Biochemistry students' difficulties with the symbolic and visual language used in Molecular Biology

Abindra Supersad Gupthar

A dissertation submitted to the University of KwaZulu-Natal, Discipline of Biochemistry, Pietermaritzburg Campus , in part fulfilment of the academic requirements for the degree of Doctor of Philosophy.

Pietermaritzburg, 2007

ABSTRACT

This study reports on recurring difficulties experienced by undergraduate students with respect to understanding and interpretation of certain symbolism, nomenclature, terminology, shorthand notation, models and other visual representations employed in the field of Molecular Biology to communicate information. Based on teaching experience and guidelines set out by a four-level methodological framework, data on various topic-related difficulties was obtained by inductive analyses of students' written responses to specifically designed, free-response and focused probes. In addition, interviews, think-aloud exercises and student-generated diagrams were also used to collect information. Both unanticipated and recurring difficulties were compared with scientifically correct propositional knowledge, categorized and subsequently classified. Students were adept at providing the meaning of the symbol "Δ" in various scientific contexts; however, some failed to recognize its use to depict the deletion of a leucine biosynthesis gene in the form, Δ *leu*. "Hazard to leucine", "change to leucine" and "abbreviation for isoleucine" were some of the erroneous interpretations of this polysemic symbol. Investigations on these definitions suggest a constructivist approach to knowledge construction and the inappropriate transfer of knowledge from prior mental schemata. The symbol, "**::**", was poorly differentiated by students in its use to indicate gene integration or transposition and in tandem gene fusion. Idiosyncratic perceptions emerged suggesting that it is, for example, a proteinaceous component linking genes in a chromosome or the centromere itself associated with the mitotic spindle or "electrons" between genes in the same way that it is symbolically shown in Lewis dot diagrams which illustrate covalent bonding between atoms. In an oligonucleotide shorthand notation, some students used valency to differentiate the phosphite trivalent form of the phosphorus atom from the pentavalent phosphodiester group, yet the concept of valency was poorly understood. By virtue of the visual form of a shorthand notation of the 3,5 phosphodiester link in DNA, the valency was incorrectly read. VSEPR theory and the Octet Rule were misunderstood or forgotten when trying to explain the valency of the phosphorus atom in synthetic oligonucleotide intermediates. Plasmid functional domains were generally well-understood although restriction mapping appeared to be a cognitively demanding task. Rote learning and

substitution of definitions were evident in the explanation of promoter and operator functions. The concept of gene expression posed difficulties to many students who believed that genes contain the entity they encode. Transcription and translation of in tandem gene fusions were poorly explained by some students as was the effect of plasmid conformation on transformation and gene expression. With regard to the selection of transformants or the hybridoma, some students could not engage in reasoning or lateral thinking as protoconcepts and domain-specific information were poorly understood. A failure to integrate and reason with factual information on phenotypic traits, media components and biochemical pathways were evident in written and oral presentations. DNA-strand nomenclature and associated function were problematic to some students as they failed to differentiate coding strand from template strand and were prone to interchange the labelling of these. A substitution of labels with those characterizing DNA replication intermediates demonstrated erroneous information transfer. DNA replication models posed difficulties integrating molecular mechanisms and detail with line drawings, coupled with inaccurate illustrations of sequential replication features. Finally, a remediation model is presented, demonstrating a shift in assessment score dispersion from a range of 0 - 4.5 to 4 - 9 when learners are guided metacognitively to work with domain-specific or critical knowledge from an information bank. The present work shows that varied forms of symbolism can present students with complex learning difficulties as the underlying information depicted by these is understood in a superficial way. It is imperative that future studies be focused on the standardization of symbol use, perhaps governed by convention that determines the manner in which threshold information is disseminated on symbol use, coupled by innovative teaching strategies which facilitate an improved understanding of the use of symbolic representations in Molecular Biology. As Molecular Biology advances, it is likely that experts will continue to use new and diverse forms of symbolic representations to explain their findings. The explanation of futuristic Science is likely to develop a symbolic language that will impose great teaching challenges and unimaginable learning difficulties to new generation teachers and learners, respectively.

DECLARATION

I hereby declare that the work forming the basis of this thesis is my own and has not been submitted for any degree or examination at any other university.

