis*. *J. Plant Physiol. 137* : 208

Scrutton, N., Berry, A. and Perham, R.N. (1990) Redesign of the coenzyme specificity of a dehydrogenase by protein engineering. *Nature 343* : 38-43

Seitz, E.W. and Hochster, R.M. (1964) Free proline in normal and in crown-gall tumour tissue of tomato and tobacco. *Life Sci. 3* : 1033-1037

Sells, G.D. and Koepe, D.E. (1981) Oxidation of proline by mitochondria isolated from water-stressed maize shoots. *Plant Physiol. 68* : 1058-1063

Setter, T.L. and Greenway, H. (1979) Growth and osmoregulation of *Chlorella emersonii* in NaCl and neutral osmotica. *Aust. J. Plant Physiol. 6* : 47-60

Shah, C.B. and Loomis, R.S. (1965) Ribonucleic acid and protein metabolism in sugar beet during drought. *Physiol. Plant. 18* : 240-254

Shah, S.H., Wainwright, S.J. and Merrett, M.J. (1990) The interaction of sodium and calcium chlorides and light on growth, potassium nutrition and proline accumulation in callus cultures of *Medicago sativa* L. *New Phytol. 116* : 37-45

Sharp, P.A. (1987) Splicing of messenger RNA precursors. *Science 235* : 766-771

Sharp, P.M. and Devine, K.M. (1989) Codon usage and gene expression in *Dictyostelium discoideum* : Highly expressed genes do 'prefer' optimal codons. *Nucl. Acids Res. 17* : 5029-5039

Sharp, R.E. and Davies, W.J. (1985) Root growth and water uptake by maize plants in drying soil. *J. Exp. Bot. 170* : 1441-1456

Sharp, P.M., Crowe, E., Higgins, D.G., Shields, D.C., Wolfe, K.H. and Wright, F. (1988) Codon usage patterns in *Escherichia coli, Saccharomyces cerevisiae, Schizosaccharomyces pombe, Drosophila melanogaster* and *Homo sapiens* : A review of the considerable within-species diversity. *Nucl. Acids Res. 16* : 8207-8211

Sheridan, W.F. (1988) Maize developmental genetics : Genes of morphogenesis. *Annu. Rev. Genet. 22* : 353-385

Shetty, K. and Asano, Y. (1991) The influence of organic nitrogen sources on the induction of embryogenic callus in *Agrostis alba* L. *J. Plant Physiol. 139* : 82-85

Shetty, K., Shetty, G.A., Nakazakim, Y., Yoshioka, Y., Asano, Y. and Oosawa, K. (1992) Stimulation of benzyladenine-induced *in vitro* shoot organogenesis in *Cucumis melo* L. by proline, salicylic acid and aspirin. *Plant Sci. 84* : 193-199

Shetty, K. and McKersie, B.D. (1993) Proline, thioproline and potassium mediated stimulation of somatic embryogenesis in alfalfa (*Medicago sativa* L.) *Plant Sci. 88* : 185-193

Sinclair, T.R. and Ludlow, M.M. (1985) Who taught plants thermodynamics? The unfulfilled potential of plant water potential. *Austr. J. Plant Physiol. 12* : 213-217

Singh, T.N., Aspinall, D. and Paleg, L.G. (1972) Proline accumulation and varietal adaptability to drought in barley : a potential metabolic measure of drought resistance. *Nature (New Biol.) 236* : 188-190

Singh, T.N., Aspinall, D., Paleg, L.G. and Boggess, S.F. (1973) Stress metabolism. II. Changes in proline concentration in excised plant tissues. *Aust. J. Biol. Sci. 26* : 57-63

Singh, N.K., Handa, A.K., Hasegawa, P.M. and Bressan, R.A. (1985) Proteins associated with adaptation of cultured tobacco cells to NaCl. *Plant Physiol. 79* : 126-137

Skriver, K. and Mundy, J. (1990) Gene expression in response to abscisic acid and osmotic stress. *Plant Cell 2* : 503-512

Skubatz, H., Meeuse, B.J.D. and Bendich, A.J. (1989) Oxidation of proline and glutamate by mitochondria of the inflorescence of voodoo lily (*Sauromatum guttatum*). *Plant Physiol. 91* : 530-535

Small, W.C. and Jones, M.E. (1990) Pyrroline-5-carboxylate dehydrogenase of the mitochondrial matrix of rat liver : Purification, physical and kinetic characteristics. *J. Biol. Chem. 265* : 18668-18672

Smirnoff, N. (1993) The role of active oxygen in the response of plants to water deficit and dessication. *New Phytol. 125* : 27-58

Smirnoff, N. and Cumbes, Q.J. (1989) Hydroxyl radical scavenging activity of compatible solutes. *Phytochemistry. 28* : 1057-1060

Smith, J.E. (1962) Mitochondrial transamination in sunflower hypocotyls. *Biochem. Biophys. Acta 57* : 183-185

Smith, C.E. and Brown, A.W. (1981) The regulation of oat coleoptile phosphoenolpyruvate carboxylase and malic enzyme activities by H^+ and metabolites. Kinetic evidence for and against a metabolic pH-stat. *Can J. Bot. 59* : 1397-1404

Smith, M.E. and Greenberg, D.M. (1956) Characterisation of an enzyme reducing pyrroline-5-carboxylate to proline. *Nature 177* : 1130

Smith, M.E. and Greenberg, D.M. (1957) Preparation and properties of partially purified glutamic semialdehyde reductase. *J. Biol. Chem. 226* : 317-327

Smith, C.J., Deutch, A.H. and Rushlow, K.E. (1984) Purification and characterisation of a γ-glutamyl kinase involved in *Escherichia coli* proline biosynthesis. *J. Bacteriol. 157* : 545-551

Smith, J.K., Schloss, J.V. and Mazur, B.J. (1989) Functional expression of plant acetolactate synthase genes in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA 86* : 4179-4183

Smith, R.J., Downing, S.J., Phang, J.M., Lodato, R.F. and Aoki, T.T. (1980) Pyrroline-5-carboxylate synthase activity in mammalian cells. *Proc. Natl. Acad. Sci. USA 77* : 5221-5225

Sneath, P.H.A. and Sokal, R.R. (1973) Numerical Taxonomy. Freeman, San Francisco, p. 573

Snustad, D.P., Hunsberger, J.P., Chereskin, B.M. and Messing, J. (1988) Maize glutamine synthetase cDNAs : Isolation by direct genetic selection in *E. coli*. *Genetics 120* : 1111-1124

Somerville, C.R. and Ogren, W.L. (1982) Isolation of photorespiration mutants in *Arabidopsis thaliana*. *In* Edelman, M., Hallick, R.B. and Chua, N.-H. (eds) Methods in Chloroplast Biology. Elsevier Biomedical Press, Amsterdam, pp. 129-138

Southern, E.M. (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol. 98* : 503-517

Spector, M.P., Aliabadi, Z., Gonzales, T. and Forster, J.W. (1986) Global control in *Salmonella typhimurium* : Two-dimensional electrophoretic analysis of starvation-, anaerobiosis-, and heat shock-inducible proteins. *J. Bacteriol. 168* : 420-424

Splittstoesser, W.E. and Fowden, L. (1973) Ornithine transaminase from *Cucurbita maxima* cotyledons. *Phytochemistry 12* : 785-790

Splittstoesser, S.A. and Splittstoesser, W.E. (1973) Pyrroline-5-carboxylate reductase from *Cucurbita* cotyledons. *Phytochemistry 12* : 1565-1568

Stackebrandt, E. (1988) Phylogenetic relationships vs. phenotypic diversity : How to achieve a phylogenetic classification system of the eubacteria. *Can. J. Microbiol. 34* : 552-556

States, D.J. and Gish, W. (1994) Combined use of sequence similarity and codon bias for coding region identification. *J. Comp. Biol. 1* : 39-50

Steponkus, P.L. (1984) Role of the plasma membrane in freezing injury and cold acclimation. *Annu. Rev. Plant Physiol. 35* : 543-584

Stewart, C.R. (1978) Role of carbohydrates in proline accumulation in wilted barley leaves. *Plant Physiol.* ***61*** : 775-778

Stewart, C.R. (1980) The mechanism of abscisic acid-induced proline accumulation in barley leaves. *Plant Physiol.* ***66*** : 230-233

Stewart, C.R. (1981) Proline accumulation : Biochemical aspects. *In* Paleg, L.G. and Aspinall, D. (eds) The Physiology and Biochemistry of Drought Resistance in Plants. Academic Press, Sydney, pp. 243-259

Stewart, I. (1962) The effect of minor element deficiencies on free amino acids in citrus leaves. *Proc. Am. Soc. Hort. Sci.* ***81*** : 244-249

Stewart, C.R. and Lai, E.Y. (1974) Δ^1-pyrroline-5-carboxylic acid dehydrogenase in mitochondrial preparations from plant seedlings. *Plant Sci. Lett.* ***3*** : 173-181

Stewart, C.R. and Lee, J.A. (1974) The role of proline accumulation in halophytes. *Planta* ***120*** : 279-289

Stewart, C.R. and Boggess, S.F. (1977) The effect of wilting on the conversion of arginine, ornithine, and glutamate to proline in bean leaves. *Plant Sci. Lett.* ***8*** : 147-153

Stewart, C.R. and Voetberg, G. (1985) Relationship between stress-induced ABA and proline accumulations and ABA-induced proline accumulation in excised barley leaves. *Plant Physiol.* ***79*** : 24-27

Stewart, C.R. and Voetberg, G. (1987) Abscisic acid accumulation is not required for proline accumulation in wilted leaves. *Plant Physiol.* ***83*** : 747-791

Stewart, C.R. Morris, C.J. and Thompson, J.F. (1966) Changes in amino acid content of excised leaves during incubation. II. Role of sugar in the accumulation of proline in wilted leaves. *Plant Physiol.* ***41*** : 1585-1590

Stewart, C.R., Voetberg, G. and Rayapati, P.J. (1986) The effects of benzyladenine, cycloheximide, and cordycepin on wilting-induced abscisic acid and proline accumulations and abscisic acid- and salt-induced proline accumulation in barley leaves. *Plant Physiol.* ***82*** : 703-707

Stewart, C.R., Boggess, S.F., Aspinall, D. and Paleg, L.G. (1977) Inhibition of proline oxidation by water stress. *Plant Physiol.* 59 : 930-932

Stitt, M. and ap Rees, T. (1980) Carbohydrate breakdown by chloroplasts of *Pisum sativum*. *Biochem. Biophys. Acta 627* : 131-143

Storey, R. and Wyn Jones, R.G. (1977) Quaternary ammonium compounds in plants in relation to salt resistance. *Phytochemistry 16* : 447

Strecker, H.J. (1957) The interconversion of glutamic acid and proline. I. The formation of Δ^1-pyrroline-5-carboxylic acid from glutamic acid in *Escherichia coli*. *J. Biol. Chem. 225* : 825-834

Strecker, H.J. (1965) Purification and properties of rat liver ornithine δ-transaminase. *J. Biol. Chem. 240* : 1225-1230

Strecker, H.J. (1971) The preparation of animal proline oxidase (rat liver), and its use for the preparation of Δ^1-pyrroline-5-carboxylate. *Methods Enzymol. 17* : 251-263

Stryer, L. (1988) Biochemistry (Third Edition). W.H. Freeman & Co., New York

Stuart, D.A. and Strickland, S.G. (1984) Somatic embryogenesis from cell cultures of *Medicago sativa* L. II. The interaction of amino acids with ammonium. *Plant Sci. 34* : 175-181

Sugiura, M. (1992) The chloroplast genome. *In* Schilperoort, R. and Dure, L. (eds) 10 Years Plant Molecular Biology, Kluwer, Dordrecht, pp.149-168

Sugiura, M. and Kisumi, M. (1985) Osmoregulation in proline-producing strain of *Serratia marcescens*. *J. Gen. Microbiol. 131* : 2515-2520

Sullivan, M.L. and Green, P.J. (1993) Post-transcriptional regulation of nuclear-encoded genes in higher plants : The roles of mRNA stability and translation. *Plant Mol. Biol. 23* : 1091-1104

Szoke, A., Miao, G.-H., Hong, Z. and Verma, D.P.S. (1992) Subcellular location of Δ^1-pyrroline-5-carboxylate reductase in root/nodule and leaf of soybean. *Plant Physiol. 99* : 1642-1649

Tabor, S. and Richardson, C.C. (1987) DNA sequence analysis with a modified bacteriophage T7 DNA polymerase. *Proc. Natl. Acad. Sci. USA 84* : 4767-4771

Tabor, S. and Richardson, C.C. (1989) Selective inactivation of the exonuclease activity of bacteriophage T7 DNA polymerase by *in vitro* mutagenesis. *J. Biol. Chem. 264* : 6447-6458

Tal, M., Heikin, H. and Dehan, K. (1978) Salt tolerance in the wild relatives of the cultivated tomato : Responses of callus tissues of *Lycopersicon esculentum, L. peruvianum* and *Solanum pennellii* to high salinity. *Z. Pflanzenphysiol.* ***86*** : 231-240

Tal, M., Katz, A., Heikin, H. and Dehan, K. (1979) Salt tolerance of the wild relatives of the cultivated tomato : Proline accumulation in *Lycopersicon esculentum* Mill., *L. peruvianum* Mill. and *Solanum pennelli* cor. treated with NaCl and polyethylene glycol. *New Phytol.* ***82*** : 349-355

Tarczynski, M.C., Jensen, R.G. and Bohnert, H.J. (1993) Stress protection of transgenic tobacco by production of the osmolyte mannitol. *Science* ***259*** : 508-510

Taylor, A.A. and Stewart, C.R. (1981) Tissue and subcellular localisation of enzymes of arginine metabolism in *Pisum sativum*. *Biochem. Biophys. Res. Commun.* ***101*** : 1281-1289

Theillet, C., Depeyroux, F., Fizman, M., Reigner, R. and Exnault, R. (1982) Influence of the excision shock on the protein metabolism of *Vicia faba* L. meristematic root cells. *Planta* ***155*** : 478-485

Thomas, J.C., De Armond, R.L. and Bohnert, H.J. (1992) The influence of NaCl on growth, proline, and phosphoenolpyruvate carboxylase levels in *Mesembryanthemum crystallinum* suspension cultures. *Plant Physiol.* ***98*** : 626-631

Thompson, J.F., Stewart, C.R. and Morris, C.J. (1966) Changes in amino acid content of excised leaves during incubation. I. the effect of water content of leaves and atmospheric oxygen level. *Plant Physiol.* ***41*** : 1578-1584

Timpa, J.D., Burke, J.J., Quisenberry, J.E. and Wendt, C.W. (1986) Effects of water stress on the organic acid and carbohydrate compositions of cotton plants. *Plant Physiol.* ***82*** : 724-728

Tingey, S.V., Tsai, F.-Y., Edwards, J.W., Walker, E.L. and Coruzzi, G.M. (1988) Chloroplast and cytosolic glutamine synthetase are encoded by homologous nuclear genes which are differentially expressed *in vivo*. *J. Biol. Chem.* ***263*** : 9651-9657

Tolbert, N.E. (1981) Metabolic pathways in peroxisomes and glyoxysomes. *Annu. Rev. Biochem.* ***50*** : 133-157

Tomenchok, D.M. and Brandriss, M.C. (1987) Gene-enzyme relationships in the proline biosynthetic pathway of *Saccharomyces cerevisiae*. *J. Bacteriol.* ***169*** : 5364-5372

Tonegawa, S., Maxam, A.M., Tizard, R., Bernard, O. and Gilbert, W. (1978) Sequence of a mouse germ-line gene for a variable region of an immunoglobulin light chain. *Proc. Natl. Acad. Sci. USA 75* : 14185-1489

Traut, T.W. (1986) Do exons code for structural and functional units in proteins? *Proc. Natl. Acad. Sci. USA 85* : 2944-2948

Treichel, S. (1975) The effect of NaCl on the concentration of proline in different halophytes. *Z. Pflanzenphysiol. 76* : 56-68

Treichel, S. (1986) The influence of NaCl on Δ^1-pyrroline-5-carboxylate reductase in proline-accumulating suspension cultures of *Mesembryanthemum nodiflorum* and other halophytes. *Physiol. Plant. 67* : 173-181

Trigiano, R.N. and Conger, B.V. (1987) Regulation of growth and somatic embryogenesis by proline and serine in suspension cultures of *Dactylis glomerata*. *J. Plant Physiol. 130* : 49-55

Tsaneva, I.R. and Weiss, B. (1990) *soxR*, a locus governing a superoxide response regulon in *Escherichia coli* K-12. *J. Bacteriol. 172* : 4197-4205

Tully, R.E., Hanson, A.D. and Nelsen, C.E. (1979) Proline accumulation in water-stressed barley leaves in relation to translocation and the nitrogen budget. *Plant Physiol. 63* : 518-523

Turner, N.C. and Jones, M.M. (1981) Turgor maintenance by osmotic adjustment : A review and evaluation. *In* Paleg, L.G. and Aspinall, D. (eds) Physiology and Biochemistry of Drought Resistance in Plants. Academic Press, Sydney, pp. 15-23

Turpin, D.H. and Weger, H.G. (1990) Interactions between photosynthesis, respiration and nitrogen assimilation. *In* Dennis, D.T. and Turpin, D.H. (eds) Plant Physiology, Biochemistry and Molecular Biology. Longmans, London, p. 428-429

Uchimiya, H., Kidou, S., Shimazaki, T., Aotsuka, S., Takamatsu, S., Nishi, R., Hashimoto, H., Matsubayashi, Y., Kidou, N., Umeda, M. and Kato, A. (1992) Random sequencing of cDNA libraries reveals a variety of expressed genes in cultured cells of rice (*Oryza sativa* L.) *Plant J. 2* : 1005-1009

Umbarger, H.E. (1987) Biosynthesis of the branched chain amino acids. *In* Neidhardt, F.C., Ingraham, J.L., Low, K.B., Magasanik, B., Schaechter, M. and Umbarger, H. (eds) *Escherichia coli* and *Salmonella typhimurium* : Cellular and Molecular Biology, Vol. 1. American Society for Microbiology, Washington, D.C., pp. 352-367

Urao, T., Katagiri, T., Mizoguchi, T., Yamaguchi-Shinozaki,K., Hayashida, N. and Shinozaki, K. (1994) Two genes that encode Ca^{2+}-dependent protein kinases are induced by drought and high-salt stresses in *Arabidopsis thaliana*. *Mol. Gen. Genet.* ***244*** : 331-340

Valle, D., Downing, S.J. and Phang, J.M. (1973) Proline inhibition of pyrroline-5-carboxylate reductase : Differences in enzymes obtained from animal and tissue culture sources. *Biochem. Biophys. Res. Commun.* ***54*** : 1418-1424

Valle, D., Blaese, R.M. and Phang, J.M. (1975) Increased sensitivity of lymphocyte Δ^1-pyrroline-5-carboxylate reductase to inhibition by proline with transformation. *Nature* ***253*** : 214-216

Valvekens, D., van Montagu, M. and van Lijsebetens, M. (1988) *Agrobacterium tumefaciens*-mediated transformation of *Arabidopsis thaliana* root explants by using kanamycin selection. *Proc. Natl. Acad. Sci. USA* ***85*** : 5536-5540

Van Bogelen, R.A., Kelley, P.M. and Neidhardt, F.C. (1987) Differential induction of heat shock, SOS, and oxidative stress regulons and accumulation of nucleotides in *Escherichia coli*. *J. Bacteriol.* ***169*** : 26-32

Van Rensburg, L., Krüger, G.H.J. and Krüger, H. (1993) Proline accumulation as drought tolerance selection criterion : Its relationship to membrane integrity and chloroplast ultrastructure in *Nicotiana tabacum* L. *J. Plant Physiol.* ***141*** : 188-194

Van Santen, V.L. and Spritz, R.A. (1987) Splicing of plant pre-mRNAs in animal systems and *vice versa*. *Gene* ***56*** : 253-265

Van Volkenburgh, E. and Boyer, J.S. (1985) Inhibitory effects of water deficit on maize leaf elongation. *Plant Physiol.* ***77*** : 190-194

Vartanian, N., Marcotte, L. and Giraudat, J. (1994) Drought rhizogenesis in *Arabidopsis thaliana* : Differential responses of hormonal mutants. *Plant Physiol.* ***104*** : 761-767

Venekamp, J.H. (1989) Regulation of cytosol acidity in plants under conditions of drought. *Physiol. Plant.* ***76*** : 112-117

Venekamp, J.H. and Koot, J.T.M. (1984) The distribution of free amino acids, especially proline, in the organs of field bean plants, *Vicia faba* L., during development in the field. *J. Plant Physiol. 116* : 343-349

Venekamp, J.H. and Koot, J.T.M. (1988) The sources of free proline and asparagine in field bean plants, *Vicia faba* L. during and after a short period of water withholding. *J. Plant Physiol. 132* : 102-109

Venekamp, J.H., Lampe, J.E.M. and Koot, J.T.M. (1989) Organic acids as sources for drought-induced proline synthesis in field bean plants, *Vicia faba* L. *J. Plant Physiol. 133* : 654-659

Verbruggen, N. and Jacobs, M. (1987) *Arabidopsis* mutants resistant to proline analogues : Isolation and preliminary characterisation. *Arab. Inf. Serv. 23* : 15-23

Verbruggen, N., Villarroel, R. and Van Montagu, M. (1993) Osmoregulation of a pyrroline-5-carboxylate reductase gene in *Arabidopsis thaliana*. *Plant Physiol. 103* : 771-781

Verma, D.P.S., Hu, C.-A.A., Delauney, A.J., Miao, G.H. and Hong, Z. (1992) Deciphering proline biosynthesis pathways in plants by direct, trans-, and co-complementation in bacteria. *In* Singh, B.K., Flores, H.E. and Shannon, J.C. (eds) Biosynthesis and Molecular Regulation of Amino Acids in Plants. American Society of Plant Physiologists, Rockville, MD, pp.128-138

Vernon, D.M. and Bohnert, J.J. (1992) A novel methyl transferase induced by osmotic stress in the facultative halophyte *Mesembryanthemum crystallinum*. *EMBO J. 11* : 2077-2085

Vernon, D.M., Ostrem, J.A. and Bohnert, H.J. (1993) Stress perception and response in a facultative halophyte : The regulation of salinity-induced genes in *Mesembryanthemum crystallinim*. *Plant Cell Environ. 16* : 437-444

Vierling, E. (1991) The roles of heat shock proteins in plants. *Annu. Rev. Plant Physiol. Plant Mol. Biol. 42* : 579-620

Visser, B. (1994) Technical aspects of drought tolerance. *Biotech. Dev. Monitor 18* : 5

Voetberg, G.S. and Stewart, C.R. (1984) Steady state proline levels in salt-shocked barley leaves. *Plant Physiol. 76* : 567-576

Voetberg, G.S. and Sharp, R.E. (1991) Growth of the maize primary root at low water-potentials. III. Role of increased proline deposition in osmotic adjustment. *Plant Physiol. 96* : 125-130

Vogel, H.J. and Davis, B.D. (1952) Glutamic γ-semialdehyde and Δ^1-pyrroline-5-carboxylic acid, intermediates in the biosynthesis of proline. *J. Am. Chem. Soc. 74* : 109-112

Vogel, R.H. and Kopac, M.J. (1959) Glutamic γ-semialdehyde in arginine and proline biosynthesis of *Neurospora* : A mutant tracer analysis. *Biochem. Biophys. Acta 36* : 505-510

Von Heijne, G. and Nishikawa, K. (1991) Chloroplast transit peptides : The perfect random coil? *FEBS Lett. 278* : 1-3

Von Heijne, G., Steppuhn, J. and Herrmann, R.G. (1989) Domain structure of mitochondrial and chloroplast targetting peptides. *Eur. J. Biochem. 180* : 535-545

Wadano, A. (1980) Proline synthesis from glutamate in the mitochondria isolated from a blowfly, *Aldrichina grahami*. *Experientia 36* : 1028-1029

Wadman van Schravendijk, H. and Van Andel, O.M. (1985) Interdependence of growth, water relations and abscisic acid level in *Phaseolus vulgaris* during water-logging. *Physiol. Plant. 63* : 215-220

Wagner, K.G. and Backer, A.I. (1992) Dynamics of nucleotides in plants studied on a cellular basis. *Int. Rev. Cytol. 134* : 1-84

Wakabayashi, Y. and Jones, M.E. (1983) Pyrroline-5-carboxylate synthesis from glutamate by rat intestinal mucosa. *J. Biol. Chem. 258* : 3865-3872

Wakabayashi, Y., Henslee, J.G. and Jones, M.E. (1983) Pyrroline-5-carboxylate synthesis from glutamate by rat intestinal mucosa : Subcellular localisation and temperature stability. *J. Biol. Chem. 258* : 3873-3882

Wakabayashi, Y., Yamada, E., Hasegawa, T. and Yamada, R. (1991) Enzymolgical evidence for the indispensibility of small intestine in the synthesis of arginine from glutamate. I. Pyrroline-5-carboxylate synthase. *Arch. Biochem. Biophys. 291* : 1-8