Signed _____

at _____ this _____ day of _____ 2007

DEDICATION

To all my students

CONTENT

PREFACE

I express my appreciation and thanks to Professor Trevor R. Anderson for his insightful analyses and guidance of my work. I joined his Science Education Research Group (SERG) with a great deal of enthusiasm, knowing very little about the research approach to Science Education. My research approach has been influenced immensely by Trevor and others belonging to SERG. I am truly grateful for the comments, criticisms and suggestions made by Konrad Schönborn, Lynn du Plessis, Sheelagh Halstead and Don Whitfield. I also thank Bashan L. Govender for assisting me with computer-related difficulties. I declare to all SERG members that my sleepy mood at late afternoon meetings was really due to fatigue and hidden health complications, and not disinterest in any particular person, topic or education theme. I actually developed many aspects of my work during SERG meetings.

I thank the University of KwaZulu-Natal for the fee remission afforded to me as a student-staff member and the National Research Foundation (NRF) (GUN Number 2053218) for financial support through a grant given to Trevor Anderson.

LIST OF FIGURES

LIST OF TABLES

Introduction

Molecular Biology is a dynamic, modern discipline of Science having emerged strongly over the last twenty five years. It has made huge impacts on a number of fields including Medicine, Agriculture, Forensic Science, Palaeontology, Veterinary Science, Fermentation Technology and Food Science. Being interdisciplinary in nature, it incorporates a profusion of concepts mainly from Biochemistry, Biology, Genetics, Chemistry and Physics (Hill *et al*., 2000; Martinez-Gracia *et al*., 2003; Dawson and Schibeci, 2003). It is, therefore, imperative that relevant foundation knowledge or factual information, including terminology and symbolism, be understood in all of these disciplines to enhance our interpretation and understanding of the molecular phenomena affecting different cell types. Certainly, Molecular Biology offers greater depth in our understanding of complex cell processes, *inter alia*, to do with metabolism or mechanisms associated with replicative DNA synthesis, transcription and translation (Chapters 2 and 3).

Students may be fascinated or bewildered by concepts such as genetic engineering, cloning, heterologous gene expression, cell fusion, gene therapy, transgenic plants, hybrid cereal grain or transposition of genes. But regardless of student attitude to these phenomena, teachers should readily expect that simple concepts will not always be clearly understood and should, therefore, be regarded as epistemological impediments to learners (Duit and Treagust, 2003). Invariably, the assimilation of more complex, related concepts could be a difficult task, affecting knowledge generation at large. Students will be always be influenced by their prior learning experiences that have led to misconceptions and syntheses of information which is scientifically unacceptable (Von Glasersfeld, 1992). Erroneous information transfer, reasoning difficulties, superficial understanding, difficulties with the conceptualization of information and poor mental imagery or visualization are some of the problems that students present when subject to a learning exercise (Grayson, 1995; 1996; 2004; Schönborn *et al*., 2006a). Molecular Biology may be considered a

cognitively demanding discipline of Science as it requires the learning, integration and application of theory, and many laboratory techniques, when planning and executing laboratory experiments or manipulations (Caldwell *et al*., 2004). In addition, extensive use of terminology, acronyms and symbolic language is made in Molecular Biology which has been shown in other fields (Garnett *et al*., 1985; Fisher *et al*., 2000; Talbot, 2001) to be extremely confusing to learners.

Over the past forty or so years, there has been extensive research done on students' alternative conceptions and reasoning difficulties in the disciplines of Physics (Harrison *et al*., 1999; Grayson, 2004), Chemistry (Garnett *et al*., 1985; Kogut, 1996; Taber, 1998) and Biology (Kindfield, 1994; Fisher *et al*., 2000; DebBurman, 2002). In contrast, only a limited number of studies have been reported on student difficulties with Biochemistry (Anderson and Grayson, 1994; Talbot, 2001; Feinman, 2004) and the rapidly developing field of Molecular Biology (Schönborn *et al*., 2002a). More recently, socio-ethical responses (Sadler and Ziedler, 2004; Sadler and Ziedler, 2005a; 2005b) to Molecular Biology have prompted the analyses of undergraduate syllabi and subsequent restructuring of curricula (Lock and Miles, 1993; Chen and Raffan, 1999; Hill *et al*., 2000; Martinez-Gracia *et al*., 2003). Details on these matters are presented in Chapter 1.

Aim and research questions of this study

In view of the above, this study reports further on recurring student difficulties encountered in Molecular Biology at a South African university, namely, the former University of Durban-Westville, currently known as the University of KwaZulu-Natal. This study aimed to answer the following research questions:

1. What is the nature and diversity of the symbolic language used in Molecular Biology?

2. To what extent do Biochemistry students find the symbolic and visual language used in Molecular Biology difficult to understand?
3. If so, what is the nature of such difficulties, and
4. What steps could be taken to remediate the difficulties for improved understanding and learning of Molecular Biology?

To address these research questions, the following research areas, which the author had suspected from his 15 years of teaching Molecular Biology might show problems, were investigated:

i. Symbols denoting gene deletion (Δ) and gene fusion or transposition (**::**), respectively.
ii. Shorthand notation of oligonucleotides with specific reference to the identification of phosphite and phosphodiester groups using valency.
iii. Representation of plasmid maps and restriction mapping.
iv. Gene markers: the visual impact, their role in cell fusion technology and the selection of transformants.
v. Nomenclature and function of nucleic acid templates.
vi. DNA replication intermediates and related models.

In addition, this study presents a model aimed at assisting with the remediation of symbolism-related difficulties encountered in Molecular Biology.

CHAPTER 1

LITERATURE SURVEY AND THEORETICAL FRAMEWORK

1.1 Introduction

Fundamental to Science Education research is the understanding and assessment of the learning process. There is a great deal that could be learnt from relevant philosophy, theories and models on the process of learning. Essentially, these form important components of conceptual parameters of Epistemology (Novak, 1984). It is clear, however, that such conceptual parameters alone are insufficient to provide all the answers to the many learning problems encountered in different progressive disciplines. This is especially true for a dynamic, rapidly developing discipline such as Molecular Biology. For this reason, methodological strategies describing the use of oral and written probes (Chapter 3) are routinely adopted in Science Education research (McMillan and Schumacher, 1993; de Jong, 2000; Cohen *et al*., 2000). It is important that methodological strategies adopted generate useful and valid information on the topic which is subject to investigation. As illustrated by Gowin's heuristic Vee device or model (1.6), central to both conceptual (theoretical) and methodological strategies, which flank the left and right arms of the Vee, respectively, are objects, events or the subject matter, depicted at the apex of the Vee, on which observations are made and records of information are documented (Gowin, 1981, cited by Novak, 1984). In essence, Gowin's heuristic Vee indicates interplay of conceptual and methodological elements which are pivotal to the field of Epistemology. Substituting similar elements of this study in the Vee model, the understanding and visual communication of symbolic representations in Molecular Biology would constitute "conceptual parameters". The interplay of these "conceptual parameters" with methodological strategies, namely "investigative probe work" (Chapter 3) is the basis of knowledge generation on learning difficulties. The

learning difficulties are analysed in response to the probes using written and oral communications by the students. Currently, there exists interesting literature (Chapter 2) on the nature and diversity of symbolic language used in Molecular Biology. The current work presents conceptual and reasoning difficulties with some of examples of symbolism which is encountered in the Molecular Biology coursework presented to undergraduate students.

1.2 Selected theories on the learning process

Given the complexity and number of theories on the learning process, only selected theories and factual information relevant to this work will be briefly presented here. Theories provide useful indicators of factors that affect the learning process. Hence, there is tremendous value in assessing the information conveyed by theories, regardless of the field of study. One such factor is the development of cognitive ability among learners (Nurrenbern, 2001; Kirschner, 2002). Certainly, Molecular Biology is a cognitively-demanding subject which presents learning challenges of varying degrees (Chapters 4 to 8). Some concepts may be understood easily while others to do with symbolism and visual language tend to pose greater difficulties. In addition, the author presents factual information from the selected theories which find use in the development of a remediation strategy (Chapter 8).

The established Piagetian theory concerning genetic epistemology infers that the cognitive ability of children is an internal trait, inherited from parents and this variable ability is used in the various developmental stages of the child. Children differ in their ability to understand more complex attributes of life or develop more complex skills as they age. This proficiency manifests in a hierarchical manner where individuals, through intellectual development, accomplish successively more complex skills and operations. The theory also indicates that the direct observation of behaviour tells one about an individual's "internal mental structure" or

"schema"(Stewart, 1985; Carey, 1986; Nurrenbern, 2001). According to Piaget, intellectual growth is linked to the social, linguistic and physical "milieu" of an individual and influenced by processes of organization and adaptation (Nurrenbern, 2001). Mental schemata are idiosyncratic and accommodate knowledge change or assimilation differently in individuals. This theory indicates that cognitive growth occurs independently of any formal mediation such as schooling although some researchers have expressed that a supportive environment does impact positively on intellectual development (Chiras, 1992; Nakhleh, 1994; Gorodetsky and Keiny, 1995; Darby, 2005). Therefore, it is reasonable to expect that mental schemata on subject matter concerning Molecular Biology can be differently constructed. In all Science Education research, it is imperative to place emphasis on careful investigations on the expression of information from different mental schemata so that teachers can address difficulties that students experience and adjust teaching to accommodate individual differences.