Walker, M.A. and Dumbroff, E.B. (1981) Effects of salt stress on abscisic acid and cytokinin levels. *Z. Pflanzenphysiol. 101* : 661-667

Walsh, K.A., Ericsson, L.H., Parmelee, D.C. and Titani, K. (1981) Advances in protein sequencing. *Annu. Rev. Biochem. 50* : 261-284

Walter, M.H. and Hahlbrock, K. (1985) Synthesis of characteristic proteins in nutrient-depleted cell suspension cultures of parsley. *Planta 166* : 194-200

Walton, E.F., Clark, C.J. and Boldingh, H.L. (1991) Effect of hydrogen cyanamide on amino acid profiles in kiwifruit buds during bud-break. *Plant Physiol. 97* : 1256-1259

Wample, R.L. and Bewley, J.D. (1975) Proline accumulation in flooded and wilted sunflower and the effects of benzyladenine and abscisic acid. *Can. J. Bot. 53* : 2893-2896

Wang, C.Y., Cheng, S.H. and Kao, C.H. (1982) Senescence in rice leaves. VII. Proline accumulation in senescing excised leaves. *Plant Physiol. 69* : 1348-1349

Wang, S.-S. and Brandriss, M.C. (1986) Proline utilisation in *Saccharomyces cerevisiae* : Analysis of the cloned *PUT1* gene. *Mol. Cell. Biol. 6* : 2638-2645

Wang, S.-S. and Brandriss, M.C. (1987) Proline utilisation in *Saccharomyces cerevisiae* : Sequence, regulation, and mitochondrial localisation of the *PUT1* gene product. *Mol. Cell Biol. 7* : 4431-4440

Waterston, R., Martin, C., Craxton, M., Huynh, C., Coulson, A., Hillier, L., Durdin, R., Green, P., Shownkeen, R., Halloran, N., Metzstein, M., Hawkins, T., Wilson, R., Berks, M., Du, Z., Thomas, K., Thierry-Mieg, J. and Sulston, J. (1992) A survey of expressed genes in *Caenorhabditis elegans*. *Nature Genet. 1* : 114-123

Webster, P.L. (1980) "Stress" protein synthesis in pea root meristem cells? *Plant Sci. Lett. 20* : 141-145

Webster, P.L. and Van't Hof, J. (1973) Polyribosomes in proliferating and non-proliferating root meristem cells. *Am. J. Bot. 60* : 117-121

Weeda, E.C., de Kort, A.D. and Beenakkers, A.M.T. (1979) Fuels for energy metabolism in the Colorado potato beetle, *Leptinotarsa decemlineata* Say. *J. Insect Physiol. 25* : 951-955

Weigel, D., Alvarez, J., Smyth, D.R., Yanofsky, M.F. and Meyerowitz, E.M. (1992) *LEAFY* controls floral meristem identity in *Arabidopsis*. *Cell 69* : 843-859

Weimberg, R., Lerner, H.R. and Poljakoff-Mayber, A. (1982) A relationship between potassium and proline accumulation in salt-stressed *Sorghum bicolor*. *Physiol. Plant. 55* : 5-10

Weimberg, R., Lerner, H.R. and Poljakoff-Mayber, A. (1984) Changes in growth and water-soluble salt concentrations in *Sorghum bicolor* stressed with sodium and potassium salts. *Physiol. Plant. 62* : 472-480

Weiner, S.J., Kollman, P.A., Case, D.A., Singh, U.C., Ghio, C., Alagona, G., Profeta, S. and Weiner, P. (1984) A new force field for molecular mechanical simulation of nucleic acids and proteins. *J. Am. Chem. Soc. 106* : 765-784

Weisemann, J.M. and Matthews, B.J. (1993) Identification and expression of a cDNA from *Daucus carota* encoding a bifunctional aspartokinase-homoserine dehydrogenase. *Plant Mol. Biol. 22* : 301-312

Weretilnyk, E.A. and Hanson, A.D. (1990) Molecular cloning of a plant betaine-aldehyde dehydrogenase, an enzyme implicated in adaptation to salinity and drought. *Proc. Natl. Acad. Sci. USA 87* : 2745-2749

Westgate, M.E. and Boyer, J.S. (1985) Osmotic adjustment and the inhibition of leaf, root, stem and silk growth at low water potentials in maize. *Planta 164* : 540-549

Wetlaufer, D.B. (1973) Nucleation, rapid folding, and globular intrachain regions in proteins. *Proc. Natl. Acad. Sci. U.S.A. 70* : 697-701

Whatmore, A.M., Chudek, J.A. and Reed, R.H. (1990) The effects of osmotic upshock on the intracellular solute pools of *Bacillus subtilis*. *J. Gen. Microbiol. 136* : 2527-2535

White, J.A. and Scandalios, J.G. (1988) Molecular biology of intracellular protein trafficking. *Physiol. Plant. 74* : 397-408

Whitehead, T.P., Thorpe, G.H.G., Carter, T.J.N., Groucutt, C. and Kricka, L.J. (1983) Enhanced luminescence procedure for sensitive determination of peroxidase-labelled conjugates in immunoassay. *Nature 305* : 158-159

Wiebauer, K., Herrero, J.-J. and Filipowicz, W. (1988) Nuclear pre-mRNA processing in plants : Distinct modes of 3' splice-site selection in plants and animals. *Mol. Cell. Biol. 8* : 2042-2051

Wierenga, R.K., Terpstra, P. and Hol, W.G.J. (1986) Prediction of the occurrence of the ADP-binding βαβ-fold in proteins, using an amino acid sequence fingerprint. *J. Mol. Biol. 187* : 101-107

Wilkinson, D.G. (1992) The theory and practice of *in situ* hybridisation. *In* Wilkinson, D.G. (ed) *In situ* Hybridisation : A Practical Approach. IRL Press, Oxford, pp.1-13

Williamson, C.L. and Slocum, R.D. (1992) Molecular cloning and evidence for osmoregulation of the Δ^1-pyrroline-5-carboxylate reductase (*proC*) gene in pea (*Pisum sativum* L.) *Plant Physiol. 100* : 1464-1470

Wilson, A.C., Carlson, S.S. and White, J.J. (1977) Biochemical evolution. *Annu. Rev. Biochem. 46* : 573-639

Wilson, B.J., Gray, A.C. and Matthews, B.J. (1991) Bifunctional protein in carrot contains both aspartokinase and homoserine dehydrogenase activities. *Plant Physiol. 97* : 1323-1328

Winicov, I., Waterborg, J.H., Harrington, R.E. and McCoy, T.J. (1989) Messenger RNA induction in cellular salt tolerance of alfalfa (*Medicago sativa*). *Plant Cell Rep. 8* : 6-11

Wise, R.R. and Naylor, A.W. (1987) Chilling-enhanced photooxidation. the peroxidative destruction of lipids during chilling injury to photosynthesis and ultrastructure. *Plant Physiol. 83* : 272-277

Withers, L.A. and King, P.J. (1979) Proline : a novel cryoprotectant for the freeze preservation of cultured cells of *Zea mays* L. *Plant Physiol. 64* : 675-678

Woese, C.R. and Fox, G.F. (1977) Phylogenetic structure of the prokaryotic domain : The primary kingdoms. *Proc. Natl. Acad. Sci. USA 74* : 5088-5090

Wolfenden, R., Andersen, L., Cullis, P. and Southgate, C. (1981) Affinities of amino acid side chains for solvent water. *Biochemistry 20* : 849-855

Wrench, P., Brady, C.J. and Hinde, R.W. (1980) Interaction of slicing and osmotic stress on proline metabolism in Jerusalem artichoke tuber tissue. *Aust. J. Plant Physiol. 7* : 149-157

Wrench, P., Wright, L., Brady, C.J. and Hinde, R.W. (1977) The source of carbon for proline synthesis in osmotically stressed artichoke tuber slices. *Aust. J. Plant Physiol. 4* : 703-711

Wright, F. (1990) The 'effective number of codons' used in a gene. *Gene 87* : 23-29

Wyn Jones, R.G. and Storey, R. (1978) Salt stress and comparitive physiology in the Gramineae. II. Glycinebetaine and proline accumulation in two salt and water-stressed barley cultivars. *Aust. J. Plant Physiol. 5* : 817-829

Wyn Jones, R.G. and Gorham, J. (1983) Aspects of salt and drought tolerance in higher plants. *In* Kosnge, T., Meredith, C.P. and Hollander, A. (eds) Genetic Engineering of Plants : An Agricultural Perspective. Plenum Press, New York, pp. 355-

Yamaguchi-Shinozaki, K. and Shinozaki, K. (1994) A novel *cis*-acting element in an *Arabidopsis* gene is involved in responsiveness to drought, low-temperature, or high-salt stress. *Plant Cell 6* : 251-264

Yancey, P.H., Clark, M.E., Hand, S.C., Bowlus, R.D. and Somero, G.N. (1982) Living with water stress : Evolution of osmolyte systems. *Science 217* : 1214-1222

Yanofsky, M.F., Ma, H., Bowman, J.L., Drews, G.N., Feldman, K.A. and Meyerowitz, E.M. (1990) The protein encoded by the *Arabidopsis* homeotic gene *agamous* resembles transcription factors. *Nature 346* : 35-39

Yeh, G.C. and Phang, J.M. (1981) The stimulation of purine nucleotide production by pyrroline-5-carboxylic acid in human erythrocytes. *Biochem. Biophys. Res. Comm. 103* : 118-124

Yeh, G.C. and Phang, J.M. (1988) Stimulation of phosphoribosyl pyrophosphate and purine nucleotide production by pyrroline-5-carboxylate in human erythrocytes. *J. Biol. Chem. 263* : 13083-13089

Yeh, G.C., Harris, S.C. and Phang, J.M. (1981) Pyrroline-5-carboxylate reductase in human erythrocytes : A comparison of differential regulation. *J. Clin. Invest. 67* : 1042-1046

Yura, T. and Vogel, H.J. (1959) Pyrroline-5-carboxylate reductase of *Neurospora crassa* : partial purification and some properties. *J. Biol. Chem. 234* : 335-342

Zaug, A.J., Grabowski, P.J. and Cech, T.R. (1983) Autocatalytic cyclisation of an excised intervening sequence RNA is a cleavage-ligation reaction. *Nature 301* : 578-583

Zhang, H.Q., Croes, A.F. and Linskens, H.F. (1982) Protein synthesis in germinating pollen of *Petunia* : Role of proline. *Planta 154* : 199-203

Zrenner, R. and Stitt, M. (1991) Comparison of the effect of rapidly and gradually developing water stress on carbohydrate metabolism in spinach leaves. *Plant Cell Environ. 14* : 939-946

APPENDIX 1

Nutrient solution for growth of *A. thaliana* (Somerville and Ogren, 1982)

Stock Solutions	ml stock / l of irrigation solution
1 M KNO_3	5.0
1 M KH_2PO_4 , pH 5.6	2.5
1 M $MgSO_4$	2.0
1 M $Ca(NO_3)_2$	2.0
1.8 % Sequestrene	2.8
Micronutrient Mix *	1.0

*Micronutrient Mix

70.0 mM H_3BO_3
14.0 mM $MnCl_2$
0.5 mM $CuSO_4$
1.0 mM $ZnSO_4$
0.2 mM $NaMoO_4$
10.0 mM NaCl
0.01 mM $CoCl_2$

Proline assays

Acid Ninhydrin Reagent

1.25 g of ninhydrin is dissolved by warming gently in 30 ml of glacial acetic acid and 20 ml of 6 M phosphoric acid with agitation. The reagent is prepared fresh before use.

Glycerol Freezer Store Solution

65% (v/v) glycerol

100 mM $MgSO_4$

25 mM Tris-HCl (pH 8.0)

Media for bacterial growth

All quantities provided are those needed to prepare 1 l of medium. Media are solidified with the addition of 15 g of agar per litre.

LB broth

10 g tryptone

10 g yeast extract

10 g NaCl

Superbroth

20 g yeast extract

35 g tryptone

5 g NaCl

1 ml 5M NaOH

DNA extraction techniques

Plasmid Miniprep

STE buffer

0.1 M NaCl

10 mM Tris-HCl (pH 8.0)

1 mM EDTA

Solution I

50 mM glucose

10 mM Tris-HCl (pH 8.0)

1 mM Na_2EDTA

Solution II

0.2 M NaOH

1% SDS

4 M potassium acetate - 2 M acetic acid

Mix 4 vol of 5 M potassium acetate with 1 vol of 10 M (57%) glacial acetic acid.

88% isopropanol - 0.2 M potassium acetate

Mix 1 vol of 5 M potassium acetate, 2 vol of dH_2O and 22 vol of isopropanol.

Preparation of DNase-free RNase

DNase-free RNase is prepared as described by Felliciello and Chinali (1993). RNase A (Boehringer Mannheim) is dissolved in water (1.5 $mg.ml^{-1}$). After addition of 0.9 vol of 0.2 M HCl, 1 ml aliquots of the solution are transferred into 2 ml microfuge tubes and incubated for 5 min in boiling water. Tubes are chilled on ice and RNase solution is neutralised at 4°C by adding to each tube at 2 min intervals and with rapid mixing 0.1 ml of 0.2 M Tris-HCl (pH 7.6) and 5 x 0.05 ml aliquots of 80 mM NaOH. The final solution, containing 1 $mg.ml^{-1}$ of RNase in 15 mM NaCl/15 mM Tris-HCl (pH 7.6), is stored in aliquots at -20°C.

TE buffer

10 mM Tris-HCl (pH 8.0)

1 mM Na_2EDTA

M13 precipitation solution

3.5 M ammonium acetate (pH 7.5)

20% PEG-6000

Equilibration of phenol

Crystalline phenol is melted at 68°C and 8-hydroxyquinoline added to a final concentration of 0.1%. An equal volume of 0.1 M Tris-HCl (pH 9.5) is added to the melted phenol and the mixture stirred on a magnetic stirrer for 20 min. The two phases are left to separate. The pH of the aqueous phase is determined using universal indicator paper (Whatman). The aqueous phase is removed, fresh 0.1 M Tris-HCl (pH 9.5) added and the processes of mixing and separation of the two phases repeated until the aqueous phase reaches pH 8.0. Phenol is stored in a light-tight container at 4°C under 0.1 vol of 0.1 M Tris-HCl (pH 8.0).

Genomic DNA Extraction

Plant Genomic DNA Extraction Buffer

7.0 M urea

0.3 M NaCl

0.05 M Tris-HCl (pH 8.0)

0.02 M Na_2EDTA

2% sarcosyl

Electrophoresis

50X TAE Buffer Stock

121 g of Tris, 28.5 ml of glacial acetic acid and 50 ml of 0.5 M Na_2EDTA (pH 8.0) are made up to a final volume of 500 ml. When diluted appropriately, this provides a working solution of 0.04 M Tris-acetate, 0.002 M Na_2EDTA (pH 8.0).

10X TBE Buffer Stock

108 g of Tris, 55 g of boric acid and 9.3 g of $Na_2EDTA.2H_2O$ are made up to 1 l with dH_2O. This provides a stock solution of pH 8.3.

10X Loading Buffer for DNA Agarose Gels

50 mM NaOH

1 mM Na_2EDTA

2.5% glycerol

0.25% bromophenol blue

Molecular Weight Marker III

MWMIII (Boehringer Mannheim) comprises fragments generated by digestion of wild-type lambda phage with *Hin*dIII and *Eco*RI. Fragment sizes are 21 226, 5 148, 4 973, 3 530, 2 027, 1 904, 1 584, 1 375, 947, 831, 564 and 125 bp.

Quantification of DNA

The concentration of DNA is calculated using the formula:

$$\mu g\ DNA.ml^{-1} = A_{260} \times \text{dilution factor} \times \Delta E$$

where $\Delta E = 50\ \mu g.ml^{-1}$ for double-stranded DNA

$\Delta E = 40\ \mu g.ml^{-1}$ for single-stranded DNA

Sequencing

All reagents used are part of the Sequenase™ Version 2.0 DNA Sequencing Kit (US Biochemical Corp.)

5X Sequenase Buffer

200 mM Tris-HCl, pH 7.5

100 mM $MgCl_2$

250 mM NaCl

5X dGTP Labelling Mix

7.5 μM dGTP

7.5 μM dCTP

7.5 μM dTTP

Enzyme Dilution Buffer

10 mM Tris-HCl, pH 7.5

5 mM DTT

0.5 $mg.ml^{-1}$ BSA

Sequenase Version 2.0 T7 DNA Polymerase

13 $units.\mu l^{-1}$ in

20 mM KPO_4, pH 7.4

1 mM DTT

0.1 mM EDTA

50% glycerol

ddGTP Termination Mix

80 μM dGTP, 80 μM dATP, 80 μM dCTP, 80 μM dTTP

8 μM ddGTP, 50 mM NaCl

ddATP Termination Mix

80 µM dGTP, 80 µM dATP, 80 µM dCTP, 80 µM dTTP

8 µM ddATP, 50 mM NaCl

ddTTP Termination Mix

80 µM dGTP, 80 µM dATP, 80 µM dCTP, 80 µM dTTP

8 µM ddTTP, 50 mM NaCl

ddCTP Termination Mix

80 µM dGTP, 80 µM dATP, 80 µM dCTP, 80 µM dTTP

8 µM ddCTP, 50 mM NaCl

Stop Solution

95% formamide

20 mM EDTA

0.05% bromophenol blue

0.05% xylene cyanol FF

Southern hybridisation

Denaturation solution

1.5 M NaCl

0.5 M NaOH

Neutralisation Solution

1.5 M NaCl

0.5 M Tris-HCl (pH 7.5)

20X SSC

0.3 M Na3 citrate

3 M NaCl

Adjust to pH 7.0

Primary wash buffer

6M urea

0.4% SDS

0.1x SSC

Secondary wash buffer

2x SSC

In situ hybridisation

DIG dNTP Labelling Mixture (Boehringer Mannheim)

1 mM dATP, 1 mM dCTP, 1 mM dGTP, 0.65 mM dTTP,

0.35 mM DIG-dUTP.

DIG Hybridisation buffer

50% deionised formamide

5X SSC

2% (w/v) blocking agent (Bochringer Mannheim)

0.1% sarcosine

0.02% SDS

PBS (pH 7.0)

130 mM NaCl

7 mM Na_2HPO_4

3 mM NaH_2PO_4

3x PBS is a three-times concentrated solution of PBS.

4% Paraformaldehyde solution

Four g of paraformaldehyde is dissolved in 100 ml of PBS (pH 7.0).

0.1 M Triethanolamine buffer

12.8 ml of triethanolamine is dissolved in 900 ml of dH_2O. The pH of the solution is adjusted to pH 8.0 using NaOH. The solution is made up to a final volume of 1 l with dH_2O.

Buffer 1

0.1 M maleic acid

0.15 M NaCl

Adjust to pH 7.5 with 5 M NaOH.

Buffer 2

1% blocking agent (Boehringer Mannheim)

Buffer 3

100 mM Tris-HCl (pH 9.5)

100 mM NaCl

APPENDIX 2

Numerical data used in the prediction of secondary structures of P5CRs from *Arabidopsis* (Verbruggen *et al.*, 1993), soybean (Delauney and Verma, 1990), pea (Williamson and Slocum, 1992), human (Dougherty *et al.*, 1992) and *E. coli* (Deutch *et al.*, 1982) are presented in this Appendix. Predictions were made using the methods of Chou and Fasman (1978a, 1978b) and Garnier *et al.* (1978). The numerical data analysed using these strategies was generated by the computer programs SEQAID II (D.D. Rhoads and D.J. Roufa, Kansas State University) and PREDICT7 (Càrmenes *et al.*, 1989).

i) Prediction of secondary structure using the method of Chou and Fasman (1978a, 1978b)

Pa and Pb values are the conformational parameters for each amino acid residue calculated by considering the relative frequency of a given amino acid within a given type of secondary structure in proteins of known structure and sequence (Chou and Fasman, 1974a). Pa and Pb are therefore measures of a given amino acid's preference to be found in α-helix or β-sheet respectively. The conformational parameter Pt reflects the preference of an amino acid to be found in a β-turn (Chou and Fasman, 1977).

In all analyses, six residues were averaged for α-helix formation (Pa avg), four residues were averaged for β-sheet formation (Pb avg) and four residues were averaged for the determination of β-turns (Pt avg). In the data presented in this Appendix, asterisks indicate when the hexa- or tetrapeptide average calculated from Pa, Pb or Pt values is equal to or exceeds 1.000.

As outlined by Chou and Fasman (1974b), the assignments listed below are made for the different amino acids participating in α-helix and β-sheet formation.

Ha assignments are : H, strong α former; h, α former; I, weak α former; i, α indifferent; b, α breaker; B, strong α breaker. I assignments are also given to Pro and Asn (near the N-terminal helix) as well as Arg (near the C-terminal helix).

Hb assignments are : H, strong β former; h, β former; I, weak β former; i, β indifferent; b, β breaker; B, strong β breaker. b assignment is also given to Trp (near the C-terminal β-region).

An abbreviated set of rules (Fasman, 1985) outlined below was used in all predictions. These are :

1) A cluster of four helical residues (H or h) out of six along the protein sequence initiates a helix. The helix segment is extended in both directions until at least two addjacent tetrapeptide breakers (Pa < 1.00) are reached. Proline cannot occur in the inner helix or at the C-terminal helical end but can occur within the first three residues at the N-terminal end. The inner helix is defined as one omitting the three helical end residues at both the amino and carboxyl ends. Any segment that is at least six residues long with an average Pa value greater than 1.03 and average Pa greater than Pb is predicted as helical.

2) A cluster of three β formers (H or h) out of five residues along the sequence will initiate a β-sheet. The β-sheet is propagated in both directions until terminated by a set of tetrapeptide breakers (average Pb < 1.00). Adjacent α-regions that have higher α- than β-potential (Pa > Pb), can also terminate β-propagation. Any segment with an average Pb value greater than 1.05 as well as an average Pb greater than Pa is predicted as β-sheet. β-sheet formation is unfavourable if the entire segment contains one or more β-sheet breakers (b or B) and less than one half β-sheet formers.

3) Tetrapeptides with Pt greater than 1.00 and an average Pt value greater than both Pa and Pb are predicted as β-turns.

4) Any segment containing overlapping α and β regions is helical if the average Pa is greater than the average Pb or β-sheet if the average Pb is greater than the average Pa.