The acquisition of cognitive skills or competence may be passed on from adults to children as explained by Vygotsky (Vygotsky, 1978 [original work published in 1926], cited by Taylor, 1994; du Toit, 1997). Adults help children to interpret and conceptualize information and solve problems. In essence, children reflect on various issues and develop innate intelligence and skills of their own. Adults play a meta-cognitive role in this regard. However, children will differ in their ability to benefit from this assisted or mediated learning process. The difference in development potential between unassisted and assisted learners is referred to as the "zone of proximal development". Hence the offer of remediation to learners (Chapter 8) could be beneficial once the nature of learning difficulties is established. It remains the aim of this writer and teacher to assist students overcome their learning difficulties and develop as competent Molecular Biologists.

Unassisted learning may develop a feeling of independence among learners, however, this may also prove to be frustrating when difficulties are encountered. Information

could be learnt incorrectly or interpreted in ways that are useless, radical or unacceptable, as explained by the theory of constructivism (Garnett *et al*., 1985; Driver and Bell, 1986; Von Glasersfeld, 1992). Conversely, when information is transmitted from teacher to student, effective learning may still be unachievable (Treagust *et al*., 1996; Grayson, 1996; Duit and Treagust, 2003). Instead, students tend to integrate factual information into existing knowledge domains or schemata in an idiosyncratic manner (Pearsall, 1997). Unique knowledge structures, based on prior learning, preconceptions and personal experiences in particular environments, may be formulated and perceived to be meaningful. These could differ from what is actually true as expressed by positivists on particular topics (Winberg, 1997). Students may show tendencies to "go beyond the information given" (Bruner, 1973, cited by Cooper, 2005) and should therefore be guided to learn with meaningful inquiry from translated information formats provided by the teacher. This should be appropriate to the learners' state of understanding so as to promote a continuum in terms of the learning process, as described in Bruner's theory of constructivism (Bruner, 1960; 1973; cited by Cooper, 2005; cited in "Teachnology": The Web Portal for Educators, 2005). Curricula should be organised in a "spiral manner" to enhance continued learning based on what the student already knows (Taber, 2003) and concepts should be re-visited to reinforce learning as shown in a remediation model in this study (Chapter 8). Bruner further postulates that intellectual development is influenced by three stages of learning; firstly, the enactive, followed by the iconic and the symbolic. Enactive learning about the world is facilitated through actions on physical objects and the outcomes of these actions. Iconic learning is facilitated through visual impact of models and pictures (Rotbain *et al*., 2006; Beltramini *et al*., 2006; Crisp and Sweiry, 2006). Symbolic learning, as expected of Molecular Biology students and a focus of this thesis, is characterised by a learner's development of capacity to think in abstract terms (Bruner, 1960, 1973; cited by Cooper, 2005; Mathewson, 2005). Predisposition to a learning environment involves tasks which are multi-faceted, demanding various levels of cognition (Kirschner, 2002; Kearney, 2004; Darby, 2005; Gilbert, 2005; 1.4.2).