5) Conformational parameters for coil are not employed. Coil is predicted by default.

ii) Prediction of secondary structure using the method of Garnier *et al.* (1978)

Prediction of secondary structure using the method of Garnier *et al.* (1978) was by computer-generated interpretation of the numerical data conducted by the computer program PREDICT7 (Cármenes *et al.*, 1989). The numerical parameters Ph, Ps, Ptu and Pco are probablility coefficients for the formation of α-helix, β-sheet, turns and coil structures respectively.

i) Numerical data used in the prediction of the secondary structure of P5CR from *Arabidopsis thaliana* (Verbruggen *et al.*, 1993).

residue	Pa	Pa Avg	Ha	Pb	Pb avg	Hb	Pt	Pt avg	Ph	Ps	Ptu	Pco
1 M	1.20	1.110*	h	1.67	1.187*	H	0.55	0.645	73	53	-63	-61
2 E	1.53	1.008*	H	0.26	0.925	B	0.82	0.872	46	55	-57	-94
3 I	1.00	0.995	I	1.60	1.260*	H	0.54	0.802	-6	62	-136	-63
4 L	1.34	1.083*	H	1.22	1.015*	h	0.57	1.057*	-35	18	-236	-35
5 P	0.59	0.992	B	0.62	0.952	b	1.56	1.075*	21	-38	-124	8
6 I	1.00	1.080*	I	1.60	0.862	H	0.54	0.890	22	-78	-166	27
7 P	0.59	1.092*	B	0.62	0.642	b	1.56	1.092*	133	-108	-129	-12
8 A	1.45	1.183*	H	0.97	0.807	I	0.64	0.862	145	-123	-20	10
9 E	1.53	1.030*	H	0.26	0.750	B	0.82	0.940	148	-85	-57	-44
10 S	0.79	0.962	I	0.72	1.097*	b	1.35	0.867	114	-27	-29	-20
11 F	1.12	0.997	h	1.28	1.120*	h	0.64	0.915	69	41	27	-34
12 K	1.07	0.898	I	0.74	1.120*	b	0.95	0.915	91	62	45	-3
13 V	1.14	0.962	h	1.65	1.335*	H	0.53	0.812	92	73	0	10
14 G	0.53	0.860	B	0.81	1.125*	I	1.54	1.065*	69	38	-18	-6
15 F	1.12	0.950	h	1.28	1.165*	h	0.64	0.840	76	6	-53	-51
16 I	1.00	0.963	I	1.60	1.047*	H	0.54	1.065*	61	-58	-96	-23
17 G	0.53	1.038*	B	0.81	0.832	i	1.54	1.167*	24	-132	-58	14
18 A	1.45	1.205*	H	0.97	1.047*	I	0.64	0.945	35	-173	-40	40
19 G	0.53	1.095*	B	0.81	1.047*	i	1.54	0.945	24	-182	-58	61
20 K	1.07	1.173*	I	0.74	0.910	b	0.95	0.765	68	-123	-60	52
21 M	1.20	1.237*	h	1.67	0.905	H	0.65	0.865	118	-62	-103	24
22 A	1.45	1.168*	H	0.97	0.887	I	0.64	0.837	140	-48	-120	-10
23 E	1.53	1.015*	H	0.26	0.887	B	0.82	0.837	153	15	-152	-99
24 S	0.79	0.950	I	0.72	1.047*	b	1.35	0.895	149	8	-139	-105
25 I	1.00	1.008*	I	1.60	1.070*	H	0.54	0.942	139	-28	-131	-98
26 A	1.45	1.083*	H	0.97	1.082*	I	0.64	0.940	118	2	-100	-90
27 R	0.79	0.973	I	0.90	1.252*	i	1.05	0.912	79	4	-69	-77
28 G	0.53	0.930	B	0.81	1.270*	i	1.54	0.810	29	33	-28	-21
29 V	1.14	1.032*	h	1.65	1.247*	H	0.53	0.762	4	98	-30	0
30 V	1.14	1.065*	h	1.65	1.037*	H	0.53	1.015*	-71	163	-30	30
31 A	1.45	0.973	H	0.97	1.037*	I	0.64	1.015*	-150	172	40	50
32 S	0.79	0.830	I	0.72	1.100*	b	1.35	0.997	-214	178	51	60
33 G	0.53	0.820	B	0.81	1.075*	i	1.54	1.050*	-268	108	102	34
34 V	1.14	0.863	h	1.65	1.027*	H	0.53	1.055*	-261	73	60	85
35 L	1.34	0.840	H	1.22	0.777	h	0.57	1.300*	-278	63	-81	140
36 P	0.59	0.745	B	0.62	0.697	b	1.56	1.420*	-222	32	21	158
37 P	0.59	0.783	B	0.62	0.942	b	1.56	1.165*	-139	42	206	148
38 N	0.73	0.927	b	0.65	1.112*	b	1.51	1.097*	-91	29	187	96
39 R	0.79	0.995	I	0.90	1.250*	i	1.05	0.967	-49	44	76	-2
40 I	1.00	1.070*	I	1.60	1.267*	H	0.54	0.865	-24	162	22	-58
41 C	0.77	1.035*	I	1.30	1.280*	h	1.29	0.862	7	239	-1	-72
42 T	0.82	1.028*	I	1.20	1.132*	h	0.99	0.775	24	253	23	-3
43 A	1.45	1.115*	H	0.97	1.012*	I	0.64	0.865	20	222	5	5
44 V	1.14	0.995	h	1.65	0.932	H	0.53	1.082*	-16	148	-5	20
45 H	1.24	0.937	h	0.71	0.825	b	0.94	1.092*	-73	65	37	51
46 S	0.79	0.862	I	0.72	0.810	b	1.35	1.235*	-116	-27	76	105
47 N	0.73	0.893	b	0.65	0.855	b	1.51	1.160*	-106	-61	87	106
48 L	1.34	0.962	H	1.22	0.917	h	0.57	1.045*	-83	-72	89	95
49 N	0.73	0.925	b	0.65	0.812	b	1.51	1.260*	-81	-81	97	86
50 R	0.79	1.058*	I	0.90	1.062*	i	1.05	1.015*	-76	-31	71	48
51 R	0.79	1.058*	I	0.90	1.157*	i	1.05	0.912	-54	29	51	38
52 D	0.98	1.113*	I	0.80	0.997	I	1.43	0.855	-35	76	71	50
53 V	1.14	1.038*	h	1.65	0.977	H	0.53	0.835	4	153	5	30
54 P	1.12	1.038*	h	1.28	0.885	h	0.64	0.862	-4	146	-13	4
55 E	1.53	0.973	H	0.26	0.767	B	0.82	1.087*	-32	95	-12	-14
56 S	0.79	0.908	I	0.72	1.115*	b	1.35	1.015*	-81	23	26	10
57 F	1.12	0.963	h	1.28	1.097*	h	0.64	1.055*	-81	19	47	-1
58 G	0.53	0.908	B	0.81	1.190*	i	1.54	1.027*	-88	-92	67	44
59 V	1.14	0.957	h	1.65	1.307*	H	0.53	0.802	-49	-62	35	80
60 N	0.73	0.898	b	0.65	1.075*	b	1.51	1.007*	-61	-56	-3	86
61 V	1.14	1.032*	h	1.65	1.212*	H	0.53	0.877	-66	-42	-5	105
62 F	1.12	1.097*	h	1.28	0.980	h	0.64	1.082*	-54	-34	-5	79
63 S	0.79	1.100*	I	0.72	0.725	b	1.35	1.127*	-29	-102	-24	120
64 T	0.82	1.158*	I	1.20	0.610	h	0.99	0.995	19	-142	-52	57
65 T	0.79	1.200*	I	0.72	0.722	b	1.35	0.680	81	-122	-109	20
66 E	1.53	1.323*	H	0.26	0.955	B	0.82	0.675	113	-115	-117	-44
67 E	1.53	1.200*	H	0.26	1.075*	B	0.82	0.707	151	-95	-117	-69
68 V	1.14	1.108*	h	1.65	1.075*	H	0.53	0.707	170	-97	-85	-50
69 V	1.14	1.108*	h	1.65	0.842	H	0.53	0.912	200	-97	-65	-38
70 K	1.07	1.108*	I	0.74	0.630	b	0.95	1.137*	214	-123	-60	-53
71 E	1.53	1.097*	H	0.26	0.857	B	0.82	1.032*	201	-85	-82	-94
72 S	0.79	1.028*	I	0.72	1.205*	b	1.35	0.960	209	-72	-94	-95
73 D	0.98	1.028*	I	0.80	1.425*	I	1.43	0.757	198	-69	-84	-130
74 V	1.14	1.058*	h	1.65	1.545*	H	0.53	0.560	172	8	-105	-120
75 V	1.14	1.043*	h	1.65	1.312*	H	0.53	0.705	122	118	-120	-115
76 I	1.00	0.952	I	1.60	1.312*	H	0.54	0.765	76	157	-86	-83
77 P	1.12	0.980	h	1.28	1.097*	h	0.64	0.867	56	116	-63	-96
78 S	0.79	0.983	I	0.72	0.932	b	1.35	1.097*	41	28	-9	-20
79 V	1.14	1.042*	h	1.65	1.060*	H	0.53	1.002*	19	-47	-70	-23
80 K	1.07	1.030*	I	0.74	1.060*	b	0.95	1.002*	-7	-78	-175	-28
81 P	0.59	1.030*	B	0.62	1.287*	b	1.56	0.897	48	-68	-29	-12
82 Q	1.17	1.173*	h	1.23	1.317*	h	0.97	0.745	70	-63	34	0
83 V	1.14	1.168*	h	1.65	1.195*	H	0.53	0.740	99	-37	-75	-40
84 V	1.14	1.115*	h	1.65	1.025*	H	0.53	0.767	109	33	-110	-63
85 K	1.07	1.180*	I	0.74	1.025*	b	0.95	0.767	123	72	-105	-71
86 K	1.07	1.225*	I	0.74	1.140*	b	0.95	0.777	168	92	-135	-68
87 A	1.45	1.225*	H	0.97	1.020*	I	0.64	0.745	190	72	-167	-100
88 V	1.14	1.115*	h	1.65	1.082*	H	0.53	0.727	239	18	-220	-130
89 T	0.82	1.103*	I	1.20	0.855	h	0.99	0.832	252	-12	-222	-163

residue	Pa	Pa avg.	Ha	Pb	Pb avg.	Hb	Pt	Pt avg.	Ph	Ps	Ptu	Pco
90 E	1.53	1.190*	H	0.26	0.735	B	0.82	0.922	258	-30	-232	-199
91 L	1.34	1.067*	H	1.22	0.855	h	0.57	0.955	255	-67	-186	-198
92 K	1.07	1.022*	I	0.74	0.855	b	0.95	0.955	256	-98	-75	-128
93 S	0.79	0.965	i	0.72	0.850	b	1.35	1.055*	235	-152	-9	-33
94 K	1.07	1.012*	I	0.74	0.855	b	0.95	0.955	226	-143	0	-3
95 L	1.34	1.000*	H	1.22	0.832	h	0.57	1.095*	192	-137	4	20
96 S	0.79	1.000*	i	0.72	0.712	b	1.35	1.190*	141	-117	16	12
97 K	1.07	1.058*	I	0.74	0.932	b	0.95	0.987	91	-103	-15	-43
98 N	0.73	1.012*	b	0.65	1.052*	b	1.51	0.892	34	-51	-38	-62
99 K	1.07	1.080*	I	0.74	1.302*	b	0.95	0.647	11	32	-40	-58
100 I	1.00	1.143*	I	1.60	1.297*	H	0.54	0.747	16	117	-106	-68
101 L	1.34	1.218*	H	1.22	1.310*	h	0.57	0.745	57	173	-181	-100
102 V	1.14	1.083*	h	1.65	1.247*	H	0.53	0.762	94	198	-185	-120
103 S	0.79	1.060*	I	0.72	1.077*	b	1.35	0.790	124	183	-184	-100
104 V	1.14	1.107*	h	1.65	1.100*	H	0.53	0.837	149	153	-185	-105
105 A	1.45	1.140*	H	0.97	1.087*	I	0.64	0.840	140	92	-120	-75
106 A	1.45	1.020*	H	0.97	1.030*	I	0.64	0.917	130	2	-65	-40
107 G	0.53	0.942	B	0.81	1.092*	i	1.54	0.900	74	-47	-23	-16
108 I	1.00	1.077*	I	1.60	1.052*	H	0.54	0.892	44	-3	-41	-31
109 K	1.07	1.105*	I	0.74	0.852	b	0.95	1.115*	18	-3	-60	-3
110 L	1.34	1.102*	H	1.22	0.972	h	0.57	1.020*	-33	8	-56	25
111 N	0.73	1.148*	b	0.65	0.975	b	1.51	1.120*	-48	14	-13	46
112 D	0.98	1.158*	i	0.80	0.877	i	1.43	0.947	-32	-9	-9	40
113 L	1.34	1.083*	H	1.22	0.975	h	0.57	0.840	37	8	14	75
114 Q	1.17	1.055*	h	1.23	0.850	h	0.97	1.035	35	12	74	95
115 E	1.53	1.023*	H	0.26	0.745	B	0.82	1.177	41	-5	143	86
116 W	1.14	0.900	h	1.19	0.987	h	1.00	1.215*	35	-10	156	87
117 S	0.79	0.897	i	0.72	0.890	b	1.35	1.322*	34	-47	106	145
118 G	0.53	0.932	B	0.81	0.935	I	1.54	1.247*	17	-87	177	144
119 Q	1.17	0.975	h	1.23	1.052*	h	0.97	1.022*	43	-68	169	150
120 D	0.98	0.970	i	0.80	1.145*	I	1.43	0.915	-15	-39	156	145
121 R	0.79	1.007*	i	0.90	1.170*	I	1.05	0.820	-49	84	131	103
122 F	1.12	0.973	h	1.28	1.357*	h	0.64	0.690	-74	201	127	59
123 I	1.00	0.908	I	1.60	1.455*	H	0.54	0.692	-69	252	114	67
124 R	0.79	0.878	I	0.90	1.210*	i	1.05	0.947	-129	274	151	108
125 V	1.14	0.845	h	1.65	1.147*	H	0.53	1.062*	-171	193	100	120
126 M	1.20	0.897	h	1.67	1.035*	H	0.65	1.177*	-247	83	-32	164
127 P	0.59	0.938	B	0.62	0.772	b	1.56	1.405*	-232	-33	134	163
128 N	0.73	1.030*	b	0.65	0.860	b	1.51	1.175*	-241	-121	92	156
129 T	0.82	0.997	I	1.20	0.940	h	0.99	0.957	-191	-132	-122	107
130 P	0.59	1.115*	B	0.62	1.052*	b	1.56	0.842	-72	-103	-49	73
131 A	1.45	1.258*	H	0.97	1.100*	I	0.64	0.837	15	-33	-25	60
132 A	1.45	1.258*	H	0.97	0.922	I	0.64	0.882	115	22	-135	-16
133 V	1.14	1.148*	h	1.65	0.922	H	0.53	0.882	189	3	-185	-50
134 G	0.53	1.148*	B	0.81	0.752	i	1.54	0.910	169	-7	-148	-36
135 E	1.53	1.260*	H	0.26	0.730	B	0.82	0.862	198	0	-177	-104
136 A	1.45	1.137*	H	0.97	1.077*	I	0.64	0.790	193	-28	-190	-115
137 A	1.45	1.118*	H	0.97	1.252*	I	0.64	0.792	183	-13	-130	-70
138 S	0.79	0.965	i	0.72	1.190*	b	1.35	0.970	169	48	-99	-48
139 V	1.14	0.970	h	1.65	1.315*	H	0.53	0.775	147	108	-105	-55
140 M	1.20	0.868	h	1.67	1.105*	H	0.65	1.027*	101	118	-53	-31
141 S	0.79	0.910	i	0.72	0.987	b	1.35	1.112*	46	78	-9	40
142 L	1.34	0.915	H	1.22	1.010*	h	0.57	1.160*	-3	-12	7	80
143 G	0.53	0.947	B	0.81	0.947	i	1.54	1.177*	-71	-77	37	109
144 T	0.82	1.113*	I	1.20	1.045*	h	0.99	1.040*	-73	-87	13	147
145 G	0.53	1.140*	B	0.81	0.810	I	1.54	0.997	-71	-97	-35	144
146 A	1.45	1.140*	H	0.97	0.672	I	0.64	0.817	15	-48	-85	85
147 T	0.82	1.140*	I	1.20	0.630	h	0.99	1.015*	89	-17	-122	52
148 E	1.53	1.170*	H	0.26	0.532	B	0.82	1.152*	138	-60	-92	16
149 E	1.53	1.105*	H	0.26	0.710	B	0.82	1.107*	181	-95	-37	-24
150 D	0.98	1.092*	I	0.80	1.045*	i	1.43	1.037*	206	-119	1	-40
151 G	0.53	1.128*	B	0.81	1.257*	i	1.54	0.812	210	-142	-28	-41
152 A	1.45	1.263*	H	0.97	1.297*	I	0.64	0.587	276	-58	-100	-55
153 I	1.00	1.208*	I	1.60	1.472*	H	0.54	0.590	304	7	-196	-93
154 V	1.14	1.130*	h	1.65	1.377*	H	0.53	0.597	299	108	-240	-130
155 A	1.45	1.182*	H	0.97	1.285*	I	0.64	0.625	278	132	-220	-155
156 M	1.20	1.130*	h	1.67	1.245*	H	0.65	0.850	248	113	-198	-136
157 L	1.34	1.018*	H	1.22	1.070*	h	0.57	0.847	192	83	-181	-125
158 F	1.12	0.973	h	1.28	1.177*	h	0.64	0.837	171	26	-133	-131
159 G	0.53	0.953	B	0.81	1.060*	i	1.54	1.062*	172	-117	-78	-51
160 A	1.45	1.088*	H	0.97	1.042*	I	0.64	0.915	210	-103	-55	-35
161 V	1.14	1.025*	h	1.65	1.200*	H	0.53	0.890	184	-97	-115	-55
162 G	0.53	1.077*	B	0.81	1.092*	i	1.54	0.900	144	-102	-93	-44
163 K	1.07	1.152*	I	0.74	1.075*	b	0.95	0.752	176	-68	-125	-78
164 I	1.00	1.228*	I	1.60	1.132*	H	0.54	0.675	161	-48	-191	-68
165 L	1.34	1.240*	H	1.22	0.932	h	0.57	0.897	157	-52	-206	-63
166 K	1.07	1.217*	I	0.74	0.692	b	0.95	0.960	178	-93	-165	-93
167 A	1.45	1.225*	H	0.97	0.692	I	0.64	0.960	228	-123	-150	-105
168 D	0.98	1.147*	i	0.80	0.867	I	1.43	0.962	225	-134	-109	-80
169 E	1.53	1.225*	H	0.26	0.987	B	0.82	0.765	198	-90	-112	-107
170 K	1.07	1.160*	I	0.74	1.122*	b	0.95	0.917	211	-108	-85	-103
171 M	1.20	1.118*	h	1.67	1.180*	H	0.65	0.840	221	-22	-78	-81
172 P	1.12	1.007*	h	1.28	1.175*	h	0.64	0.810	209	16	-93	-136
173 D	0.98	1.043*	I	0.80	1.155*	I	1.43	0.897	193	31	-129	-130
174 A	1.45	1.012*	H	0.97	1.157*	I	0.64	0.925	150	27	-147	-120
175 V	1.14	0.858	h	1.65	1.220*	H	0.53	0.907	84	43	-100	-70
176 T	0.82	0.800	I	1.20	0.987	h	0.99	1.112*	-3	23	-32	-18
177 G	0.53	0.752	B	0.81	0.690	I	1.54	1.250*	-116	-7	52	74
178 L	1.34	0.762	H	1.22	0.867	h	0.57	1.202*	-168	3	114	130
179 S	0.79	0.780	I	0.72	0.765	b	1.35	1.445*	-224	-22	161	205
180 G	0.53	0.750	B	0.81	0.740	I	1.54	1.497*	-278	-67	227	214
181 S	0.79	0.828	I	0.72	0.780	b	1.35	1.272*	-259	-12	156	240
182 G	0.53	0.883	B	0.81	0.922	I	1.54	1.217*	-261	-7	2	184
183 P	0.59	1.018	B	0.62	1.120*	b	1.56	0.967	-142	22	66	128

residue	Pa	Pa avg.	Ba	Pb	Pb avg.	Bb	Pt	Pt avg.	Pb	Pa	Ptu	Pco
184 A	1.45	1.162*	H	0.97	1.285*	i	0.64	0.737	-75	72	20	45
185 T	0.61	1.087*	b	1.29	1.347*	h	1.13	0.720	0	120	-91	-46
186 I	1.00	1.240*	I	1.60	1.267*	H	0.54	0.597	64	182	-151	-93
187 F	1.12	1.315*	h	1.28	1.267*	h	0.64	0.597	121	211	-188	-106
188 L	1.34	1.352*	H	1.22	1.012*	h	0.57	0.642	157	158	-206	-109
189 A	1.45	1.370*	H	0.97	0.950	i	0.64	0.660	200	127	-185	-110
190 I	1.00	1.292*	I	1.60	1.012*	H	0.54	0.642	204	97	-211	-108
191 E	1.53	1.213*	B	0.26	0.855	B	0.82	0.667	201	28	-187	-119
192 A	1.45	1.047*	H	0.97	0.990	i	0.64	0.820	198	-53	-125	-75
193 L	1.34	0.995	H	1.22	0.950	h	0.57	1.045*	185	-102	-71	-20
194 A	1.45	1.013*	H	0.97	0.847	i	0.64	1.287*	163	-108	5	0
195 D	0.88	1.013*	I	0.80	1.017*	I	1.43	1.260*	126	-105	56	30
196 G	0.53	0.938	B	0.81	1.060*	i	1.54	1.062*	74	-82	82	54
197 G	0.53	1.073*	B	0.81	1.100*	i	1.54	0.837	49	-37	22	44
198 V	1.14	1.083*	h	1.65	1.100*	H	0.53	0.837	49	23	-65	35
199 A	1.45	1.025*	H	0.97	0.992	I	0.64	0.847	30	32	-75	35
200 A	1.45	1.038*	H	0.97	0.905	I	0.64	1.077*	-7	32	-50	30
201 G	0.53	1.020*	B	0.81	0.887	i	1.54	1.180*	-46	13	-103	44
202 L	1.34	1.173*	H	1.22	0.750	h	0.57	1.000*	-33	13	-236	25
203 P	0.59	1.173*	B	0.62	0.750	b	1.56	1.000*	83	-3	-124	-12
204 R	0.79	1.207*	i	0.90	0.837	B	1.05	0.770	117	-26	-64	-57
205 H	1.53	1.298*	H	0.26	0.917	B	0.82	0.650	168	-5	-127	-109
206 L	1.34	1.285*	H	1.22	1.032*	h	0.57	0.782	188	28	-141	-110
207 A	1.45	1.193*	H	0.97	1.032*	I	0.64	0.782	188	27	-100	-60
208 L	1.34	1.147*	H	1.22	1.032*	h	0.57	0.782	176	13	-66	-15
209 S	0.79	1.060*	I	0.72	0.907	b	1.35	0.977	149	-17	-44	35
210 L	1.34	1.118*	H	1.22	1.035*	h	0.57	0.882	140	-32	-56	10
211 A	1.45	1.118*	H	0.97	1.030*	I	0.64	0.987	150	-13	-30	-5
212 S	0.79	0.965	I	0.72	1.200*	b	1.35	0.960	129	23	-28	-5
213 Q	1.17	1.075*	h	1.23	1.325*	h	0.97	0.765	105	72	-41	-10
214 T	0.82	1.122*	I	1.20	1.220*	h	0.99	0.907	67	103	-52	12
215 V	1.14	1.122*	h	1.65	1.162*	H	0.53	0.820	59	113	-65	30
216 L	1.34	1.132*	H	1.22	0.992	h	0.57	0.847	72	68	-76	5
217 G	0.53	1.098*	B	0.81	0.987	i	1.54	0.952	89	28	-73	-21
218 A	1.45	1.142*	H	0.97	1.282*	I	0.64	0.730	128	47	-77	-28
219 A	1.45	1.078*	H	0.97	1.372*	I	0.64	0.702	160	72	-105	5
220 T	0.82	0.973	I	1.20	1.310*	h	0.99	0.880	184	68	-92	-3
221 M	1.30	0.925	h	1.67	1.195*	H	0.65	0.870	173	103	-93	-61
222 V	1.14	0.903	h	1.65	1.077*	H	0.53	0.955	134	78	-95	-90
223 S	0.79	0.920	I	0.72	0.867	b	1.35	1.207*	61	73	14	7
224 K	1.07	0.867	I	0.74	0.872	b	0.95	1.107*	8	17	80	37
225 T	0.82	0.797	I	1.20	0.865	h	0.99	1.105*	-31	-52	98	82
226 G	0.53	0.850	B	0.81	0.720	I	1.54	1.247*	-86	-102	132	96
227 K	1.07	0.985	I	0.74	0.720	b	0.95	1.247*	-117	-93	30	77
228 H	1.24	0.985	h	0.71	0.947	b	0.94	1.142*	-168	-30	-138	51
229 P	0.59	0.942	B	0.62	1.075*	b	1.56	1.050*	-137	17	1	68
230 G	0.53	1.007*	B	0.81	1.105*	I	1.54	0.897	-133	43	67	39
231 V	1.14	1.108*	h	1.65	1.102*	H	0.53	0.870	-60	83	-35	-15
232 L	1.34	1.055*	H	1.22	0.890	h	0.57	1.095*	-78	133	-51	-18
233 K	1.07	0.963	I	0.74	0.907	b	0.95	1.065*	-92	142	-5	-33
234 D	0.98	0.883	I	0.80	1.112*	I	1.43	1.095*	-107	136	61	0
235 D	0.98	0.808	I	0.80	1.092*	I	1.43	1.075*	-125	156	84	25
236 V	1.14	0.733	h	1.65	1.047*	H	0.53	1.107*	-156	163	70	60
237 T	0.82	0.680	I	1.20	0.837	h	0.99	1.360*	-191	153	53	122
238 S	0.79	0.680	I	0.72	0.740	b	1.35	1.497*	-239	68	16	185
239 P	0.59	0.715	B	0.62	0.860	b	1.56	1.407*	-197	17	186	183
240 G	0.53	0.858	B	0.81	1.005*	I	1.54	1.265*	-211	-32	245	189
241 G	0.53	0.858	B	0.81	1.202*	I	1.54	1.015*	-181	-47	135	148
242 T	0.82	0.960	I	1.20	1.242*	h	0.99	0.790	-136	18	33	137
243 T	0.82	1.030*	I	1.20	1.145*	h	0.99	0.927	-96	113	-32	102
244 I	1.00	1.148*	I	1.60	1.257*	H	0.54	0.812	-29	162	-66	47
245 A	1.45	1.205*	H	0.97	1.035*	I	0.64	0.912	30	157	-100	20
246 G	0.53	1.218*	B	0.81	0.887	I	1.54	0.957	44	93	-93	9
247 V	1.14	1.308*	h	1.65	0.960	H	0.53	0.715	109	63	-150	16
248 H	1.24	1.207*	h	0.71	0.612	b	0.94	0.787	120	30	-183	-4
249 E	1.53	1.132*	H	0.26	0.620	B	0.82	0.790	133	-35	-147	-44
250 L	1.34	1.063*	H	1.22	0.757	h	0.57	0.970	140	-77	-51	-5
251 E	1.53	0.972	H	0.26	0.632	B	0.82	1.165*	141	-115	53	-7
252 K	1.07	0.958	I	0.74	0.887	b	0.95	1.120*	147	-168	105	22
253 G	0.53	0.917	B	0.81	0.937	I	1.54	1.145*	130	-172	92	74
254 S	0.79	1.052*	i	0.72	0.967	b	1.35	0.920	159	-142	31	60
255 P	1.12	1.120*	h	1.28	1.087*	H	0.64	0.830	179	-64	-68	-36
256 R	0.79	1.055*	I	0.90	1.072*	I	1.05	0.812	171	-41	-86	-67
257 A	1.45	1.165*	H	0.97	1.265*	I	0.64	0.712	148	-28	-83	-50
258 T	0.82	1.113*	I	1.20	1.185*	h	0.99	0.930	119	-32	-62	-8
259 L	1.34	1.167*	H	1.22	1.127*	h	0.57	0.842	107	-47	-76	5
260 M	1.20	1.185*	h	1.67	1.238*	H	0.65	0.832	118	-12	-83	-41
261 N	0.73	1.227*	b	0.65	1.230*	b	1.51	0.802	152	-46	-128	-64
262 A	1.45	1.347*	H	0.97	1.310*	I	0.64	0.585	210	-3	-180	-98
263 V	1.14	1.283*	h	1.65	1.310*	H	0.53	0.585	264	23	-260	-135
264 V	1.14	1.225*	h	1.65	1.140*	H	0.53	0.612	304	23	-275	-155
265 A	1.45	1.167*	H	0.97	0.912	I	0.64	0.717	310	-23	-260	-165
266 A	1.45	1.057*	H	0.97	0.895	I	0.64	0.820	285	-73	-210	-120
267 A	1.45	1.070*	H	0.97	0.832	I	0.64	0.997	225	-68	-150	-83
268 K	1.07	1.052*	I	0.74	0.815	b	0.95	1.100*	183	-98	-95	-28
269 H	0.79	1.005*	I	0.90	0.695	I	1.05	1.067*	116	-96	-54	-32
270 S	0.79	1.068*	I	0.72	0.775	b	1.35	0.947	101	-67	-19	5
271 H	0.79	1.068*	I	0.90	0.775	I	1.05	0.947	64	-46	-14	-37
272 E	1.53	1.124*	H	0.26	0.857	B	0.82	0.927	38	-45	-2	-14
273 L	1.34	1.022*	H	1.22	0.972	h	0.57	1.060*	15	-57	29	25
274 S	0.79	0.917	I	0.72	0.890	b	1.35	1.223*	-26	-67	71	70
275 Q	1.17	0.980	h	1.23	0.975	h	0.97	1.160*	-54	-58	69	40
276 S	0.79	0.790	I	0.72	0.720	b	1.35	1.350*	-66	-32	81	100

ii) Numerical data used in the prediction of secondary structure of P5CR from soybean (Delauncy and Verma, 1990).

residue		Pa	Pa Avg	Ha	Pb	Pb avg	Hb	Pt	Pt avg	Pa	Ps	Pts	Pco
1	M	1.20	1.073*	h	1.67	1.202*	H	0.65	0.662	53	53	-38	-36
2	E	1.53	0.972	H	0.26	0.940	B	0.82	0.890	28	55	-27	-64
3	I	1.00	0.958	I	1.60	1.275*	H	0.54	0.820	-26	62	-101	-53
4	P	1.12	1.047*	h	1.28	1.030*	h	0.64	1.075*	-81	36	-188	-46
5	P	0.59	0.992	B	0.62	0.952	b	1.56	1.075*	-49	-18	-89	38
6	I	1.00	0.995	I	1.60	0.862	H	0.54	0.890	-61	-53	-101	87
7	P	0.59	0.965	B	0.62	0.642	b	1.56	1.092*	38	-73	-39	73
8	A	1.45	1.090*	H	0.97	0.810	I	0.64	0.985	35	-98	65	80
9	E	1.53	0.937	H	0.26	0.867	B	0.82	1.072*	33	-75	13	16
10	S	0.79	0.868	i	0.72	1.107*	b	1.35	1.010*	4	-27	39	30
11	Y	0.61	0.903	b	1.29	1.130*	h	1.13	1.057*	-29	35	69	19
12	T	0.82	0.890	i	1.20	1.127*	h	0.99	0.935	7	73	68	42
13	L	1.34	0.995	H	1.22	1.227*	h	0.57	0.822	30	98	19	40
14	G	0.53	0.860	B	0.81	1.125*	i	1.54	1.065*	29	73	12	-1
15	F	1.12	0.950	h	1.28	1.165*	h	0.64	0.840	59	71	-13	-31
16	I	1.00	0.963	I	1.60	1.047*	H	0.54	1.065*	56	12	-46	22
17	G	0.53	1.038*	B	0.81	0.832	i	1.54	1.167*	24	-52	-28	54
18	A	1.45	1.205*	H	0.97	1.047*	I	0.64	0.945	40	-93	-20	60
19	G	0.53	1.095*	B	0.81	1.047*	i	1.54	0.945	29	-142	-43	76
20	K	1.07	1.173*	I	0.74	0.910	b	0.95	0.765	73	-118	-60	52
21	M	1.20	1.237*	h	1.67	0.905	H	0.65	0.865	123	-62	-103	24
22	A	1.45	1.168*	H	0.97	0.887	I	0.64	0.837	150	-48	-120	-10
23	E	1.53	1.015*	H	0.26	0.887	B	0.82	0.837	163	25	-152	-99
24	S	0.79	1.002*	I	0.72	1.047*	b	1.35	0.895	159	28	-149	-105
25	I	1.00	1.060*	I	1.60	1.070*	H	0.54	0.942	154	-33	-141	-93
26	A	1.45	1.025*	H	0.97	0.912	I	0.64	0.967	138	-28	-100	-80
27	R	0.79	0.915	i	0.90	1.082*	i	1.05	0.940	94	-46	-89	-67
28	G	0.53	0.872	B	0.81	1.082*	i	1.54	0.940	39	-37	-8	-6
29	A	1.45	0.973	H	0.97	1.060*	I	0.64	0.892	5	22	20	30
30	V	1.14	0.955	h	1.65	1.020*	H	0.53	1.117*	-81	93	30	60
31	R	0.79	0.863	I	0.90	1.020*	I	1.05	1.117*	-184	124	101	58
32	S	0.79	0.830	I	0.72	1.100*	b	1.35	0.997	-259	138	101	70
33	G	0.53	0.830	B	0.81	1.075*	i	1.54	1.050*	-323	98	122	44
34	V	1.14	0.873	h	1.65	1.027*	H	0.53	1.055*	-321	83	60	115
35	L	1.34	0.850	H	1.22	0.795	h	0.57	1.260*	-338	63	-116	155
36	P	0.59	0.758	B	0.62	0.715	b	1.56	1.380*	-282	37	-44	158
37	P	0.59	0.797	B	0.62	0.960	b	1.56	1.125*	-194	57	131	133
38	S	0.79	0.940	i	0.72	1.030*	b	1.35	0.997	-119	43	116	100
39	R	0.79	0.998	i	0.90	1.150*	I	1.05	0.907	-49	29	21	3
40	I	1.00	1.073*	I	1.60	1.167*	H	0.54	0.805	-4	117	-8	-23
41	R	0.79	1.093*	I	0.90	1.180*	I	1.05	0.802	26	174	-19	-32
42	T	0.82	1.083*	I	1.20	1.132*	h	0.99	0.775	39	173	-7	7
43	A	1.45	1.170*	H	0.97	1.152*	I	0.64	0.687	35	142	-20	5
44	V	1.14	1.170*	h	1.65	1.072*	H	0.53	0.905	-1	108	-60	10
45	H	1.24	1.112*	h	0.71	0.965	b	0.94	0.915	-28	70	-48	1
46	F	1.12	1.037*	h	1.28	1.030*	h	0.64	0.840	-46	16	-38	-16
47	N	0.73	0.938	b	0.65	0.935	b	1.51	0.942	-16	-1	-43	6
48	L	1.34	1.058*	H	1.22	0.997	h	0.57	0.827	37	-22	-31	20
49	A	1.45	1.022*	H	0.97	0.895	I	0.64	1.070*	60	-73	20	50
50	R	0.79	1.035*	I	0.90	0.895	i	1.05	1.070*	44	-91	51	53
51	R	0.79	1.035*	I	0.90	0.990	i	1.05	0.967	26	-91	46	58
52	G	0.53	1.090*	B	0.81	0.830	I	1.54	0.910	-6	-97	57	79
53	A	1.45	1.090*	H	0.97	0.807	I	0.64	0.862	30	-23	35	75
54	F	1.12	1.038*	h	1.28	0.885	h	0.64	0.862	1	1	-8	54
55	E	1.53	0.988	H	0.26	0.767	B	0.82	1.087*	-42	10	3	41
56	S	0.79	0.923	I	0.72	1.115*	b	1.35	1.015*	-86	3	56	35
57	F	1.12	1.015*	h	1.28	1.235*	h	0.64	0.925	-101	41	52	-16
58	G	0.53	0.927	B	0.81	1.327*	I	1.54	0.897	-118	38	60	4
59	V	1.14	0.970	h	1.65	1.430*	H	0.53	0.655	-83	78	0	20
60	T	0.82	0.902	I	1.20	1.172*	h	0.99	0.912	-103	58	18	62
61	V	1.14	0.928	h	1.65	1.052*	H	0.53	1.002*	-146	78	-50	80
62	L	1.34	0.902	H	1.22	0.802	h	0.57	1.247*	-203	73	-121	80
63	P	0.59	0.868	B	0.62	0.697	b	1.56	1.462*	-162	-13	71	98
64	S	0.79	0.960	i	0.72	0.742	b	1.35	1.430*	-161	-107	146	130
65	N	0.73	0.960	b	0.65	0.975	b	1.51	1.225*	-121	-91	52	71
66	D	0.98	1.093*	i	0.80	1.225*	I	1.43	0.980	-80	-59	36	35
67	D	0.98	1.062*	i	0.80	1.250*	I	1.43	0.885	-35	26	-9	25
68	V	1.14	1.062*	h	1.65	1.115*	H	0.53	0.732	4	93	-65	5
69	V	1.14	1.062*	h	1.65	0.882	H	0.53	0.937	54	113	-90	-30
70	H	0.79	1.062*	I	0.90	0.670	I	1.05	1.162*	56	89	-74	-57
71	E	1.53	1.120*	H	0.26	0.857	B	0.82	1.032*	63	55	-72	-104
72	S	0.79	1.088*	i	0.72	1.205*	b	1.35	0.960	84	23	-74	-95
73	D	0.98	1.088*	i	0.80	1.437*	i	1.43	0.755	93	-9	-64	-110
74	V	1.14	1.115*	h	1.65	1.542*	H	0.53	0.540	97	28	-90	-100
75	V	1.14	1.103*	h	1.65	1.310*	H	0.53	0.745	75	118	-130	-115
76	V	1.14	1.012*	h	1.65	1.310*	H	0.53	0.745	54	158	-125	-85
77	L	1.34	1.017*	H	1.22	1.082*	h	0.57	0.850	62	123	-96	-70
78	S	0.79	1.017*	i	0.72	0.932	b	1.35	1.097*	56	63	-34	-35
79	V	1.14	1.075*	h	1.65	1.060*	H	0.53	1.002*	22	-12	-90	-43
80	K	1.07	1.063*	I	0.74	0.952	b	0.95	1.012*	-12	-33	-180	-43
81	P	0.59	1.048*	B	0.62	1.180*	b	1.56	0.907	31	-28	-49	-37
82	Q	1.17	1.140*	h	1.23	1.210*	h	0.97	0.755	40	7	-26	-60
83	L	1.34	1.135*	H	1.22	1.102*	h	0.57	0.870	62	43	-146	-110
84	V	1.14	1.043*	h	1.65	1.210*	H	0.53	0.860	59	98	-145	-143
85	K	1.07	1.032*	I	0.74	1.210*	b	0.95	0.860	66	147	-70	-93
86	D	0.98	1.077*	i	0.80	1.205*	I	1.43	0.960	80	171	1	-40
87	V	1.14	1.050*	h	1.65	1.190*	H	0.53	0.840	64	188	-15	-10
88	V	1.14	0.958	h	1.65	1.082*	H	0.53	0.850	39	158	-20	-20
89	S	0.79	0.992	i	0.72	0.970	b	1.35	0.965	14	123	-9	-8

residue	Pa	Pa avg.	Ha	Pb	Pb avg.	Hb	Pt	Pt avg.	Pb	Pb	Ptu	Pco
90 K	1.07	1.083*	I	0.74	0.945	b	0.95	1.017*	13	62	-7	-30
91 L	1.34	1.042*	H	1.22	1.065*	h	0.57	0.922	17	8	-121	-35
92 T	0.82	0.997	I	1.20	1.065*	h	0.99	0.922	27	-52	-217	-58
93 P	0.59	1.007*	H	0.62	1.065*	b	1.56	0.922	114	-48	-109	-57
94 L	1.34	1.147*	H	1.22	1.005*	h	0.57	0.770	157	-12	-68	-70
95 L	1.34	1.147*	H	1.22	0.967	h	0.57	0.862	187	8	-101	-85
96 T	0.82	1.147*	I	1.20	0.847	h	0.99	0.957	182	-12	-67	-71
97 K	1.07	1.200*	I	0.74	0.852	b	0.95	0.852	166	-23	-50	-48
98 R	1.24	1.153*	h	0.71	0.972	b	0.94	0.757	135	8	-53	-62
99 K	1.07	1.137*	H	0.74	1.207*	b	0.95	0.655	86	17	-65	-48
100 L	1.34	1.200*	H	1.22	1.202*	h	0.57	0.755	82	93	-121	-60
101 L	1.34	1.218*	H	1.22	1.310*	h	0.57	0.745	112	148	-130	-65
102 V	1.14	1.083*	h	1.65	1.247*	H	0.53	0.762	137	158	-135	-90
103 S	0.79	1.030*	I	0.72	1.077*	h	1.35	0.790	164	138	-154	-70
104 V	1.14	1.077*	h	1.65	1.100*	H	0.53	0.837	194	73	-180	-110
105 A	1.45	1.110*	H	0.97	0.987	I	0.64	0.952	195	-3	-145	-100
106 A	1.45	1.047*	H	0.97	0.930	I	0.64	1.030*	180	-68	-107	-90
107 G	0.53	0.968	B	0.81	0.992	i	1.54	1.012*	124	-137	-53	-36
108 T	0.82	1.103*	I	1.20	0.975	h	0.99	0.865	97	-97	-42	-31
109 E	1.07	1.162*	I	0.74	0.875	b	0.95	0.975	78	-88	-70	-3
110 L	1.34	1.338*	H	1.22	0.995	h	0.57	0.880	42	-42	-56	7
111 E	1.07	1.205*	I	0.74	0.997	b	0.95	0.980	91	7	-20	17
112 D	0.98	1.268*	I	0.80	0.877	i	1.43	0.947	63	36	-29	10
113 L	1.34	1.153*	H	1.22	0.975	h	0.57	0.840	142	88	-81	5
114 Q	1.17	1.092*	h	1.23	0.912	h	0.97	0.857	140	72	-46	-5
115 E	1.53	1.060*	H	0.26	0.807	B	0.82	1.000*	126	20	23	-14
116 W	1.14	0.937	h	1.19	0.905	h	1.00	1.172*	105	-25	56	22
117 A	1.45	0.933	H	0.97	0.807	i	0.64	1.280*	88	-108	40	85
118 G	0.53	0.858	B	0.81	0.790	i	1.54	1.382*	52	-152	142	129
119 N	0.73	0.902	b	0.65	0.907	h	1.51	1.157*	62	-136	157	161
120 D	0.98	0.970	I	0.80	1.145*	I	1.43	0.515	20	-94	141	135
121 R	0.79	1.007*	I	0.90	1.170*	I	1.05	0.820	-9	34	115	88
122 F	1.12	0.973	h	1.28	1.357*	h	0.64	0.690	-39	151	117	44
123 I	1.00	0.908	I	1.60	1.455*	H	0.54	0.592	-44	202	84	37
124 R	0.79	0.878	I	0.90	1.210*	I	1.05	0.847	-109	244	101	58
125 V	1.14	0.845	h	1.65	1.147*	H	0.53	1.062*	-166	203	55	70
126 M	1.20	0.897	h	1.67	1.035*	H	0.65	1.177*	-247	83	-9	124
127 P	0.59	0.938	B	0.62	0.772	b	1.56	1.405*	-232	-28	124	153
128 N	0.73	1.030*	b	0.65	0.860	b	1.51	1.175*	-241	-106	117	176
129 T	0.82	0.997	I	1.20	0.940	h	0.99	0.957	-191	-112	-82	127
130 P	0.59	1.055*	B	0.62	1.052*	b	1.56	0.862	-72	-78	-9	78
131 A	1.45	1.198*	H	0.97	1.100*	I	0.64	0.837	19	-3	15	40
132 A	1.48	1.198*	H	0.97	1.165*	I	0.64	0.920	105	52	-90	-15
133 V	1.14	1.088*	h	1.65	1.155*	H	0.53	0.920	148	33	-135	-40
134 G	0.53	1.088*	B	0.81	0.995	i	1.54	0.947	119	38	-88	-26
135 Q	1.17	1.200*	h	1.23	0.972	h	0.97	0.900	130	62	-125	-69
136 A	1.45	1.137*	H	0.97	1.077*	I	0.64	0.790	100	52	-150	-55
137 A	1.45	1.118*	H	0.97	1.252*	I	0.64	0.792	80	77	-110	-30
138 S	0.79	0.965	I	0.72	1.190*	b	1.35	0.970	61	153	-99	-20
139 V	1.14	0.922	h	1.65	1.315*	H	0.53	0.775	49	218	-100	-55
140 M	1.20	0.863	h	1.67	1.105*	H	0.65	1.027*	11	203	-23	-1
141 S	0.79	0.905	I	0.72	0.890	b	1.35	1.350*	-29	113	56	105
142 L	1.34	0.910	H	1.22	0.890	h	0.57	1.350*	-58	3	69	135
143 G	0.53	0.942	B	0.81	0.827	i	1.54	1.267*	-106	-82	72	139
144 G	0.53	1.108*	B	0.81	0.925	i	1.54	1.130*	-113	-127	37	109
145 S	0.79	1.183*	I	0.72	0.787	b	1.35	0.950	-89	-72	-31	175
146 A	1.45	1.140*	H	0.97	0.672	I	0.64	0.817	8	-3	-110	95
147 T	0.82	1.020*	I	1.20	0.630	h	0.99	1.015*	74	-2	-97	57
148 E	1.53	1.050*	H	0.26	0.532	B	0.82	1.152*	98	-35	-42	26
149 E	1.53	0.962	H	0.26	0.630	B	0.82	1.320*	116	-75	48	11
150 D	0.98	0.948	I	0.80	0.965	I	1.43	1.255*	116	-104	116	49
151 G	0.53	0.980	B	0.81	1.165*	i	1.54	1.032*	95	-112	112	64
152 N	0.73	1.115*	b	0.65	1.205*	b	1.51	0.807	135	-61	52	61
153 I	1.00	1.180*	I	1.60	1.350*	H	0.54	0.572	169	43	-66	2
154 I	1.00	1.182*	I	1.60	1.255*	H	0.54	0.680	171	107	-131	-53
155 A	1.45	1.067*	H	0.97	1.175*	I	0.64	0.705	158	132	-130	-90
156 Q	1.17	0.992	h	1.23	1.135*	h	0.97	0.930	120	142	-66	-40
157 L	1.34	0.885	H	1.22	1.007*	h	0.57	1.020*	47	128	-31	5
158 F	1.12	0.848	h	1.28	1.102*	h	0.64	1.017*	-14	111	22	-1
159 C	0.53	0.820	B	0.81	0.985	i	1.54	1.242*	-58	-2	102	89
160 S	0.79	0.922	I	0.72	0.967	b	1.35	1.055*	-29	-17	161	160
161 I	1.00	0.968	I	1.60	1.187*	H	0.54	0.882	-29	-33	104	132
162 G	0.53	1.043*	B	0.81	1.085*	i	1.54	1.007*	-1	-72	87	126
163 K	1.07	1.210*	I	0.74	1.067*	b	0.95	0.860	48	-63	55	112
164 I	1.00	1.287*	I	1.60	1.125*	H	0.54	0.782	66	-73	-1	87
165 W	1.14	1.298*	h	1.19	0.750	h	1.00	0.852	77	-70	-114	24
166 K	1.07	1.210*	I	0.74	0.557	b	0.95	0.807	128	-58	-170	-18
167 A	1.45	1.218*	H	0.97	0.557	I	0.64	0.807	180	-98	-90	-25
168 E	1.53	1.140*	H	0.26	0.637	B	0.82	0.930	188	-125	-77	-39
169 E	1.53	1.127*	H	0.26	0.892	B	0.82	0.885	126	-115	-22	-67
170 K	1.07	1.038*	I	0.74	1.027*	b	0.95	1.027*	139	-138	46	-23
171 V	0.61	0.997	b	1.29	1.085*	h	1.13	0.960	156	-95	89	24
172 P	1.12	0.983	h	1.28	1.162*	h	0.64	0.812	172	-29	32	-46
173 D	0.98	1.020*	I	0.80	1.142*	I	1.43	0.900	163	11	-39	-75
174 A	1.45	0.988	H	0.97	1.145*	I	0.64	0.937	125	52	-87	-95
175 I	1.00	0.835	I	1.60	1.207*	H	0.54	0.910	48	82	-41	-48
176 T	0.82	0.800	I	1.20	0.987	h	0.99	1.112*	-48	68	-2	-3
177 G	0.53	0.752	B	0.81	0.890	i	1.54	1.350*	-171	33	92	99
178 L	1.34	0.762	H	1.22	0.867	h	0.57	1.202*	-218	28	149	155
179 S	0.79	0.780	I	0.72	0.768	b	1.35	1.448*	-274	-2	176	225
180 G	0.53	0.750	B	0.81	0.740	i	1.54	1.497*	-328	-42	247	234
181 S	0.79	0.852	I	0.72	0.780	b	1.35	1.272*	-314	-22	186	270
182 G	0.53	0.822	B	0.81	0.922	i	1.54	1.217*	-316	-2	17	194
183 P	0.59	0.957	B	0.62	1.132*	b	1.56	0.965	-192	42	81	118

residue	Pa	Pa avg.	Ha	Pb	Pb avg.	Hb	Pt	Pt avg.	Ph	Ps	Ptu	Pco
184 A	1.45	1.100*	H	0.97	1.300*	i	0.64	0.857	-120	92	30	45
185 Y	0.61	1.025*	b	1.29	1.362*	h	1.13	0.840	-45	140	-71	-36
186 V	1.14	1.178*	h	1.65	1.282*	H	0.53	0.717	17	198	-125	-75
187 Y	0.61	1.230*	b	1.29	1.270*	h	1.13	0.720	70	225	-141	-71
188 L	1.34	1.352*	H	1.22	1.012*	h	0.57	0.642	117	183	-186	-95
189 A	1.45	1.370*	H	0.97	0.950	i	0.64	0.660	170	167	-185	-130
190 I	1.00	1.292*	I	1.60	1.012*	H	0.54	0.642	189	112	-221	-143
191 E	1.53	1.213*	H	0.26	0.855	B	0.82	0.667	201	90	-222	-169
192 A	1.45	1.047*	H	0.97	0.990	I	0.64	0.820	198	37	-175	-130
193 L	1.34	0.995	H	1.22	0.950	h	0.57	1.045*	180	-22	-111	-60
194 A	1.45	1.013*	H	0.97	0.847	I	0.64	1.287*	158	-63	-15	-10
195 D	0.98	1.013*	f	0.80	1.017*	i	1.43	1.260*	121	-89	56	30
196 G	0.53	0.938	B	0.81	1.060*	i	1.54	1.062*	74	-82	52	54
197 G	0.53	1.073*	B	0.81	1.100*	i	1.54	0.837	49	-27	22	44
198 V	1.14	1.083*	h	1.65	1.100*	H	0.53	0.837	44	43	-60	25
199 A	1.45	1.025*	H	0.97	0.992	I	0.64	0.847	15	62	-65	25
200 A	1.45	0.947	H	0.97	0.905	I	0.64	1.077*	-37	82	-45	0
201 G	0.53	0.928	B	0.81	0.887	i	1.54	1.180*	-101	83	-103	-16
202 L	1.34	0.972	H	1.22	0.885	h	0.57	1.152*	-103	58	-216	-15
203 P	0.59	0.972	B	0.62	0.885	b	1.56	1.152*	-37	22	-49	3
204 R	0.79	1.005*	i	0.90	0.910	i	1.05	1.100*	-18	-6	36	-22
205 D	0.98	1.097*	i	0.80	0.990	i	1.43	0.980	15	6	11	-25
206 L	1.34	1.175*	H	1.22	0.970	h	0.57	0.960	25	58	-36	-20
207 S	0.79	1.083*	i	0.72	0.970	b	1.35	0.960	21	73	-19	35
208 L	1.34	1.147*	H	1.22	1.032*	h	0.57	0.782	23	78	-11	45
209 S	0.79	1.060*	i	0.72	0.907	b	1.35	0.977	11	38	1	75
210 L	1.34	1.118*	H	1.22	1.035*	h	0.57	0.882	25	13	-21	40
211 A	1.45	1.118*	H	0.97	1.030*	I	0.64	0.987	55	22	-10	15
212 S	0.79	0.965	i	0.72	1.200*	b	1.35	0.960	64	48	-26	0
213 Q	1.17	1.075*	h	1.23	1.325*	h	0.97	0.765	60	102	-61	-30
214 T	0.82	1.122*	i	1.20	1.220*	h	0.99	0.907	42	118	-92	-28
215 V	1.14	1.117*	h	1.65	1.162*	H	0.53	0.820	44	103	-105	-20
216 L	1.34	1.127*	H	1.22	0.992	h	0.57	0.847	42	58	-66	15
217 G	0.53	1.093*	B	0.81	0.867	i	1.54	1.042*	39	18	-33	19
218 A	1.45	1.137*	H	0.97	1.082*	I	0.64	0.820	68	42	-15	50
219 A	1.45	1.090*	H	0.97	1.252*	I	0.64	0.792	95	67	-20	80
220 S	0.79	0.985	i	0.72	1.190*	b	1.35	0.970	116	68	-24	75
221 M	1.20	0.942	h	1.67	1.317*	H	0.65	0.875	113	113	-48	-16
222 V	1.14	0.920	h	1.65	1.200*	H	0.53	0.960	94	98	-75	-60
223 S	0.79	0.937	i	0.72	0.990	b	1.35	1.212*	41	93	19	25
224 Q	1.17	0.903	h	1.23	0.995	h	0.97	1.112*	-15	47	84	50
225 T	0.82	0.797	i	1.20	0.865	h	0.99	1.105*	-61	-12	118	102
226 G	0.53	0.855	B	0.81	0.720	i	1.54	1.247*	-121	-37	167	131
227 K	1.07	0.990	I	0.74	0.720	b	0.95	1.247*	-137	-53	110	157
228 H	1.24	0.990	h	0.71	0.842	b	0.94	1.252*	-168	-20	-23	166
229 P	0.59	0.947	B	0.62	0.970	b	1.56	1.160*	-127	7	121	183
230 G	0.53	1.012*	B	0.81	1.000*	i	1.54	1.007*	-123	23	172	129
231 Q	1.17	1.113*	h	1.23	0.997	h	0.97	0.980	-70	57	74	60
232 L	1.34	1.055*	H	1.22	0.890	h	0.57	1.095*	-98	108	29	52
233 K	1.07	0.963	I	0.74	0.997	b	0.95	1.085*	-117	132	55	22
234 D	0.98	0.883	i	0.80	1.112*	i	1.43	1.095*	-127	156	121	60
235 D	0.98	0.808	i	0.80	1.092*	i	1.43	1.075*	-135	206	144	85
236 V	1.14	0.733	h	1.65	1.047*	H	0.53	1.107*	-161	233	130	120
237 T	0.82	0.680	i	1.20	0.837	h	0.99	1.360*	-196	208	103	172
238 S	0.79	0.680	i	0.72	0.740	b	1.35	1.497*	-249	108	66	245
239 P	0.59	0.715	B	0.62	0.860	b	1.56	1.407*	-207	47	226	223
240 G	0.53	0.753	B	0.81	1.005*	i	1.54	1.265*	-226	-22	275	219
241 G	0.53	0.753	B	0.81	1.202*	i	1.54	1.015*	-206	-37	175	174
242 T	0.82	0.832	i	1.20	1.300*	h	0.99	0.877	-171	53	78	162
243 T	0.82	0.902	i	1.20	1.202*	h	0.99	1.015*	-141	138	36	142
244 I	1.00	1.020*	I	1.60	1.302*	H	0.54	0.902	-114	207	9	107
245 T	0.82	1.077*	i	1.20	1.080*	h	0.99	1.002*	-91	233	-17	87
246 G	0.53	1.195*	B	0.81	0.845	i	1.54	0.960	-96	198	-8	79
247 I	1.00	1.228*	I	1.60	0.947	H	0.54	0.717	-29	167	-36	102
248 H	1.24	1.150*	h	0.71	0.612	b	0.94	0.787	-20	115	-83	91
249 E	1.53	1.032*	H	0.26	0.597	B	0.82	0.930	-17	65	-37	71
250 L	1.34	0.963	H	1.22	0.735	h	0.57	1.110*	-10	38	24	95
251 E	1.53	0.872	H	0.26	0.632	B	0.82	1.352*	-9	-35	128	96
252 N	0.73	0.705	b	0.65	0.887	b	1.51	1.307*	-22	-106	212	126
253 G	0.53	0.720	B	0.81	0.950	i	1.54	1.192*	-30	-147	212	179
254 G	0.53	0.855	B	0.81	0.950	i	1.54	1.192*	-13	-152	177	154
255 F	1.12	0.967	h	1.28	1.047*	h	0.64	1.055*	29	-54	102	94
256 R	0.79	0.902	i	0.90	1.032*	i	1.05	1.037*	21	-21	49	73
257 G	0.53	1.012*	B	0.81	1.225*	i	1.54	0.937	-18	-2	37	69
258 T	0.82	1.113*	i	1.20	1.185*	h	0.99	0.930	-11	58	23	72
259 L	1.34	1.167*	H	1.22	1.127*	h	0.57	0.842	2	28	-41	70
260 M	1.20	1.185*	h	1.67	1.235*	H	0.65	0.832	33	38	-63	4
261 N	0.73	1.227*	b	0.65	1.230*	b	1.51	0.802	87	9	-118	-44
262 A	1.45	1.347*	H	0.97	1.310*	I	0.64	0.585	165	27	-180	-85
263 V	1.14	1.283*	h	1.65	1.310*	H	0.53	0.585	234	53	-260	-135
264 V	1.14	1.225*	h	1.65	1.140*	H	0.53	0.612	284	43	-275	-155
265 A	1.45	1.167*	H	0.97	0.912	I	0.64	0.717	300	-33	-260	-165
266 A	1.45	1.057*	H	0.97	0.895	I	0.64	0.820	285	-73	-210	-120
267 A	1.45	1.070*	H	0.97	0.832	I	0.64	0.997	225	-68	-160	-93
268 K	1.07	1.052*	I	0.74	0.815	b	0.95	1.100*	183	-98	-115	-48
269 R	0.79	1.005*	i	0.90	0.695	i	1.05	1.067*	121	-106	-74	-52
270 S	0.79	1.048*	i	0.72	0.775	b	1.35	0.947	111	-87	-24	0
271 R	0.79	1.112*	i	0.90	0.775	i	1.05	0.947	74	-56	-14	-37
272 E	1.53	1.220*	H	0.26	0.733	B	0.82	0.913	48	-40	-27	-39
273 L	1.34	1.065*	H	1.22	0.970	h	0.57	0.960	20	-42	4	0
274 S	0.79	0.790	i	0.72	0.720	b	1.35	1.350*	-16	-57	46	50

iii) Numerical data used in the prediction of the secondary structure of P5CR from pea (Williamson and Slocum, 1992).

residue		Pa	Pa Avg	Ha	Pb	Pb avg	Hb	Pt	Pt avg	Pa	Pb	Ptu	Pcc
1	M	1.20	1.110*	h	1.67	1.187*	H	0.65	0.645	108	43	-103	-81
2	E	1.53	1.133*	H	0.26	0.925	H	0.82	0.872	91	65	-122	-144
3	I	1.00	1.010*	I	1.60	1.260*	H	0.54	0.802	44	82	-206	-143
4	L	1.34	1.007*	H	1.22	1.165*	h	0.57	0.810	-5	73	-281	-100
5	P	0.59	0.915	B	0.62	1.040*	b	1.56	1.005*	19	57	-79	-67
6	I	1.00	0.918	I	1.60	1.085*	H	0.54	0.972	12	37	19	12
7	L	1.34	0.888	H	1.22	0.865	h	0.57	1.175*	2	28	-26	30
8	S	0.79	0.888	i	0.72	0.882	b	1.35	1.315*	-29	23	56	65
9	D	0.98	0.845	i	0.80	1.002*	i	1.43	1.225*	-62	21	111	65
10	S	0.79	0.869	i	0.72	1.107*	b	1.35	1.010*	-89	58	99	90
11	Y	0.61	0.903	b	1.29	1.130*	h	1.13	1.057*	-122	90	99	59
12	T	0.82	0.890	i	1.20	1.127*	h	0.99	0.935	-86	118	78	72
13	L	1.34	0.995	H	1.22	1.227*	h	0.57	0.822	-53	128	14	45
14	G	0.53	0.860	B	0.81	1.125*	i	1.54	1.065*	-41	113	12	-1
15	P	1.12	0.950	h	1.28	1.165*	h	0.54	0.840	-1	101	-13	-31
16	I	1.00	0.963	I	1.60	1.047*	H	0.54	1.065*	21	32	-46	22
17	G	0.53	1.038*	B	0.81	0.832	I	1.54	1.167*	9	-42	-28	54
18	A	1.45	1.285*	H	0.97	1.047*	I	0.64	0.945	40	-93	-20	66
19	G	0.53	1.095*	B	0.81	1.047*	I	1.54	0.945	49	-147	-53	66
20	K	1.07	1.173*	I	0.74	0.910	b	0.95	0.765	113	-128	-80	32
21	M	1.20	1.237*	h	1.67	0.905	H	0.65	0.865	173	-77	-133	-6
22	A	1.45	1.215*	H	0.97	0.887	I	0.64	0.837	205	-68	-160	-50
23	E	1.53	1.062*	H	0.26	0.887	B	0.82	0.837	218	5	-187	-134
24	S	0.79	1.048*	I	0.72	1.007*	b	1.35	0.870	209	18	-179	-145
25	I	1.00	1.048*	I	1.60	1.030*	H	0.54	0.917	189	-53	-161	-123
26	A	1.45	1.013*	H	0.97	0.872	I	0.64	0.942	148	-68	-80	-68
27	K	1.07	0.903	I	0.74	0.810	b	0.95	1.120*	101	-109	-35	-28
28	G	0.53	0.813	B	0.81	0.850	i	1.54	1.145*	29	-107	22	39
29	A	1.45	0.915	H	0.97	0.827	I	0.64	1.097*	-5	-73	35	65
30	S	0.79	0.897	i	0.72	0.787	b	1.35	1.322*	-89	-2	36	85
31	R	0.79	0.863	i	0.90	1.020*	i	1.05	1.117*	-159	49	66	33
32	S	0.79	0.863	i	0.72	1.100*	b	1.35	0.997	-199	78	61	45
33	G	0.53	0.883	B	0.81	1.075*	I	1.54	1.050*	-238	53	107	49
34	V	1.14	0.907	h	1.65	1.052*	H	0.53	1.002*	-236	68	20	105
35	L	1.34	0.883	H	1.22	0.820	h	0.57	1.207*	-252	73	-106	125
36	P	0.59	0.850	B	0.62	0.740	b	1.56	1.327*	-177	32	71	128
37	S	0.79	0.888	i	0.72	0.985	b	1.35	1.072*	-151	58	111	115
38	S	0.79	0.998	i	0.72	1.217*	b	1.35	0.867	-89	78	21	45
39	R	0.79	1.073*	i	0.90	1.337*	I	1.05	0.777	-54	159	-9	-22
40	I	1.00	1.073*	I	1.60	1.355*	H	0.54	0.675	-24	222	-33	-48
41	V	1.14	1.028*	h	1.65	1.132*	H	0.53	0.775	-36	218	-30	-5
42	T	0.82	0.937	I	1.20	0.900	h	0.99	0.980	-81	188	-2	27
43	A	1.45	0.932	H	0.97	0.762	I	0.64	1.110*	-145	97	25	55
44	R	1.24	0.822	h	0.71	0.675	b	0.94	1.349*	-203	9	97	111
45	S	0.79	0.747	I	0.72	0.677	b	1.35	1.442*	-264	-87	86	175
46	N	0.73	0.857	b	0.65	0.722	b	1.51	1.367*	-251	-151	27	191
47	P	0.59	0.977	B	0.62	0.785	b	1.56	1.252*	-117	-123	138	188
48	S	0.79	1.065*	I	0.72	0.872	b	1.35	1.022*	-69	-87	166	160
49	R	0.79	1.188*	I	0.90	0.935	I	1.05	0.845	-14	-96	31	53
50	R	0.79	1.188*	I	0.90	1.030*	I	1.05	0.742	31	-21	-34	3
51	A	1.45	1.223*	H	0.97	0.870	I	0.64	0.685	75	17	-35	15
52	A	1.45	1.070*	H	0.97	0.807	I	0.64	0.862	120	62	-35	15
53	P	1.12	0.993	h	1.28	0.965	h	0.64	0.837	91	71	-43	-6
54	E	1.53	0.945	H	0.26	0.847	B	0.82	1.062*	48	40	-27	-4
55	S	0.79	0.880	I	0.72	1.182*	b	1.35	0.992	-11	8	11	5
56	I	1.00	0.972	I	1.60	1.302*	H	0.54	0.902	-36	32	-16	-38
57	G	0.53	0.937	B	0.81	1.315*	I	1.54	0.900	-58	38	-15	-46
58	I	1.00	0.980	I	1.60	1.417*	H	0.54	0.657	-31	87	-36	-8
59	T	0.82	0.935	I	1.20	1.197*	h	0.99	0.860	-48	103	-52	22
60	V	1.14	0.962	h	1.65	1.077*	H	0.53	0.950	-91	128	-50	45
61	L	1.34	0.935	H	1.22	0.827	h	0.57	1.195*	-138	123	9	65
62	S	0.79	0.882	I	0.72	0.722	b	1.35	1.410*	-159	38	81	120
63	S	0.79	0.960	I	0.72	0.742	b	1.35	1.430*	-161	-47	118	135
64	N	0.73	0.860	b	0.65	0.975	b	1.51	1.225*	-131	-51	112	101
65	D	0.98	1.060*	i	0.80	1.225*	i	1.43	0.980	-70	-34	76	50
66	D	0.98	1.048*	i	0.80	1.250*	i	1.43	0.885	-20	61	11	-10
67	V	1.14	1.007*	h	1.65	1.292*	H	0.53	0.687	24	103	-55	-15
68	V	1.14	1.007*	h	1.65	1.060*	H	0.53	0.892	44	128	-70	-30
69	R	0.79	1.007*	I	0.90	0.810	I	1.05	1.137*	41	104	-49	-42
70	A	1.45	1.065*	H	0.97	0.997	I	0.64	1.007*	20	77	-45	-50
71	S	0.79	1.010*	i	0.72	1.167*	b	1.35	0.980	11	58	-54	-30
72	N	0.73	1.010*	b	0.65	1.400*	b	1.51	0.775	-11	39	-63	-44
73	V	1.14	1.078*	h	1.65	1.557*	H	0.53	0.557	-11	78	-65	-40
74	V	1.14	1.067*	h	1.65	1.325*	H	0.53	0.762	-36	168	-85	-60
75	V	1.14	0.975	h	1.65	1.325*	H	0.53	0.762	-51	198	-60	-50
76	F	1.12	0.980	h	1.28	1.097*	h	0.64	0.867	-29	176	-38	-71
77	S	0.79	1.017*	I	0.72	0.932	b	1.35	1.097*	-9	103	-19	-25
78	V	1.14	1.075*	h	1.65	1.660*	H	0.53	1.002*	-26	28	-75	-33
79	K	1.07	1.063*	I	0.74	0.952	b	0.95	1.012*	-42	-38	-165	-28
80	P	0.59	1.048*	B	0.62	1.188*	b	1.56	0.907	6	-78	-19	-7
81	Q	1.17	1.140*	h	1.23	1.210*	h	0.97	0.755	35	-68	9	-25
82	L	1.34	1.135*	H	1.22	1.102*	h	0.57	0.870	72	-37	-106	-70
83	V	1.14	1.135*	h	1.65	1.210*	H	0.53	0.860	104	38	-135	-133
84	K	1.07	1.123*	I	0.74	1.210*	b	0.95	0.860	146	107	-130	-153
85	D	0.98	1.168*	i	0.80	1.330*	I	1.43	0.765	180	151	-94	-135
86	V	1.14	1.183*	h	1.65	1.315*	H	0.53	0.645	182	168	-135	-140
87	V	1.14	1.092*	h	1.65	1.207*	H	0.53	0.655	174	148	-140	-120
88	L	1.34	1.125*	H	1.22	0.980	h	0.57	0.760	160	108	-116	-113
89	K	1.07	1.125*	I	0.74	0.830	b	0.95	1.007*	138	47	-70	-103

residue	Pa	Pa avg.	Ha	Pb	Pb avg.	Hb	Pt	Pt avg.	Ph	Ps	Ptu	Pco
90 L	1.34	1.083*	H	1.22	0.950	h	0.57	0.912	105	3	-146	-88
91 K	1.07	1.038*	I	0.74	0.950	b	0.95	0.912	86	-38	-225	-98
92 P	0.59	1.023*	B	0.62	1.065*	b	1.56	0.922	134	-43	-114	-82
93 L	1.34	1.103*	H	1.22	1.095*	h	0.57	0.770	142	-12	-78	-80
94 L	1.34	1.103*	H	1.22	0.990	h	0.57	0.985	157	-27	-136	-100
95 T	0.82	1.103*	I	1.20	0.870	h	0.99	1.080*	142	-37	-102	-121
96 K	1.07	1.157*	I	0.74	0.875	b	0.95	0.975	141	-43	-80	-108
97 D	0.98	1.110*	I	0.80	0.995	I	1.43	0.880	128	-34	-49	-108
98 K	1.07	1.137*	I	0.74	1.207*	b	0.95	0.655	116	12	-70	-78
99 L	1.34	1.200*	H	1.22	1.202*	h	0.57	0.755	112	93	-126	-70
100 L	1.34	1.218*	H	1.22	1.310*	h	0.57	0.745	142	153	-166	-85
101 V	1.14	1.083*	h	1.65	1.247*	H	0.53	0.762	172	143	-185	-120
102 S	0.79	1.060*	I	0.72	1.077*	b	1.35	0.790	209	123	-194	-110
103 V	1.14	1.107*	h	1.65	1.100*	H	0.53	0.837	244	68	-210	-135
104 A	1.45	1.140*	H	0.97	1.087*	I	0.64	0.840	240	7	-170	-125
105 A	1.45	1.077*	H	0.97	1.030*	I	0.64	0.917	220	-38	-145	-120
106 G	0.53	0.998	B	0.81	1.092*	I	1.54	0.900	159	-87	-88	-81
107 I	1.00	1.133*	I	1.60	1.075*	H	0.54	0.752	129	-43	-91	-81
108 K	1.07	1.162*	I	0.74	0.875	b	0.95	0.975	103	-38	-115	-48
109 L	1.34	1.238*	H	1.22	0.995	h	0.57	0.880	52	-12	-76	-13
110 K	1.07	1.205*	I	0.74	0.997	b	0.95	0.980	61	-8	-35	2
111 D	0.98	1.268*	I	0.80	0.877	I	1.43	0.947	73	-39	-49	0
112 L	1.34	1.193*	H	1.22	0.975	h	0.57	0.840	157	-52	-86	-5
113 Q	1.17	1.177*	h	1.23	0.912	h	0.97	0.857	170	-43	-56	5
114 E	1.53	1.237*	H	0.26	0.807	B	0.82	1.000*	191	-70	-2	-4
115 W	1.14	1.113*	h	1.19	0.920	h	1.00	1.030*	200	-80	16	32
116 A	1.45	1.110*	H	0.97	0.687	I	0.64	0.985	203	-138	-25	70
117 G	0.53	1.035*	B	0.81	0.670	I	1.54	1.087*	192	-162	57	99
118 H	1.24	1.078*	h	0.71	0.787	b	0.94	0.862	195	-135	57	121
119 E	1.53	1.062*	H	0.26	1.010*	B	0.82	0.762	113	-105	23	66
120 R	0.79	1.007*	I	0.90	1.170*	I	1.05	0.820	79	-6	61	23
121 F	1.12	0.973	h	1.28	1.357*	h	0.64	0.690	44	86	102	4
122 I	1.00	0.908	I	1.60	1.455*	H	0.54	0.692	24	102	104	17
123 R	0.79	0.878	I	0.90	1.210*	I	1.05	0.947	-51	149	116	58
124 V	1.14	0.845	h	1.65	1.147*	H	0.53	1.062*	-116	103	60	75
125 M	1.20	0.897	h	1.67	1.035*	H	0.65	1.177*	-202	18	-8	124
126 P	0.59	0.938	B	0.62	0.772	b	1.56	1.405*	-202	-58	124	153
127 N	0.73	1.030*	b	0.65	0.860	b	1.51	1.175*	-226	-116	117	176
128 T	0.82	0.997	I	1.20	0.940	h	0.99	0.957	-191	-112	-82	127
129 P	0.59	1.055*	B	0.62	1.052*	b	1.56	0.842	-72	-78	-9	78
130 A	1.45	1.198*	H	0.97	1.100*	I	0.64	0.837	10	-3	15	40
131 A	1.45	1.198*	H	0.97	1.165*	I	0.64	0.920	105	52	-90	-15
132 V	1.14	1.088*	h	1.65	1.165*	H	0.53	0.920	149	33	-135	-40
133 G	0.53	1.088*	B	0.81	0.995	I	1.54	0.947	119	38	-98	-26
134 Q	1.17	1.200*	h	1.23	0.972	h	0.97	0.900	130	62	-126	-65
135 A	1.45	1.137*	H	0.97	1.077*	I	0.64	0.790	100	62	-150	-55
136 A	1.45	1.118*	H	0.97	1.252*	I	0.64	0.792	85	77	-110	-30
137 S	0.79	0.965	I	0.72	1.190*	b	1.35	0.970	76	143	-89	-10
138 V	1.14	0.922	h	1.65	1.315*	H	0.53	0.775	74	198	-80	-35
139 M	1.20	0.973	h	1.67	1.105*	H	0.65	1.027*	46	193	-13	19
140 S	0.79	1.015*	I	0.72	0.890	b	1.35	1.250*	21	108	26	90
141 L	1.34	1.020*	H	1.22	0.952	h	0.57	1.072*	7	-2	24	110
142 G	0.53	1.052*	B	0.81	0.890	I	1.54	1.090*	-16	-107	12	99
143 G	0.53	1.218*	B	0.81	0.987	I	1.54	0.952	2	-152	-38	109
144 A	1.45	1.293*	H	0.97	0.850	I	0.64	0.772	60	-108	-107	100
145 A	1.45	1.293*	H	0.97	0.672	I	0.64	0.817	130	-38	-185	45
146 T	0.82	1.173*	I	1.20	0.630	h	0.99	1.015*	204	-32	-177	2
147 E	1.53	1.260*	H	0.26	0.572	B	0.82	0.927	223	-50	-157	-64
148 E	1.53	1.172*	H	0.26	0.670	B	0.82	1.100*	236	-75	-107	-109
149 D	0.98	1.048*	I	0.80	0.910	I	1.43	1.037*	241	-119	-19	-70
150 A	1.45	1.080*	H	0.97	1.110*	I	0.64	0.815	234	-138	15	-15
151 N	0.73	1.062*	b	0.65	1.047*	b	1.51	0.992	225	-106	-3	16
152 L	1.34	1.127*	H	1.22	1.192*	h	0.57	0.857	220	-12	-46	-5
153 I	1.00	0.992	I	1.60	1.192*	H	0.54	0.857	186	42	-56	-38
154 S	0.79	0.957	I	0.72	1.112*	b	1.35	0.882	152	63	-64	-35
155 Q	1.17	0.992	h	1.23	1.135*	h	0.97	0.930	105	92	-1	5
156 L	1.34	0.885	H	1.22	1.007*	h	0.57	1.025*	37	103	19	35
157 F	1.12	0.840	h	1.28	1.102*	h	0.64	1.017*	-19	106	37	9
158 G	0.53	0.820	B	0.81	0.985	I	1.54	1.242*	-68	33	92	79
159 S	0.79	0.922	I	0.72	0.967	b	1.35	1.095*	-49	13	141	145
160 I	1.00	0.968	I	1.60	1.187*	H	0.54	0.892	-54	17	99	132
161 G	0.53	1.043*	B	0.81	1.085*	I	1.54	1.007*	-31	-12	102	116
162 K	1.07	1.118*	I	0.74	1.067*	b	0.95	0.860	23	7	80	82
163 I	1.00	1.103*	I	1.60	1.125*	H	0.54	0.782	31	12	34	27
164 W	1.14	1.115*	h	1.19	0.925	h	1.00	1.005*	27	5	-64	-36
165 K	1.07	1.112*	I	0.74	0.827	b	0.95	1.112*	43	-18	-100	-38
166 A	1.45	1.120*	H	0.97	0.827	I	0.64	1.112*	65	-83	-5	-15
167 D	0.98	1.042*	I	0.80	0.905	I	1.43	1.112*	80	-124	36	10
168 D	0.98	1.120*	I	0.80	1.025*	I	1.43	0.915	30	-104	56	7
169 K	1.07	1.123*	I	0.74	1.025*	b	0.95	0.915	63	-88	35	17
170 P	1.12	1.082*	h	1.28	1.082*	h	0.64	0.837	96	-9	37	19
171 P	1.12	0.983	h	1.28	1.162*	h	0.64	0.812	111	56	-23	-56
172 D	0.98	1.020*	I	0.80	1.142*	I	1.43	0.900	100	61	-64	-75
173 A	1.45	0.988	H	0.97	1.145*	I	0.64	0.927	65	67	-87	-80
174 I	1.00	0.835	I	1.60	1.207*	H	0.54	0.910	-4	52	-26	-18
175 T	0.82	0.800	I	1.20	0.987	h	0.99	1.112*	-73	18	18	22
176 G	0.53	0.752	B	0.81	0.890	I	1.54	1.250*	-171	-27	112	119
177 L	1.34	0.762	H	1.22	0.867	h	0.57	1.202*	-203	-17	159	165
178 S	0.79	0.780	I	0.72	0.765	b	1.35	1.445*	-259	-32	176	225
179 G	0.53	0.750	B	0.81	0.740	I	1.54	1.497*	-313	-52	247	234
180 S	0.79	0.828	I	0.72	0.780	b	1.35	1.272*	-294	-12	186	270
181 G	0.53	0.798	B	0.81	0.922	I	1.54	1.217*	-291	-2	27	204
182 P	0.59	0.933	B	0.62	1.120*	b	1.56	0.967	-172	42	91	128

residue	Pa	Pa avg.	Ha	Pb	Pb avg.	Hb	Pt	Pt avg.	Ph	Ps	Ptu	Pco
183 A	1.45	1.077•	H	0.97	1.287•	I	0.64	0.860	-110	92	40	50
184 Y	0.61	1.002•	b	1.29	1.350•	h	1.13	0.842	-45	140	-61	-36
185 I	1.00	1.155•	I	1.60	1.270•	H	0.54	0.720	9	197	-111	-73
186 Y	0.61	1.230•	b	1.29	1.270•	h	1.13	0.720	60	225	-141	-71
187 L	1.34	1.352•	H	1.22	1.012•	h	0.57	0.642	102	183	-166	-80
188 A	1.45	1.370•	H	0.97	0.950	I	0.64	0.660	155	167	-165	-110
189 I	1.00	1.292•	I	1.60	1.012•	H	0.54	0.642	169	112	-211	-123
190 E	1.53	1.213•	H	0.26	0.855	B	0.82	0.667	176	100	-202	-149
191 A	1.45	1.047•	H	0.97	0.990	I	0.64	0.820	178	27	-145	-100
192 L	1.34	0.995	H	1.22	0.950	h	0.57	1.045•	170	-32	-91	-40
193 A	1.45	1.013•	H	0.97	0.847	I	0.64	1.287•	153	-63	-5	-10
194 D	0.98	1.013•	i	0.80	1.017•	i	1.43	1.260•	121	-89	56	30
195 G	0.53	0.938	B	0.81	1.060•	i	1.54	1.062•	74	-82	52	54
196 G	0.53	1.073•	B	0.81	1.100•	i	1.54	0.837	49	-27	22	44
197 V	1.14	1.083•	h	1.65	1.100•	H	0.53	0.837	44	43	-60	25
198 A	1.45	1.025•	H	0.97	0.992	I	0.64	0.847	20	62	-65	25
199 A	1.45	0.947	H	0.97	0.905	I	0.64	1.077•	-22	72	-35	10
200 G	0.53	0.928	B	0.81	0.887	i	1.54	1.180•	-76	63	-83	4
201 L	1.34	1.082•	H	1.22	0.885	h	0.57	1.152•	-68	48	-206	5
202 P	0.59	1.082•	B	0.62	0.885	b	1.56	1.152•	13	17	-79	-12
203 R	0.79	1.115•	i	0.90	0.972	i	1.05	0.922	47	-11	-9	-47
204 D	0.98	1.207•	i	0.80	1.052•	i	1.43	0.802	95	1	-49	-65
205 L	1.34	1.285•	H	1.22	1.032•	h	0.57	0.782	120	53	-111	-70
206 A	1.45	1.193•	H	0.97	1.032•	I	0.64	0.782	125	67	-95	-40
207 L	1.34	1.147•	H	1.22	1.032•	h	0.57	0.782	118	73	-76	-5
208 S	0.79	1.060•	i	0.72	0.907	b	1.35	0.977	91	33	-49	35
209 L	1.34	1.118•	H	1.22	1.035•	h	0.57	0.882	90	8	56	10
210 A	1.45	1.118•	H	0.97	1.030•	I	0.64	0.987	105	17	-30	-5
211 S	0.79	0.965	i	0.72	1.200•	b	1.35	0.960	99	38	-26	-5
212 Q	1.17	1.075•	h	1.23	1.325•	h	0.97	0.765	85	82	-61	-30
213 T	0.82	1.122•	i	1.20	1.220•	h	0.99	0.907	62	108	-92	-28
214 V	1.14	1.117•	h	1.65	1.162•	H	0.53	0.820	59	108	-105	-20
215 L	1.34	1.127•	H	1.22	0.992	h	0.57	0.847	62	68	-56	25
216 G	0.53	1.145•	B	0.81	0.867	i	1.54	1.042•	74	28	-23	39
217 A	1.45	1.193•	H	0.97	1.082•	I	0.64	0.820	113	47	-35	55
218 A	1.45	1.175•	H	0.97	1.082•	I	0.64	0.820	150	57	-95	25
219 S	0.79	1.065•	i	0.72	1.140•	b	1.35	0.907	176	23	-114	-10
220 M	1.20	1.022•	h	1.67	1.265•	H	0.65	0.712	178	48	-115	-76
221 A	1.45	1.000•	H	0.97	1.027•	I	0.64	0.887	153	32	-135	-110
222 T	0.82	0.965	i	1.20	0.987	h	0.99	1.112•	104	43	-57	-23
223 L	1.34	0.927	H	1.22	0.872	h	0.57	1.102•	52	3	24	40
224 S	0.79	0.792	i	0.72	0.745	b	1.35	1.195•	16	-52	91	90
225 G	0.53	0.855	B	0.81	0.720	i	1.54	1.247•	-43	-67	147	101
226 K	1.07	0.990	I	0.74	0.720	b	0.95	1.247•	-72	-73	90	132
227 H	1.24	0.990	h	0.71	0.842	b	0.94	1.252•	-123	-70	-53	146
228 P	0.59	0.947	B	0.62	0.970	b	1.56	1.160•	-97	-63	71	133
229 G	0.53	1.012•	B	0.81	1.000•	i	1.54	1.007•	-113	-27	112	59
230 Q	1.17	1.113•	h	1.23	0.997	h	0.97	0.980	-75	37	14	0
231 L	1.34	1.055•	H	1.22	0.890	h	0.57	1.095•	-103	93	-1	22
232 K	1.07	0.963	I	0.74	0.997	b	0.95	1.085•	-117	127	55	22
233 D	0.98	0.883	i	0.80	1.112•	i	1.43	1.095•	-127	156	121	60
234 D	0.98	0.808	i	0.80	1.092•	i	1.43	1.075•	-135	206	144	85
235 V	1.14	0.733	h	1.65	1.047•	H	0.53	1.107•	-161	233	130	120
236 T	0.82	0.680	i	1.20	0.837	h	0.99	1.360•	-191	203	103	172
237 S	0.79	0.680	i	0.72	0.740	b	1.35	1.497•	-239	98	56	235
238 P	0.59	0.715	B	0.62	0.860	b	1.56	1.407•	-197	32	206	203
239 G	0.53	0.858	B	0.81	1.005•	i	1.54	1.265•	-211	-32	245	189
240 G	0.53	0.858	B	0.81	1.202•	i	1.54	1.015•	-181	-47	135	149
241 T	0.82	0.960	i	1.20	1.242•	h	0.99	0.790	-136	18	33	137
242 T	0.82	1.030•	i	1.20	1.145•	h	0.99	0.927	-96	113	-32	102
243 I	1.00	1.148•	I	1.60	1.257•	H	0.54	0.812	-29	162	-66	47
244 A	1.45	1.205•	H	0.97	1.035•	I	0.64	0.912	30	157	-100	20
245 G	0.53	1.218•	B	0.81	0.857	i	1.54	0.957	39	103	-93	9
246 V	1.14	1.308•	h	1.65	0.960	H	0.53	0.715	104	93	-140	20
247 H	1.24	1.207•	h	0.71	0.612	b	0.94	0.787	115	40	-163	16
248 E	1.53	1.088•	H	0.26	0.620	B	0.82	0.790	118	5	-127	-24
249 L	1.34	1.020•	H	1.22	0.757	h	0.57	0.970	110	-17	-51	-5
250 E	1.53	0.928	H	0.26	0.655	B	0.82	1.212•	96	-60	68	8
251 K	1.07	0.762	I	0.74	0.910	b	0.95	1.167•	87	-118	150	42
252 G	0.53	0.720	B	0.81	0.950	i	1.54	1.192•	45	-132	152	109
253 G	0.53	0.855	B	0.81	0.950	i	1.54	1.192•	32	-137	122	99
254 P	1.12	0.967	h	1.28	1.047•	h	0.64	1.055•	54	-54	42	44
255 R	0.79	0.902	i	0.90	1.032•	i	1.05	1.037•	31	-31	-1	23
256 G	0.53	1.012•	B	0.81	1.225•	i	1.54	0.937	-18	-22	7	39
257 T	0.82	1.113•	i	1.20	1.185•	h	0.99	0.930	-11	8	8	62
258 L	1.34	1.167•	H	1.22	1.127•	h	0.57	0.842	2	-2	-46	65
259 M	1.20	1.185•	h	1.67	1.235•	H	0.65	0.832	33	28	-63	4
260 N	0.73	1.227•	b	0.65	1.230•	b	1.51	0.802	87	9	-118	-44
261 A	1.45	1.347•	H	0.97	1.310•	I	0.64	0.585	165	27	-180	-85
262 V	1.14	1.283•	h	1.65	1.310•	H	0.53	0.585	234	53	-260	-135
263 V	1.14	1.225•	h	1.65	1.140•	H	0.53	0.612	284	43	-275	-155
264 A	1.45	1.167•	H	0.97	0.912	I	0.64	0.717	300	-33	-260	-165
265 A	1.45	1.057•	H	0.97	0.895	I	0.64	0.820	285	-73	-210	-120
266 A	1.45	1.070•	H	0.97	0.832	I	0.64	0.997	225	-68	-160	-93
267 K	1.07	1.052•	I	0.74	0.815	b	0.95	1.100•	183	-98	-115	-48
268 R	0.79	1.005•	i	0.90	0.695	i	1.05	1.067•	121	-106	-74	-52
269 S	0.79	1.048•	i	0.72	0.775	b	1.35	0.947	111	-87	-24	0
270 R	0.79	1.112•	i	0.90	0.775	i	1.05	0.947	74	-56	-14	-37
271 E	1.53	1.220•	H	0.26	0.733	B	0.82	0.913	48	-40	-27	-39
272 L	1.34	1.065•	H	1.22	0.970	h	0.57	0.960	20	-42	4	0
273 S	0.79	0.790	i	0.72	0.720	b	1.35	1.350•	-16	-57	46	50

iv) Numerical data used in the prediction of the secondary structure of PSCR from *Escherichia coli* (Deutch *et al.*, 1982).

residue		Pa	Pa avg	Ba	Pb	Pb avg	Bb	Pt	Pt avg	Ph	Ps	Ptu	Pco
1	M	1.20	1.067*	h	1.67	0.852	H	0.65	0.842	243	-72	-103	-81
2	E	1.53	1.053*	B	0.26	0.835	B	0.82	0.815	203	-70	-52	-102
3	K	1.07	0.965	I	0.74	0.972	b	0.95	0.998	151	-38	10	-76
4	K	1.07	0.875	I	0.74	1.107*	b	0.95	0.917	111	-8	55	-48
5	I	1.00	0.825	I	1.60	1.322*	H	0.54	0.815	69	37	44	-28
6	G	0.53	0.747	B	0.81	1.125*	I	1.54	1.065*	12	78	17	-31
7	F	1.12	0.780	h	1.28	1.247*	h	0.64	1.002*	6	111	12	-56
8	I	1.00	0.793	I	1.60	1.130*	H	0.54	1.227*	-19	87	-16	-18
9	G	0.53	0.715	B	0.81	0.892	I	1.54	1.470*	-86	23	52	19
10	C	0.77	0.805	I	1.30	1.107*	b	1.29	1.247*	-113	-1	119	43
11	G	0.53	0.918	B	0.81	0.985	I	1.54	1.310*	-141	-82	147	139
12	N	0.72	0.997	b	0.65	0.967	h	1.51	1.162*	-98	-81	142	141
13	M	1.20	1.098*	h	1.67	1.047*	H	0.65	0.965	-62	-37	87	104
14	G	0.53	0.987	B	0.81	1.030*	I	1.54	0.917	-56	-22	52	56
15	K	1.07	0.987	I	0.74	1.132*	b	0.95	0.875	8	82	-10	2
16	A	1.45	1.032*	H	0.97	1.150*	I	0.64	0.823	28	117	-70	-30
17	I	1.00	0.957	I	1.60	1.110*	H	0.54	1.047*	31	117	-91	-28
18	L	1.34	1.032*	H	1.22	1.015*	h	0.57	1.055*	12	163	-101	-50
19	G	0.53	0.940	B	0.81	1.110*	I	1.54	1.047*	-38	93	-58	-46
20	G	0.53	0.940	B	0.81	1.150*	I	1.54	0.822	-53	83	-18	29
21	L	1.34	1.047*	H	1.22	1.127*	h	0.57	0.775	-53	98	-41	60
22	I	1.00	1.013*	I	1.60	1.025*	H	0.54	1.017*	-69	87	-11	77
23	A	1.45	1.070*	H	0.97	0.932	I	0.64	1.125*	-82	117	45	85
24	S	0.79	0.927	I	0.72	1.102*	b	1.35	1.097*	-99	128	81	100
25	G	0.53	0.883	B	0.81	1.227*	I	1.54	0.902	-151	113	87	59
26	Q	1.17	0.990	h	1.23	1.180*	h	0.97	0.907	-167	122	69	80
27	V	1.14	0.962	h	1.65	1.075*	H	0.53	1.050*	-221	108	-5	120
28	L	1.34	0.962	H	1.22	0.970	h	0.57	1.160*	-253	88	-36	165
29	P	0.59	0.928	B	0.62	1.065*	b	1.56	1.152*	-197	122	141	158
30	G	0.53	0.932	B	0.81	1.207*	I	1.54	1.012*	-168	158	227	139
31	Q	1.17	0.980	h	1.23	1.417*	h	0.97	0.760	-150	237	159	120
32	I	1.00	0.883	I	1.60	1.432*	H	0.54	0.800	-179	277	149	92
33	W	1.14	0.848	h	1.19	1.332*	h	1.00	0.912	-208	305	111	77
34	V	1.14	0.757	h	1.65	1.190*	H	0.53	1.052*	-236	318	103	105
35	Y	0.61	0.730	b	1.29	0.957	h	1.13	1.257*	-285	255	119	124
36	T	0.82	0.807	I	1.20	0.790	h	0.99	1.365*	-331	153	58	177
37	P	0.59	0.860	B	0.62	0.690	b	1.56	1.475*	-307	17	171	223
38	S	0.79	1.003*	I	0.72	0.720	b	1.35	1.322*	-264	-107	146	255
39	P	0.59	1.113*	B	0.62	0.952	b	1.56	1.117*	-102	-138	171	183
40	D	0.98	1.238*	I	0.80	1.040*	I	1.43	0.887	-15	-114	146	82
41	K	1.07	1.282*	I	0.74	1.082*	b	0.95	0.690	88	-3	-25	7
42	V	1.14	1.267*	h	1.65	1.202*	H	0.53	0.595	154	83	-105	-50
43	A	1.45	1.272*	H	0.97	0.967	I	0.64	0.697	183	137	-145	-85
44	A	1.45	1.217*	H	0.97	0.925	I	0.64	0.895	205	122	-170	-105
45	L	1.34	1.063*	H	1.22	0.990	h	0.57	0.977	187	73	-156	-90
46	R	1.24	1.007*	h	0.71	1.005*	b	0.94	0.995	177	45	-78	-54
47	D	0.98	0.522	I	0.80	1.030*	I	1.43	1.145*	133	-44	16	0
48	Q	1.17	1.000*	h	1.23	1.230*	h	0.97	0.922	100	-13	44	15
49	F	1.12	1.047*	h	1.28	1.085*	h	0.64	1.057*	61	6	42	34
50	G	0.53	1.115*	B	0.81	1.007*	I	1.54	1.057*	19	3	32	69
51	I	1.00	1.158*	I	1.60	1.047*	H	0.54	0.832	41	-23	4	112
52	N	0.73	1.233*	b	0.65	0.712	b	1.51	0.902	54	-56	-18	141
53	A	1.45	1.307*	H	0.97	0.738	I	0.64	0.862	115	-98	-10	145
54	A	1.45	1.320*	H	0.97	0.730	I	0.64	0.862	155	-123	-20	120
55	E	1.53	1.268*	H	0.26	0.795	B	0.82	0.945	188	-150	-27	91
56	S	0.79	1.255*	I	0.72	0.795	b	1.35	0.945	214	-157	-34	70
57	A	1.45	1.318*	H	0.97	1.037*	I	0.64	0.740	283	-148	-65	-5
58	Q	1.17	1.243*	h	1.23	1.027*	h	0.97	0.740	308	-73	-76	-40
59	E	1.53	1.290*	H	0.26	1.027*	B	0.82	0.740	321	-30	-103	-69
60	V	1.14	1.198*	h	1.65	1.362*	H	0.53	0.670	327	-27	-130	-80
61	A	1.45	1.175*	H	0.97	1.192*	I	0.64	0.697	318	12	-105	-75
62	Q	1.17	1.100*	h	1.23	1.150*	h	0.97	0.895	303	57	-76	-70
63	I	1.00	1.092*	I	1.60	1.242*	H	0.54	0.787	269	73	-86	-28
64	A	1.45	1.167*	H	0.97	1.242*	I	0.64	0.787	255	82	-80	-40
65	D	0.98	1.167*	I	0.80	1.320*	I	1.43	0.787	240	81	-44	-45
66	I	1.00	1.193*	I	1.60	1.362*	H	0.54	0.690	216	127	-96	-73
67	I	1.00	1.205*	I	1.60	1.205*	H	0.54	0.615	186	162	-146	-98
68	F	1.12	1.137*	h	1.28	1.217*	h	0.64	0.612	156	148	-108	-91
69	A	1.45	1.038*	H	0.97	1.082*	I	0.64	0.690	145	57	-60	-35
70	A	1.45	0.963	H	0.97	0.995	I	0.64	0.920	130	-23	-40	-10
71	V	1.14	0.922	h	1.65	0.955	H	0.53	1.145*	84	-97	-80	2
72	K	1.07	0.898	I	0.74	0.942	b	0.95	1.147*	28	-143	-105	42
73	P	0.59	0.898	B	0.62	1.175*	b	1.56	1.072*	63	-148	21	43
74	G	0.53	0.990	B	0.81	1.420*	I	1.54	0.817	34	-97	52	-1
75	I	1.00	1.125*	I	1.60	1.402*	H	0.54	0.670	81	12	-126	-123
76	M	1.20	1.090*	h	1.67	1.415*	H	0.65	0.667	83	78	-218	-151
77	I	1.00	1.145*	I	1.60	1.302*	H	0.54	0.647	66	137	-166	-96
78	K	1.07	1.145*	I	0.74	1.082*	b	0.95	0.850	66	177	-158	-73
79	V	1.14	1.103*	h	1.65	0.962	H	0.53	0.817	39	153	-170	-55
80	L	1.34	1.045*	H	1.22	0.950	h	0.57	0.820	42	113	-151	-55
81	S	0.79	0.953	I	0.72	0.945	b	1.35	0.925	41	58	-109	-25
82	E	1.53	1.045*	B	0.26	0.945	B	0.82	0.925	56	20	-79	-74
83	I	1.00	0.912	I	1.60	1.060*	H	0.54	1.057*	59	-8	-51	-40
84	T	0.82	0.923	I	1.20	0.965	h	0.99	1.065*	37	-32	-32	-23
85	S	0.79	0.950	I	0.72	0.837	b	1.35	1.195*	7	-37	16	20
86	S	0.79	0.950	I	0.72	0.832	b	1.35	1.095*	-1	-47	21	50
87	L	1.34	1.042*	H	1.22	0.852	h	0.57	1.115*	-13	-97	44	80
88	N	0.73	1.008*	b	0.65	0.727	b	1.51	1.318*	-41	-136	97	78
89	K	1.07	1.077*	I	0.74	0.870	b	0.95	1.075*	-19	-123	85	67

residue	Pa	Pa avg	Ha	Pb	Pb avg	Hb	Pt	Pt avg	Ph	Ps	Ptu	Pco
90 D	0.98	1.030•	i	0.80	1.097•	i	1.43	0.970	-7	-64	61	20
91 S	0.79	1.033•	i	0.72	1.310•	b	1.35	0.745	1	63	-4	25
92 L	1.34	1.143•	H	1.22	1.310•	h	0.57	0.745	42	158	-91	-35
93 V	1.14	1.162•	h	1.65	1.405•	H	0.53	0.737	74	228	-140	-80
94 V	1.14	1.060•	h	1.65	1.235•	H	0.53	0.765	82	243	-145	-80
95 S	0.79	1.060•	i	0.72	1.065•	b	1.35	0.792	76	188	-114	-35
96 I	1.00	1.065•	I	1.60	1.087•	H	0.54	0.840	66	142	-116	-43
97 A	1.45	1.122•	H	0.97	1.100•	I	0.64	0.837	55	92	-85	-40
98 A	1.45	1.043•	H	0.97	1.157•	I	0.64	0.925	45	32	-45	-10
99 G	0.53	0.997	B	0.81	1.220•	i	1.54	0.907	9	8	-30	-11
100 V	1.14	1.132•	h	1.65	1.217•	H	0.53	0.880	37	43	-95	-30
101 T	0.82	1.183•	i	1.20	1.112•	h	0.99	0.990	49	63	-107	-13
102 L	1.34	1.178•	H	1.22	1.117•	h	0.57	0.885	87	48	-91	-10
103 D	0.98	1.197•	i	0.80	1.055•	i	1.43	0.902	128	46	-54	-55
104 Q	1.17	1.257•	h	1.23	1.080•	h	0.97	0.807	163	57	-101	-75
105 L	1.34	1.150•	H	1.22	1.015•	h	0.57	0.725	177	28	-111	-35
106 A	1.45	1.133•	H	0.97	1.015•	I	0.64	0.725	200	72	-80	-35
107 R	0.79	1.055•	i	0.90	0.975	i	1.05	0.950	217	44	-49	-42
108 A	1.45	1.055•	H	0.97	0.927	I	0.64	0.922	220	42	-15	-5
109 L	1.34	0.992	H	1.22	0.885	h	0.57	1.120•	202	-32	-1	25
110 G	0.53	0.935	B	0.81	0.805	i	1.54	1.240•	139	-102	57	64
111 H	1.24	1.013•	h	0.71	0.787	b	0.94	1.092•	140	-105	62	71
112 D	0.98	0.938	i	0.80	1.010•	i	1.43	0.992	80	-104	46	50
113 R	0.79	1.017•	i	0.90	1.210•	i	1.05	0.770	41	-16	11	-25
114 K	1.07	1.085•	I	0.74	1.210•	b	0.95	0.770	43	82	0	-48
115 I	1.00	1.005•	I	1.60	1.267•	H	0.54	0.692	21	177	4	-33
116 I	1.00	0.960	I	1.60	1.285•	H	0.54	0.720	-9	172	-1	-8
117 R	0.79	0.930	i	0.90	1.040•	i	1.05	0.975	-89	174	31	-2
118 A	1.45	0.897	H	0.97	0.977	I	0.64	1.090•	-155	87	5	25
119 M	1.20	0.897	h	1.67	1.035•	H	0.65	1.177•	-217	53	-68	79
120 P	0.59	0.920	B	0.62	0.772	b	1.56	1.405•	-187	-28	99	123
121 N	0.73	1.012•	b	0.65	0.860	b	1.51	1.175•	-201	-61	102	151
122 T	0.82	1.012•	i	1.20	1.002•	h	0.99	0.940	-156	-72	-97	97
123 P	0.59	1.117•	B	0.62	1.115•	b	1.56	0.825	-59	-18	-39	33
124 A	1.45	1.107•	H	0.97	1.122•	I	0.64	0.812	15	32	-15	20
125 L	1.34	1.065•	H	1.22	1.122•	h	0.57	0.812	77	98	-111	-10
126 V	1.14	0.978	h	1.65	1.020•	H	0.53	1.055•	79	118	-95	-10
127 N	0.73	0.920	b	0.65	1.025•	b	1.51	1.085•	37	114	-18	21
128 A	1.45	0.988	H	0.97	1.162•	I	0.64	0.955	15	102	50	60
129 G	0.53	0.883	B	0.81	1.100•	i	1.54	1.132•	-51	88	70	59
130 M	1.20	0.893	h	1.67	1.310•	H	0.65	0.880	-62	128	47	59
131 T	0.82	0.815	i	1.20	1.192•	h	0.99	0.965	-81	113	28	77
132 S	0.79	0.920	i	0.72	1.047•	b	1.35	1.107•	-109	133	49	105
133 V	1.14	1.012•	h	1.65	1.030•	H	0.53	1.147•	-141	138	-20	70
134 T	0.82	1.012•	i	1.20	0.860	h	0.99	1.175•	-161	78	-97	82
135 P	0.59	1.012•	B	0.62	0.865	b	1.56	1.070•	-117	37	26	88
136 N	0.73	1.012•	b	0.65	1.122•	b	1.51	0.812	-123	24	72	81
137 A	1.45	1.145•	H	0.97	1.260•	I	0.64	0.682	-85	37	-10	65
138 L	1.34	1.067•	H	1.22	1.172•	h	0.57	0.912	-73	93	12	100
139 V	1.14	0.980	h	1.65	0.932	H	0.53	0.975	-86	148	-55	100
140 T	0.82	1.032•	i	1.20	0.720	h	0.99	1.200•	-53	123	-147	77
141 P	0.59	1.058•	B	0.62	0.720	b	1.56	1.200•	23	62	-24	58
142 E	1.53	1.150•	H	0.26	0.807	B	0.82	0.970	53	35	36	-9
143 D	0.98	1.118•	i	0.80	0.942	i	1.43	1.122•	103	-19	-29	-50
144 T	0.82	1.077•	i	1.20	1.155•	h	0.99	0.897	137	-72	-42	-38
145 A	1.45	1.107•	H	0.97	1.160•	I	0.64	0.792	168	-28	-20	-25
146 D	0.98	1.052•	i	0.80	1.080•	i	1.43	1.010•	186	21	16	-35
147 V	1.14	1.020•	h	1.65	1.280•	H	0.53	0.787	164	73	-30	-35
148 L	1.34	0.958	H	1.22	1.187•	h	0.57	0.815	137	88	-16	-30
149 N	0.73	0.922	b	0.65	1.107•	b	1.51	0.935	94	109	17	-19
150 I	1.00	0.888	I	1.60	1.270•	H	0.54	0.880	69	112	14	-18
151 F	1.12	0.977	h	1.28	1.190•	h	0.64	0.905	36	111	22	-21
152 R	0.79	1.032•	i	0.90	1.072•	i	1.05	1.130•	16	49	51	28
153 C	0.77	1.155•	i	1.30	0.912	h	1.29	1.072•	17	-1	69	28
154 P	1.12	1.217•	h	1.28	0.830	h	0.64	0.910	61	-94	52	59
155 G	0.53	1.197•	B	0.81	0.575	I	1.54	0.955	39	-142	42	79
156 E	1.53	1.350•	H	0.26	0.785	B	0.82	0.702	123	-175	-17	41
157 A	1.45	1.350•	H	0.97	1.120•	I	0.64	0.632	148	-173	-85	15
158 E	1.53	1.207•	H	0.26	1.120•	B	0.82	0.632	181	-170	-107	-29
159 V	1.14	1.152•	h	1.65	1.120•	H	0.53	0.632	215	-147	-125	-85
160 I	1.00	1.128•	I	1.60	0.862	H	0.54	0.890	267	-133	-101	-83
161 A	1.45	1.168•	H	0.97	0.880	I	0.64	0.917	248	-103	-145	-65
162 E	1.53	1.025•	H	0.26	1.037•	B	0.82	0.892	186	-60	-212	-69
163 P	0.59	0.960	B	0.62	1.150•	b	1.56	0.922	181	-78	-39	-42
164 M	1.20	1.052•	h	1.67	1.150•	H	0.65	0.922	126	-57	52	-16
165 I	1.00	0.940	I	1.60	1.145•	H	0.54	0.892	34	12	-91	-43
166 H	1.24	0.963	h	0.71	1.157•	b	0.94	0.890	-50	75	-198	-19
167 P	0.59	0.888	B	0.62	1.182•	b	1.56	1.040•	-62	107	-69	-7
168 V	1.14	0.878	h	1.65	1.440•	H	0.53	0.782	-76	168	5	-25
169 V	1.14	0.820	h	1.65	1.207•	H	0.53	0.987	-101	168	-65	-45
170 G	0.53	0.762	B	0.81	0.997	i	1.54	1.240•	-201	128	47	34
171 V	1.14	0.772	h	1.65	0.975	H	0.53	1.192•	-236	108	110	110
172 S	0.79	0.823	i	0.72	0.742	b	1.35	1.397•	-269	68	111	165
173 G	0.53	0.793	B	0.81	0.717	i	1.54	1.450•	-306	-17	147	149
174 S	0.79	0.895	i	0.72	0.757	b	1.35	1.225•	-279	-42	71	165
175 S	0.79	0.950	i	0.72	0.900	b	1.35	1.170•	-259	-27	-89	135
176 P	0.59	1.018•	B	0.62	1.132•	b	1.56	0.965	-152	-43	16	83
177 A	1.45	1.107•	H	0.97	1.297•	I	0.64	0.735	-80	7	30	20
178 Y	0.61	1.032•	b	1.29	1.472•	h	1.13	0.737	5	55	-76	-51
179 V	1.14	1.185•	h	1.65	1.470•	H	0.53	0.615	79	128	-145	-85
180 F	1.12	1.237•	h	1.28	1.457•	h	0.64	0.617	131	166	-173	-116
181 M	1.20	1.250•	h	1.67	1.202•	H	0.65	0.662	178	138	-198	-136
182 F	1.12	1.292•	h	1.28	1.027•	h	0.64	0.660	236	86	-208	-166

residue	Pa	Pa avg	Ha	Pb	Pb avg	Hb	Pt	Pt avg	Ph	Ps	Ptu	Pco
183 I	1.00	1.268*	I	1.69	1.125*	H	0.54	0.662	266	2	-231	-138
184 E	1.53	1.363*	H	0.26	0.967	B	0.82	0.687	283	-100	-222	-134
185 A	1.45	1.330*	H	0.97	1.102*	I	0.64	0.840	303	-188	-185	-120
186 M	1.20	1.278*	h	1.67	1.102*	H	0.65	0.940	316	-242	-158	-91
187 A	1.45	1.302*	H	0.97	0.927	I	0.64	0.837	338	-238	-185	-110
188 D	0.98	1.148*	I	0.80	1.097*	I	1.43	0.810	328	-169	-174	-125
189 A	1.45	1.073*	H	0.97	1.202*	I	0.64	0.595	300	-118	-180	-120
190 A	1.45	1.032*	H	0.97	1.162*	I	0.64	0.820	248	-53	-150	-105
191 V	1.14	0.888	h	1.65	1.122*	H	0.53	1.045*	179	-2	-100	-55
192 L	1.34	0.830	H	1.22	1.127*	h	0.57	1.075*	97	-12	-21	10
193 G	0.53	0.848	B	0.81	0.977	I	1.54	1.322*	-11	-27	67	59
194 C	0.53	0.955	B	0.81	1.000*	I	1.54	1.200*	-48	-42	12	69
195 M	1.20	1.188*	h	1.67	1.040*	H	0.65	0.975	-32	-12	-148	74
196 P	0.59	1.010*	B	0.62	0.930	h	1.56	1.055*	58	-33	-69	48
197 B	0.79	1.090*	I	0.90	1.017*	I	1.05	0.825	101	-6	-24	-17
198 A	1.45	1.145*	H	0.97	1.115*	I	0.64	0.845	150	-3	-80	-65
199 Q	1.17	1.145*	h	1.33	1.057*	h	0.97	0.922	160	2	-56	-55
200 A	1.45	1.192*	H	0.97	1.070*	I	0.64	0.840	145	7	-35	0
201 Y	0.61	1.145*	b	1.29	1.070*	h	1.13	0.840	140	5	-31	-14
202 K	1.07	1.285*	I	0.74	0.990	b	0.95	0.717	153	-3	-60	-28
203 F	1.12	1.287*	h	1.28	1.113*	h	0.64	0.722	166	11	-78	-56
204 A	1.45	1.310*	H	0.97	1.035*	I	0.64	0.722	200	17	-115	-55
205 A	1.45	1.157*	H	0.97	1.205*	I	0.64	0.695	235	27	-160	-115
206 Q	1.17	1.047*	h	1.33	1.380*	h	0.97	0.697	250	22	-186	-130
207 A	1.45	1.093*	H	0.97	1.278*	I	0.64	0.840	225	-8	-145	-65
208 V	1.14	1.030*	h	1.65	1.212*	H	0.53	1.017*	204	-47	-140	-80
209 M	1.20	1.040*	h	1.67	1.042*	H	0.65	1.045*	183	-87	-78	-36
210 G	0.53	1.030*	B	0.81	0.810	I	1.54	1.120*	154	-132	-38	-11
211 S	0.79	1.165*	I	0.72	1.025*	b	1.35	0.897	156	-92	-49	35
212 A	1.45	1.288*	H	0.97	1.257*	I	0.64	0.892	140	-48	-85	12
213 K	1.07	1.183*	I	0.74	1.320*	b	0.95	0.875	148	-23	-110	12
214 M	1.20	1.093*	h	1.67	1.200*	H	0.65	0.842	136	48	-148	-51
215 V	1.14	1.148*	h	1.65	1.082*	H	0.53	0.727	149	88	-195	-75
216 L	1.34	1.165*	H	1.22	0.872	h	0.57	0.980	122	83	-168	-35
217 E	1.53	1.040*	H	0.26	0.632	B	0.82	1.042*	93	45	-97	-54
218 T	0.82	0.873	I	1.20	0.745	h	0.99	1.072*	110	-47	-42	-28
219 G	0.53	0.978	B	0.81	0.600	I	1.54	1.215*	92	-142	22	29
220 E	1.53	1.113*	H	0.26	0.600	B	0.82	1.215*	96	-195	-37	31
221 H	1.24	1.037*	h	0.71	0.777	b	0.94	1.170*	73	-195	-158	46
222 P	0.59	0.993	B	0.62	0.998	b	1.56	1.077*	106	-203	11	63
223 G	0.53	1.095*	B	0.81	0.635	I	1.54	0.925	110	-202	102	39
224 A	1.45	1.197*	H	0.97	0.933	I	0.64	0.897	158	-163	15	-10
225 L	1.34	1.083*	H	1.22	1.107*	h	0.57	0.980	122	-57	-26	-48
226 K	1.07	0.992	I	0.74	1.215*	b	0.95	0.895	78	62	0	-68
227 D	0.98	0.912	I	0.80	1.365*	I	1.43	0.975	38	171	51	-45
228 M	1.20	0.837	h	1.67	1.335*	H	0.65	0.955	-12	218	37	-36
229 V	1.14	0.725	h	1.65	1.072*	H	0.53	1.182*	-81	243	55	-30
230 C	0.77	0.672	I	1.30	0.862	h	1.29	1.435*	-138	204	49	28
231 S	0.79	0.680	I	0.72	0.740	b	1.35	1.497*	-194	33	11	125
232 P	0.59	0.715	B	0.62	0.860	b	1.56	1.407*	-187	-53	171	153
233 G	0.53	0.872	B	0.81	1.005*	I	1.54	1.265*	-186	-122	230	179
234 G	0.53	1.025*	B	0.81	1.202*	I	1.54	1.015*	-181	-127	110	164
235 T	0.82	1.127*	I	1.20	1.065*	h	0.99	0.835	-111	-37	-7	117
236 T	0.82	1.122*	I	1.20	1.007*	h	0.99	0.747	-41	63	-93	52
237 I	1.00	1.175*	I	1.60	1.120*	H	0.54	0.632	11	122	-131	-18
238 E	1.53	1.232*	H	0.26	0.945	B	0.82	0.760	88	135	-182	-109
239 A	1.45	1.232*	H	0.97	1.282*	I	0.64	0.687	138	117	-170	-100
240 V	1.14	1.245*	h	1.65	1.355*	H	0.53	0.570	177	98	-175	-55
241 R	0.79	1.233*	I	0.98	1.087*	I	1.05	0.742	197	49	-169	-62
242 V	1.14	1.190*	h	1.65	0.847	H	0.53	0.685	247	9	-170	-60
243 L	1.34	1.187*	H	1.22	0.620	h	0.57	0.790	277	-32	-131	-40
244 E	1.53	1.095*	H	0.26	0.517	B	0.82	1.032*	258	-100	-72	-44
245 E	1.53	1.082*	H	0.26	0.772	B	0.82	0.987	219	-158	-17	-67
246 K	1.07	1.068*	I	0.74	0.932	b	0.95	1.045*	194	-208	5	-38
247 G	0.53	1.080*	B	0.81	0.990	I	1.54	0.967	160	-252	7	-16
248 F	1.12	1.158*	h	1.28	1.030*	h	0.64	0.742	232	-189	-48	-36
249 R	0.79	1.227*	I	0.90	1.122*	I	1.05	0.715	274	-131	-149	-62
250 A	1.45	1.337*	H	0.97	1.297*	I	0.64	0.587	315	-88	-190	-55
251 A	1.45	1.295*	H	0.97	1.120*	I	0.64	0.632	325	-40	-170	-75
252 V	1.14	1.190*	h	1.65	1.120*	H	0.53	0.632	334	-52	-165	-80
253 I	1.00	1.178*	I	1.60	1.125*	H	0.54	0.662	296	-68	-201	-138
254 E	1.53	1.140*	H	0.26	1.025*	B	0.82	0.775	289	-90	-192	-179
255 A	1.45	1.085*	H	0.97	1.145*	I	0.64	0.807	288	-163	-132	-178
256 M	1.20	1.098*	h	1.67	1.227*	H	0.65	0.970	286	-167	-123	-136
257 T	0.82	1.077*	I	1.20	1.227*	h	0.99	0.970	287	-172	-132	-116
258 K	1.07	1.072*	I	0.74	0.992	b	0.95	0.927	291	-158	-120	-113
259 C	0.77	1.148*	I	1.30	0.992	h	1.29	0.927	307	-171	-81	-102
260 M	1.20	1.198*	h	1.67	0.847	H	0.65	0.942	323	-167	-113	-86
261 E	1.53	1.222*	H	0.26	0.495	B	0.82	0.985	318	-170	-147	-152
262 K	1.07	1.098*	I	0.74	0.615	b	0.95	1.017*	331	-178	-155	-163
263 S	0.79	1.098*	I	0.72	0.735	b	1.35	0.922	284	-212	-114	-100
264 E	1.53	1.098*	H	0.26	0.735	B	0.82	0.922	249	-230	-87	-127
265 K	1.07	1.012*	I	0.74	0.855	b	0.95	0.955	219	-213	-50	-88
266 L	1.34	0.997	H	1.22	0.850	h	0.57	1.055*	180	-192	-31	-36
267 S	0.79	0.883	I	0.72	0.727	b	1.35	1.217*	134	-187	26	17
268 K	1.07	0.930	I	0.74	0.730	b	0.95	1.150*	94	-123	20	2
269 S	0.79	0.790	I	0.72	0.720	b	1.35	1.350*	46	-62	31	36

v) Numerical data used in the prediction of the secondary structure of PSCR from human (Dougherty *et al.*, 1992)

residue	Pa	Pa avg	Ra	Pb	Pb avg	Rb	Pt	Pt avg	Pb	Pa	Ptu	Pco
1 M	1.20	0.963	h	1.67	1.212*	H	0.68	1.017*	-7	98	-23	-11
2 S	0.79	0.852	i	0.72	1.118*	b	1.38	1.015*	-14	118	1	25
3 V	1.14	0.962	h	1.65	1.335*	H	0.53	0.812	-31	133	-20	15
4 G	0.53	0.860	B	0.81	1.125*	i	1.54	1.065*	-56	93	7	29
5 F	1.12	0.967	h	1.28	1.165*	h	0.64	0.840	-39	91	7	9
6 I	1.00	1.003*	I	1.60	1.047*	H	0.54	1.065*	-29	42	-41	27
7 G	0.53	1.078*	B	0.81	0.955	i	1.54	1.172*	-36	-17	-23	39
8 A	1.45	1.177*	H	0.97	1.057*	I	0.64	0.930	20	-43	-8	50
9 G	0.53	1.177*	B	0.81	1.057*	i	1.54	0.930	42	-77	-43	34
10 Q	1.17	1.312*	h	1.23	1.175*	h	0.97	0.705	105	-33	-106	-5
11 L	1.34	1.358*	H	1.22	1.110*	h	0.57	0.622	137	8	-166	-20
12 A	1.45	1.313*	H	0.97	1.110*	I	0.64	0.622	180	32	-155	-70
13 F	1.12	1.160*	h	1.28	1.110*	h	0.64	0.622	222	96	-143	-101
14 A	1.45	1.160*	H	0.97	0.975	I	0.64	0.700	250	107	-105	-55
15 L	1.34	1.055*	H	1.22	0.935	h	0.57	0.925	232	33	-56	25
16 A	1.45	1.073*	H	0.97	0.950	I	0.64	0.942	190	7	25	37
17 K	1.07	1.073*	I	0.74	1.007*	b	0.95	1.030*	146	-63	40	52
18 G	0.53	0.983	B	0.81	1.065*	i	1.54	0.952	104	-82	25	39
19 P	1.12	1.085*	h	1.28	1.105*	h	0.64	0.727	111	-69	-33	-1
20 T	0.82	1.122*	i	1.20	0.987	h	0.99	0.952	124	-47	-112	-18
21 A	1.45	1.227*	H	0.97	1.100*	I	0.64	0.837	155	-38	-130	-59
22 A	1.45	1.227*	H	0.97	1.162*	I	0.64	0.820	200	-43	-140	-70
23 G	0.53	1.192*	B	0.81	1.162*	I	1.54	0.820	217	-82	-143	-76
24 V	1.14	1.282*	h	1.65	1.202*	H	0.53	0.595	289	-72	-205	-90
25 L	1.34	1.258*	H	1.22	0.967	h	0.57	0.697	322	-57	-221	-65
26 A	1.45	1.235*	H	0.97	0.847	I	0.64	0.792	340	-83	-175	-95
27 A	1.45	1.235*	H	0.97	1.005*	I	0.64	0.767	288	-33	-180	-125
28 H	1.24	1.125*	h	0.71	1.180*	b	0.94	0.770	222	5	-168	-137
29 K	1.07	1.050*	I	0.74	1.245*	b	0.95	0.695	138	2	-125	-98
30 I	1.00	0.970	I	1.60	1.240*	H	0.54	0.795	91	22	-91	-68
31 M	1.20	0.967	h	1.67	1.020*	H	0.65	0.997	43	8	-63	-46
32 A	1.45	0.967	H	0.97	0.757	I	0.64	1.225*	-10	-18	-5	-20
33 S	0.79	0.888	i	0.72	0.715	b	1.35	1.422*	-54	-57	-59	30
34 S	0.79	0.980	i	0.72	0.952	b	1.35	1.247*	-89	-107	-144	55
35 F	0.59	1.090*	B	0.62	0.972	b	1.56	1.267*	-17	-123	1	48
36 D	0.98	1.128*	i	0.80	1.122*	i	1.43	1.020*	20	-134	66	30
37 M	1.20	1.155*	h	1.67	1.165*	H	0.65	0.822	81	-77	-63	-51
38 D	0.98	1.087*	i	0.80	1.047*	i	1.43	0.907	115	-24	-109	-100
39 L	1.34	1.165*	H	1.22	1.260*	h	0.57	0.682	152	48	-153	-85
40 A	1.45	1.165*	H	0.97	1.135*	I	0.64	0.877	210	37	-165	-108
41 T	0.82	1.055*	i	1.20	1.135*	h	0.99	0.877	237	18	-192	-113
42 V	1.14	1.097*	h	1.65	1.140*	H	0.53	0.772	239	-7	-215	-130
43 S	0.79	1.107*	i	0.72	0.952	b	1.35	0.902	249	-32	-194	-110
44 A	1.45	1.063*	H	0.97	0.957	I	0.64	0.802	258	-48	-175	-115
45 L	1.34	1.012*	H	1.22	1.132*	h	0.57	0.805	217	-52	-126	-85
46 E	0.79	0.967	i	0.98	1.030*	I	1.85	1.047*	156	4	-34	-65
47 K	1.07	1.058*	I	0.74	1.217*	b	0.95	0.917	131	-13	35	-38
48 M	1.20	1.017*	h	1.67	1.217*	H	0.65	0.917	88	33	47	-1
49 G	0.53	0.915	B	0.81	1.105*	I	1.54	0.897	34	28	22	-6
50 V	1.14	1.033*	h	1.65	1.202*	H	0.53	0.760	27	73	-45	-43
51 K	1.07	0.965	I	0.74	0.945	b	0.95	1.017*	3	42	-72	-43
52 L	1.34	0.965	H	1.22	0.937	h	0.57	1.015*	-48	33	-131	10
53 T	0.82	0.997	i	1.20	0.795	h	0.99	1.250*	-71	-17	-142	72
54 P	0.59	0.997	B	0.62	0.680	b	1.56	1.240*	-4	-93	21	133
55 H	1.24	1.088*	h	0.71	0.590	b	0.94	1.055*	2	-170	112	141
56 N	0.73	1.077*	b	0.65	0.712	b	1.51	1.067*	4	-156	32	78
57 K	1.07	1.162*	I	0.74	0.962	b	0.95	0.822	33	-88	23	32
58 E	1.53	1.115*	H	0.26	1.085*	B	0.82	0.827	68	-5	3	11
59 T	0.82	1.023*	i	1.20	1.197*	h	0.99	0.857	92	38	43	27
60 V	1.14	1.077*	h	1.65	1.077*	H	0.53	0.947	92	48	-5	-5
61 Q	1.17	1.110*	h	1.23	0.865	h	0.97	1.172*	88	47	-26	-40
62 H	1.24	1.102*	h	0.71	0.970	b	0.94	1.062*	65	50	-18	-9
63 S	0.79	1.118*	i	0.72	1.097*	b	1.35	0.970	61	13	-24	10
64 D	0.98	1.228*	i	0.80	1.237*	I	1.43	0.792	108	16	-14	-25
65 V	1.14	1.255*	h	1.65	1.342*	H	0.53	0.577	159	33	-60	-40
66 L	1.34	1.243*	H	1.22	1.172*	h	0.57	0.605	170	88	-81	-65
67 P	1.12	1.118*	h	1.28	1.280*	h	0.64	0.595	176	91	-83	-91
68 L	1.34	1.138*	H	1.22	1.145*	h	0.57	0.672	180	53	-81	-65
69 A	1.45	1.082*	H	0.97	0.995	I	0.64	0.920	165	22	-10	-35
70 V	1.14	1.007*	h	1.65	0.830	H	0.53	0.995	107	-2	-45	-3
71 K	1.07	0.915	I	0.74	0.917	b	0.95	0.997	38	-33	-90	37
72 P	0.59	0.923	B	0.62	1.132*	b	1.56	0.895	18	-33	41	83
73 H	1.24	0.992	h	0.71	1.132*	b	0.94	0.895	-43	-5	97	46
74 I	1.00	1.008*	I	1.60	1.275*	H	0.54	0.820	-84	42	-96	-28
75 I	1.00	1.005*	I	1.60	1.275*	H	0.54	0.820	-104	92	-246	-58
76 P	0.59	1.093*	B	0.62	1.180*	b	1.56	0.827	42	147	-94	-42
77 P	1.12	1.162*	h	1.28	1.225*	h	0.64	0.795	-16	146	-8	-31
78 I	1.00	1.063*	I	1.60	0.970	H	0.54	0.840	16	112	-111	-38
79 L	1.34	1.138*	H	1.22	0.970	h	0.57	0.840	47	63	-106	-15
80 D	0.98	1.078*	I	0.80	0.867	i	1.43	1.062*	50	-19	6	30
81 E	1.53	1.082*	H	0.26	0.910	B	0.82	0.885	66	-55	18	36
82 I	1.00	1.062*	I	1.60	1.045*	H	0.54	1.037*	89	-108	-6	27
83 G	0.53	1.078*	B	0.81	1.045*	I	1.54	1.037*	92	-147	-8	44
84 A	1.45	1.122*	H	0.97	0.907	I	0.64	0.857	143	-168	-10	40
85 D	0.98	1.087*	i	0.80	0.865	i	1.43	1.055*	208	-144	-54	5
86 I	1.00	1.090*	I	1.60	0.890	H	0.54	0.960	221	-93	-96	-23
87 E	1.53	1.113*	H	0.26	0.667	B	0.82	1.060*	218	-79	-62	-69
88 D	0.98	1.048*	I	0.80	1.002*	i	1.43	0.990	193	-54	31	-70
89 H	0.79	1.017*	i	0.90	1.215*	i	1.85	0.765	144	-6	31	-77

residue		Pa	Pa avg	Ba	Pb	Pb avg	Bb	Pt	Pt avg	Ph	Pa	Ptu	Pco
90	H	1.24	1.013*	h	0.71	1.402*	b	0.94	0.635	105	45	42	-49
91	I	1.00	1.048*	i	1.60	1.405*	H	0.54	0.737	74	77	14	-18
92	V	1.14	1.133*	h	1.65	1.330*	H	0.53	0.925	74	128	-25	-10
93	V	1.14	1.022*	h	1.65	1.160*	H	0.53	0.952	64	133	-15	-15
94	S	0.79	1.022*	I	0.72	0.990	b	1.35	0.980	21	93	11	-15
95	C	0.77	1.027*	I	1.30	1.012*	h	1.29	1.027*	-8	59	-21	-52
96	A	1.45	1.063*	H	0.97	1.100*	I	0.64	0.837	-15	32	-15	-55
97	A	1.45	0.955	H	0.97	1.157*	I	0.64	0.925	5	-28	-20	-30
98	G	0.53	0.845	H	0.81	1.315*	I	1.54	0.980	4	-47	-10	-16
99	V	1.14	0.923	h	1.65	1.292*	H	0.53	0.852	49	23	-65	-25
100	T	0.82	0.988	I	1.20	1.060*	h	0.99	1.057*	64	73	-107	-23
101	I	1.00	1.030*	I	1.60	1.160*	H	0.54	0.945	66	72	-121	-23
102	S	0.79	1.042*	I	0.72	0.825	b	1.35	1.015*	71	83	-114	-35
103	S	0.79	1.133*	I	0.72	0.830	b	1.35	0.915	111	33	-84	-35
104	I	1.00	1.133*	I	1.60	0.835	H	0.54	0.815	116	-38	-76	-48
105	E	1.53	1.208*	H	0.26	0.740	B	0.82	0.822	88	-58	-42	-37
106	K	1.07	1.140*	I	0.74	0.855	b	0.95	0.955	64	-78	30	-31
107	K	1.07	1.093*	I	0.74	0.912	b	0.95	0.877	41	-83	50	12
108	L	1.34	1.013*	H	1.22	1.047*	h	0.57	0.800	25	-17	24	20
109	S	0.79	1.032*	I	0.72	0.967	b	1.35	0.920	-1	3	31	35
110	A	1.45	0.998	H	0.97	0.942	I	0.64	0.972	-7	-13	20	5
111	P	1.12	0.880	h	1.28	0.943	h	0.64	0.972	-29	-29	-28	-6
112	R	0.79	0.882	H	0.90	0.777	I	1.05	1.202*	-89	-71	-69	8
113	P	0.59	0.927	B	0.62	0.777	b	1.56	1.202*	-57	-68	26	53
114	A	1.45	0.960	H	0.97	1.035*	I	0.64	0.945	-80	-73	5	75
115	P	0.59	0.847	B	0.62	1.192*	b	1.56	0.920	-12	-38	96	63
116	R	0.79	0.948	H	0.90	1.262*	I	1.05	0.792	-14	34	121	18
117	V	1.14	0.953	h	1.65	1.362*	H	0.53	0.882	-11	138	25	-25
118	I	1.00	0.885	I	1.60	1.387*	H	0.54	0.882	-34	202	49	-28
119	R	0.79	0.855	I	0.90	1.207*	I	1.05	0.995	-89	259	71	-27
120	C	0.77	0.822	I	1.30	1.205*	h	1.29	1.110*	-133	219	127	-17
121	M	1.28	0.883	h	1.67	1.180*	H	0.65	1.038*	-182	113	137	44
122	T	0.82	0.873	I	1.20	0.917	h	0.99	1.262*	-230	-7	141	87
123	N	0.73	0.927	I	0.65	1.030*	H	1.51	1.147*	-276	-26	57	106
124	T	0.82	0.937	I	1.20	1.280*	h	0.99	0.902	-286	-7	-77	93
125	P	0.59	1.055*	B	0.62	1.392*	b	1.56	0.787	-162	72	-4	68
126	V	1.14	1.045*	h	1.65	1.462*	H	0.53	0.660	-81	193	10	55
127	V	1.14	1.097*	h	1.65	1.115*	H	0.53	0.732	-6	223	-95	20
128	V	1.14	1.043*	h	1.65	0.985	H	0.53	0.985	69	206	-90	10
129	S	0.79	1.043*	I	0.90	0.735	I	1.05	1.012*	46	179	-39	-7
130	E	1.53	1.013*	H	0.26	0.810	B	0.82	0.997	33	118	-22	-39
131	G	0.53	1.000*	B	0.81	1.157*	I	1.54	0.925	-3	23	30	-6
132	A	1.45	1.048*	H	0.97	1.277*	I	0.64	0.822	23	12	10	45
133	T	0.82	0.895	I	1.20	1.277*	h	0.99	0.822	32	43	-17	42
134	V	1.14	0.895	h	1.65	1.277*	H	0.53	0.822	27	78	-30	45
135	Y	0.61	0.912	b	1.29	1.067*	h	1.13	1.075*	20	105	17	24
136	A	1.45	1.052*	H	0.97	1.045*	I	0.64	1.040*	38	77	45	55
137	T	0.82	1.085*	I	1.20	0.980	h	0.99	1.115*	24	28	56	62
138	G	0.53	1.058*	B	0.81	0.922	I	1.54	1.027*	-11	-27	42	99
139	T	0.82	1.225*	I	1.20	1.027*	h	0.99	0.885	-16	-37	18	92
140	H	1.24	1.252*	h	0.71	1.140*	b	0.94	0.770	-3	16	-58	91
141	A	1.45	1.133*	H	0.97	1.027*	I	0.64	0.740	0	57	-80	40
142	Q	1.17	1.023*	h	1.33	0.985	h	0.97	0.937	70	72	-76	10
143	V	1.14	1.052*	h	1.65	0.880	H	0.53	1.080*	89	38	-45	25
144	E	1.53	1.062*	H	0.26	0.692	B	0.82	1.210*	93	-28	13	26
145	D	0.98	1.062*	I	0.80	0.932	I	1.43	1.147*	93	-79	61	40
146	G	0.53	1.083*	B	0.81	1.150*	I	1.54	0.962	100	-122	57	69
147	S	0.79	1.228*	I	0.90	1.012*	I	1.05	0.772	154	-46	6	33
148	L	1.34	1.320*	H	1.22	1.095*	h	0.57	0.752	220	-12	-81	5
149	N	1.20	1.228*	h	1.67	1.095*	H	0.65	0.752	233	28	-73	-21
150	E	1.53	1.165*	H	0.26	0.982	B	0.82	0.732	214	40	-87	-64
151	Q	1.17	1.100*	h	1.23	1.097*	h	0.97	0.865	161	32	-71	-90
152	L	1.34	0.993	H	1.22	1.090*	h	0.57	0.870	90	43	-46	-40
153	L	1.34	0.957	H	1.22	1.197*	h	0.57	0.860	30	48	27	-5
154	S	0.79	0.862	I	0.72	1.095*	b	1.35	1.102*	32	23	98	85
155	T	0.82	0.867	I	1.20	1.235*	h	0.99	0.925	32	48	128	117
156	V	1.14	0.985	h	1.65	1.260*	H	0.53	1.000*	19	28	110	130
157	G	0.53	0.985	B	0.81	1.147*	I	1.54	1.115*	-31	3	102	129
158	P	1.12	1.152*	h	1.28	1.010*	h	0.64	0.935	-44	46	55	94
159	C	0.77	1.220*	I	1.30	1.102*	h	1.29	0.907	-13	59	-41	48
160	T	0.82	1.255*	I	1.20	0.842	h	0.99	0.790	34	28	-72	82
161	H	1.53	1.342*	H	0.28	0.607	B	0.82	0.747	118	-15	-92	-14
162	V	1.14	1.253*	h	1.65	0.742	H	0.53	0.900	207	-72	-105	-55
163	H	1.53	1.227*	H	0.26	0.635	B	0.82	0.918	246	-105	-93	-84
164	H	1.53	1.213*	H	0.26	0.970	B	0.82	0.840	262	-120	-67	-119
165	D	0.98	1.148*	I	0.80	1.108*	I	1.43	0.992	279	-149	-29	-75
166	L	1.34	1.122*	H	1.22	1.147*	h	0.57	0.795	293	-82	-61	-60
167	I	1.00	0.987	I	1.60	1.255*	H	0.54	0.785	287	-33	-71	-83
168	D	0.98	1.043*	I	0.80	1.155*	I	1.43	0.897	356	26	-99	-95
169	A	1.45	1.012*	H	0.97	1.157*	I	0.64	0.925	190	62	-92	-75
170	V	1.14	0.858	h	1.65	1.220*	H	0.53	0.907	94	103	-65	-45
171	T	0.82	0.860	I	1.20	0.987	h	0.99	1.112*	-8	113	-27	-13
172	G	0.53	0.752	B	0.81	0.890	I	1.54	1.250*	-148	78	72	84
173	L	1.34	0.762	H	1.22	0.867	h	0.57	1.202*	-198	63	139	145
174	S	0.79	0.780	I	0.72	0.765	b	1.35	1.445*	-244	38	178	215
175	G	0.53	0.750	B	0.81	0.740	I	1.54	1.497*	-283	-22	237	214
176	S	0.79	0.963	I	0.72	0.780	b	1.35	1.272*	-264	8	156	250
177	G	0.53	0.958	B	0.81	0.922	I	1.54	1.217*	-271	3	2	199
178	P	0.59	1.007*	B	0.62	0.962	b	1.56	0.992	-147	32	76	168
179	A	1.45	1.150*	H	0.97	1.127*	I	0.64	0.782	-80	72	45	85
180	Y	0.61	1.132*	b	1.29	1.185*	h	1.13	0.858	-5	90	-66	-16
181	A	1.45	1.152*	H	0.97	1.105*	I	0.64	0.727	55	132	-102	-75
182	P	1.12	1.192*	b	1.28	1.167*	h	0.64	0.710	101	156	-133	-116

residue	Pa	Pa avg	Ha	Pb	Pb avg	Hb	Pt	Pt avg	Ph	Ps	Ptu	Pco
276 Q	1.17	1.098*	h	1.23	1.055*	h	0.97	1.102*	143	12	9	-5
277 V	1.14	1.070*	h	1.65	0.990	H	0.53	1.020*	147	3	-65	20
278 S	0.79	1.058*	I	0.72	0.820	b	1.35	1.047*	159	-17	-159	50
279 P	0.59	1.105*	B	0.62	1.040*	b	1.56	0.845	236	-23	-14	58
280 A	1.45	1.143*	H	0.97	1.070*	I	0.64	0.692	260	-33	55	65
281 A	1.45	1.068*	H	0.97	1.012*	I	0.64	0.770	285	-33	-5	15
282 I	1.00	1.050*	I	1.60	1.070*	H	0.54	0.857	251	-18	-41	-31
283 K	1.07	1.047*	I	0.74	1.070*	b	0.95	0.857	213	37	-82	-101
284 K	1.07	1.047*	I	0.74	1.190*	b	0.95	0.762	173	82	-105	-118
285 T	0.82	1.058*	I	1.20	1.205*	h	0.99	0.882	157	43	-122	-123
286 I	1.00	1.100*	I	1.60	1.090*	H	0.54	0.872	141	32	-166	-133
287 L	1.34	1.157*	H	1.22	1.102*	h	0.57	0.870	112	43	-161	-145
288 D	0.98	1.097*	I	0.80	0.982	I	1.43	0.965	80	46	-54	-128
289 K	1.07	1.065*	I	0.74	1.087*	b	0.95	0.750	46	67	-40	-88
290 V	1.14	0.985	h	1.65	1.102*	H	0.53	0.870	2	78	-10	-28
291 K	1.07	1.037*	I	0.74	0.870	b	0.95	1.075*	-22	77	25	22
292 L	1.34	0.947	H	1.22	0.840	h	0.57	1.227*	-48	53	44	50
293 D	0.98	0.860	I	0.80	0.777	I	1.43	1.245*	-70	26	16	55
294 S	0.79	0.938	I	0.72	0.780	b	1.35	1.272*	-111	3	-104	65
295 P	0.59	1.030*	B	0.62	0.900	b	1.56	1.182*	-52	-28	16	38
296 A	1.45	1.063*	H	0.97	0.987	I	0.64	0.952	-65	-13	58	25
297 G	0.53	0.920	B	0.81	1.050*	I	1.54	0.935	-86	-17	-3	14
298 T	0.82	0.963	I	1.20	1.027*	h	0.99	0.887	-73	18	-27	7
299 A	1.45	0.915	H	0.97	0.882	I	0.64	1.030*	-65	42	-15	55
300 L	1.34	0.880	H	1.22	0.820	h	0.57	1.207*	-68	93	-36	85
301 S	0.79	0.793	I	0.72	0.717	b	1.35	1.450*	-94	78	-59	135
302 P	0.59	0.840	B	0.62	0.715	b	1.56	1.347*	-44	32	121	123
303 S	0.79	0.965	I	0.72	0.860	b	1.35	1.205*	-44	-12	201	165
304 G	0.53	1.057*	B	0.81	0.865	I	1.54	1.105*	-76	-57	135	94
305 H	1.24	1.067*	h	0.71	0.907	b	0.94	0.862	-88	-40	62	66
306 T	0.82	0.992	I	1.20	1.095*	h	0.99	0.770	-123	33	-12	19
307 K	1.07	0.987	I	0.74	0.950	b	0.95	0.912	-119	87	5	7
308 L	1.34	1.032*	H	1.22	0.990	h	0.57	0.937	-123	118	-36	30
309 L	1.34	1.050*	H	1.22	0.865	h	0.57	1.132*	-138	133	-106	45
310 P	0.59	0.925	B	0.62	0.865	b	1.56	1.132*	-59	97	36	33
311 R	0.79	1.068*	I	0.90	0.982	I	1.05	0.902	-33	64	76	18
312 S	0.79	1.025*	I	0.72	0.882	b	1.35	1.030*	-9	28	-4	15
313 L	1.34	1.072*	H	1.22	0.945	h	0.57	0.852	-13	28	-91	10
314 A	1.45	1.012*	H	0.97	0.842	I	0.64	1.095*	-35	2	-150	40
315 P	0.59	0.924	B	0.62	0.785	b	1.56	1.172*	11	-33	16	53
316 A	1.45	1.007*	H	0.97	0.830	I	0.64	1.140*	-10	-93	85	60
317 G	0.53	0.860	B	0.81	0.783	I	1.54	1.307*	-41	-112	22	11
318 K	1.07	1.025*	I	0.74	0.770	b	0.95	1.190*	-2	-63	20	7
319 D	0.98	0.980	I	0.80	0.800	I	1.43	1.430*	10	-34	21	20

residue	Pa	Pa avg	Ba	Pb	Pb avg	Bb	Pt	Pt avg	Ph	Ps	Ptu	Pco
183 T	0.82	1.230*	i	1.20	1.047*	h	0.99	0.907	137	98	-157	-113
184 A	1.45	1.335*	H	0.97	0.990	I	0.64	0.820	150	67	-155	-90
185 L	1.34	1.257*	H	1.22	1.052*	h	0.57	0.802	162	23	-156	-86
186 D	0.98	1.122*	I	0.80	0.990	i	1.43	0.820	178	1	-104	-89
187 A	1.45	1.047*	H	0.97	0.990	I	0.64	0.820	208	-43	-130	-68
188 L	1.34	0.995	H	1.22	0.950	h	0.57	1.045*	207	-82	-111	-25
189 A	1.45	0.950	H	0.97	0.847	I	0.64	1.287*	190	-83	-45	-30
190 D	0.98	0.908	I	0.80	1.017*	i	1.43	1.260*	143	-104	21	-19
191 G	0.53	0.833	B	0.81	1.002*	i	1.54	1.140*	79	-82	32	-1
192 G	0.53	0.968	B	0.81	1.217*	i	1.54	0.917	19	-42	37	19
193 V	1.14	0.978	h	1.65	1.217*	H	0.53	0.917	-11	63	-18	32
194 K	1.07	0.920	I	0.74	1.110*	b	0.95	0.927	-37	82	0	27
195 M	1.20	0.873	h	1.67	1.080*	H	0.65	1.080*	-69	103	17	19
196 G	0.53	0.897	B	0.81	0.887	i	1.54	1.180*	-116	78	-58	9
197 L	1.34	1.050*	H	1.22	0.910	b	0.57	1.057*	-103	53	-200	0
198 P	0.59	1.017*	B	0.62	0.910	b	1.56	1.057*	-22	2	-69	-12
199 R	0.79	1.050*	I	0.90	0.997	I	1.05	0.827	13	4	-19	-42
200 R	0.79	1.142*	I	0.90	1.185*	I	1.05	0.697	58	44	-84	-87
201 L	1.34	1.098*	H	1.22	1.185*	h	0.57	0.697	87	103	-91	-85
202 A	1.45	1.117*	H	0.97	1.185*	I	0.64	0.687	90	142	-55	-20
203 V	1.14	1.070*	h	1.65	1.142*	H	0.53	0.922	88	133	-60	-10
204 R	0.79	1.122*	I	0.90	0.975	I	1.05	0.950	41	94	-34	-12
205 L	1.34	1.213*	H	1.22	1.057*	h	0.57	0.930	32	48	-66	-25
206 G	0.53	1.213*	B	0.81	0.995	i	1.54	0.947	44	-7	-63	-38
207 A	1.45	1.213*	H	0.97	1.097*	I	0.64	0.705	133	-8	-95	-55
208 Q	1.17	1.212*	h	1.23	1.160*	h	0.97	0.687	183	-8	-181	-75
209 A	1.45	1.260*	H	0.97	1.055*	I	0.64	0.838	183	2	-188	-66
210 L	1.34	1.197*	H	1.22	1.055*	h	0.57	0.830	187	18	-191	-75
211 L	1.34	1.173*	H	1.22	0.992	h	0.57	0.847	227	-22	-141	-55
212 G	0.53	1.173*	B	0.81	0.872	I	1.54	0.942	252	-77	-83	-46
213 A	1.45	1.308*	H	0.97	1.087*	I	0.64	0.720	298	-103	-95	-48
214 A	1.45	1.273*	H	0.97	1.159*	I	0.64	0.702	295	-58	-125	-63
215 K	1.07	1.163*	I	0.74	1.212*	b	0.95	0.685	296	-23	-145	-83
216 M	1.20	1.240*	h	1.67	1.205*	H	0.65	0.682	256	38	-123	-56
217 L	1.34	1.235*	H	1.22	0.967	h	0.57	0.857	227	13	-141	-35
218 L	1.34	1.218*	H	1.22	0.727	h	0.57	0.920	162	33	-98	-19
219 H	1.24	1.093*	b	0.71	0.730	b	0.94	1.020*	145	-6	-38	6
220 S	0.79	0.978	I	0.72	0.730	b	1.35	1.020*	114	-77	1	65
221 E	1.53	1.038*	H	0.26	0.705	B	0.82	1.072*	113	-100	43	21
222 Q	1.17	1.007*	h	1.23	0.842	h	0.97	1.252*	58	-108	19	40
223 H	1.24	0.990	b	0.71	0.842	b	0.94	1.252*	-15	-105	-33	86
224 F	0.59	0.947	B	0.62	0.978	b	1.56	1.160*	-4	-98	111	113
225 G	0.53	0.970	B	0.81	1.008*	i	1.54	1.087*	-25	-82	167	94
226 Q	1.17	1.072*	h	1.23	0.997	h	0.97	0.980	0	-8	74	60
227 L	1.34	1.008*	H	1.22	0.852	h	0.57	1.115*	-48	63	69	82
228 K	1.07	0.917	I	0.74	0.960	b	0.95	1.105*	-97	117	135	107
229 D	0.98	0.837	I	0.80	0.955	I	1.43	1.205*	-147	161	196	148
230 N	0.73	0.762	b	0.65	0.935	b	1.51	1.185*	-191	194	167	151
231 V	1.14	0.728	h	1.65	0.927	H	0.53	1.197*	-206	193	160	150
232 S	0.79	0.780	I	0.72	0.717	b	1.35	1.450*	-219	148	126	210
233 S	0.79	0.765	I	0.72	0.748	b	1.35	1.497*	-229	33	66	250
234 P	0.59	0.820	B	0.62	0.802	b	1.56	1.320*	-147	-53	191	208
235 G	0.53	0.928	B	0.81	0.947	I	1.54	1.177*	-116	-97	207	169
236 G	0.53	1.082*	B	0.81	1.149*	I	1.54	0.927	-51	-67	85	104
237 A	1.45	1.217*	H	0.97	1.138*	I	0.64	0.777	20	2	-45	50
238 T	0.82	1.182*	I	1.20	1.138*	H	0.99	0.777	74	83	-117	2
239 I	1.00	1.235*	I	1.60	1.135*	H	0.54	0.672	111	117	-146	-38
240 H	1.24	1.292*	h	0.71	0.802	b	0.94	0.772	120	145	-183	-94
241 A	1.45	1.340*	H	0.97	1.137*	I	0.64	0.670	90	137	-180	-85
242 L	1.34	1.230*	H	1.22	1.200*	h	0.57	0.652	82	123	-166	-45
243 H	1.24	1.095*	h	0.71	0.988	b	0.94	0.710	70	90	-113	-14
244 V	1.14	0.977	h	1.65	0.962	H	0.53	0.817	57	63	-85	0
245 L	1.34	0.973	H	1.22	0.752	h	0.57	1.070*	37	43	-31	40
246 E	1.53	0.882	H	0.26	0.650	B	0.82	1.312*	13	-5	43	16
247 S	0.79	0.758	I	0.72	0.905	b	1.35	1.267*	-8	-72	101	50
248 C	0.53	0.850	B	0.81	0.950	I	1.54	1.192*	-33	-137	137	89
249 G	0.53	0.885	B	0.81	0.927	I	1.54	1.145*	-23	-137	107	74
250 F	1.12	1.063*	h	1.20	1.030*	h	0.64	0.982	19	-29	22	29
251 R	0.79	0.998	I	0.90	1.015*	I	1.05	0.885	34	14	-19	13
252 S	0.79	1.108*	I	0.72	1.190*	b	1.35	0.757	29	58	-39	30
253 L	1.34	1.167*	H	1.22	1.172*	h	0.57	0.797	27	73	-46	35
254 L	1.34	1.198*	H	1.22	1.118*	h	0.57	0.815	12	63	-36	35
255 I	1.00	1.217*	I	1.60	1.217*	H	0.54	0.805	9	42	-46	17
256 N	0.73	1.182*	b	0.65	0.882	b	1.51	0.875	27	-36	-68	-9
257 A	1.45	1.188*	H	0.97	0.962	I	0.64	0.857	95	-53	-50	-15
258 V	1.14	1.113*	h	1.65	0.900	H	0.53	0.835	139	-12	-65	-30
259 E	1.53	1.055*	H	0.26	0.812	B	0.82	1.025*	163	-10	-42	-39
260 A	1.45	0.937	H	0.97	1.147*	I	0.64	0.955	143	-9	0	-30
261 S	0.79	0.827	I	0.72	1.130*	b	1.35	1.057*	114	-7	41	5
262 C	0.77	0.950	I	1.30	1.250*	h	1.29	0.987	95	44	54	-22
263 I	1.00	1.045*	I	1.60	1.150*	H	0.54	0.907	104	17	22	2
264 R	0.79	1.073*	I	0.90	0.815	I	1.05	0.977	101	5	-34	-32
265 T	0.82	1.073*	I	1.20	0.895	h	0.99	0.857	124	-22	-17	-26
266 R	0.79	1.137*	I	0.90	0.992	I	1.05	0.852	99	-51	16	-7
267 E	1.53	1.247*	H	0.26	0.857	B	0.82	0.927	83	-68	-2	1
268 L	1.34	1.155*	H	1.22	1.210*	h	0.57	0.885	55	-62	39	45
269 Q	1.17	1.127*	h	1.23	1.147*	h	0.97	0.902	38	-58	74	85
270 S	0.79	1.187*	I	0.72	1.040*	b	1.35	1.017*	42	-67	81	130
271 M	1.20	1.250*	h	1.67	1.187*	H	0.65	0.922	71	-57	27	89
272 A	1.45	1.240*	H	0.97	0.815	I	0.64	0.955	115	-68	-15	35
273 D	0.98	1.130*	I	0.80	0.880	I	1.43	1.047*	170	-49	-4	5
274 Q	1.17	1.065*	h	1.23	1.092*	h	0.97	0.822	165	-38	14	30
275 H	1.53	1.112*	H	0.26	0.965	B	0.82	0.917	153	-20	-2	1