Ausubel's assimilation theory of learning relates to all disciplines including a relatively new discipline such as Molecular Biology. It describes key concepts which guide research in teaching and learning (Ausubel, 1960; cited by Novak, 1984). The theory reflects on both meaningful and rote learning. With regard to the former, Ausubel defines meaningful learning as the "non arbitrary, substantive, non verbatim incorporation of new knowledge into cognitive structure" (cited by Novak, 1984). Cognitive structure bears reference to schemata or the knowledge frameworks a learner already has from a young age. Learners must relate in a conscious way so as to integrate the new knowledge with the existing knowledge framework or lattice and relate this knowledge to that which is known (Carey, 1986; Pines and West, 1986; Pearsall *et al*., 1997; Kirschner, 2002). Ausubel considers rote learning as a continuum of the knowledge integration process, however, indicates that in this instance learners incorporate knowledge into schemata in an "arbitrary, verbatim, non substantive way" (cited by Novak, 1984). Ausubel introduces the term "subsumption" to describe the idiosyncratic nature of meaningful learning where knowledge is subsumed or integrated into a "cognitive interaction product" (cited by Novak, 1984). This product may be influenced by concepts or misconceptions previously integrated into the knowledge framework. In time, individuals acquire further knowledge about concepts, making them more precise, exclusive and inclusive. This is referred to as "progressive differentiation". "Superordinate learning" may also be encountered during subsumption when more general concepts learnt, indicate meaningful relationships between two or more concepts from an existing knowledge framework (Cho *et al*., 1985). However, learners may experience "cognitive dissonance" or "a negative emotional response" when learning from different texts which present contradictory or different meanings for a single or more concepts (Novak, 1977, cited by Cho *et al*., 1985). The theory also makes reference to the "advanced organiser" who facilitates brief, initial learning of smaller components of knowledge that serve as a bridge between the learner's existing knowledge framework and what has to be learnt. The learning of vital, smaller amounts of information could promote the subsumption of the more complex knowledge (Ausubel, 1960; cited by Novak, 1984)

as shown in a remediation strategy in this study. Smaller amounts of information could be placed in certain knowledge domains (Novak, 1984; Wandersee *et al*., 1994; Herron, 1996) as indicated in the remediation model of this study (Chapter 8). This may involve the hierarchical organisation of facts or protoconcepts (Herron, 1996; Fisher *et al*., 2000) which may be accessed from subordinate knowledge domains, followed by the more complex information in supraordinate domains (Chapter 8).

1.3 Generic features of memory models

The generic features of simple and more elaborate memory models represent three brain areas associated with a sensory register, short-term memory (STM) and long-term memory (LTM), respectively (Stewart and Atkin, 1982; Stewart, 1985). In addition, processes involving information transfer from one brain area to another may also be indicated. The sensory register receives the information input which is briefly maintained for approximately 0.1 to 0.5 seconds. It functions in pattern recognition, feature extraction and perceptual processing. STM or "working memory" is a limited information store which retains the immediate interpretation of events or data encoded by the pattern recognition system of the sensory register. STM holds information for a few seconds to a few minutes while re-organizing and more permanent storage occurs in LTM. STM information may be lost from its "echo box" if not retrieved quickly for rehearsal (Stewart, 1985). LTM, which receives the information from STM, has a permanent and vast storage capacity, coupled with search mechanisms for information retrieval (Stewart, 1985). Of interest to cognitive scientists is the process of subsumption in LTM.

Various forms of symbolism in Molecular Biology convey a wealth of information which must be learnt meaningfully. Often the ability to re-call such information from memory characterizes the learner's interpretation skill. Hence memory is linked to knowledge subsumption and conceptualization (1.3.1).

1.3.1 The subsumption of knowledge and conceptualisation

In accordance with Vygotsky's theory of learning (1.2), children learn from adults once elementary verbal and practical communications are understood. In this mediated learning process, questions are asked, responses are given and concepts are formed. The interaction with intelligent peers, friends, the environment, electronic game stations, educational toys, radio, television, DVD recordings and computers, for example, assists in developing this informal or spontaneous knowledge integration (Mayer, 2003; Kearney, 2004; Parslow, 2004; Darby, 2005). However, recent work presented by Alexander (2007) supports the argument that efforts to bridge the cognitive and sociocultural orientations towards conceptual change are either unnecessary or unachievable.

The expansion of a schema or schemata, however, may not proceed smoothly through formal instruction. When new information is taught, which differs from what was previously believed, a conflict situation could arise regarding the acceptance of concepts or ideas. From an ontological perspective, it is natural to find individuals questioning reality and being recalcitrant to discard or extinguish old ideas and previous beliefs (Duit and Treagust, 2003). Sometimes a "peripheral conceptual change" could be achieved where parts of the old knowledge merge with the new to generate "a hybrid idea" (Gilbert *et al*., 1982; Duit and Treagust, 2003) or to enforce a clearer view (Duit and Treagust, 2003; Van Zele *et al*., 2004). A "formal - symbolic" situation manifests when spontaneous or informal knowledge contributes to minimal interaction with the new knowledge that appears complex and more defined in structure (Pines and West, 1984). In cases where the spontaneous knowledge is extensive, instruction on particular concepts may be superfluous or prompt a feeble re-evaluation of beliefs (Georghiades, 2000). An early conceptual change model indicated that dissatisfaction with a prior conception could initiate dramatic conceptual change, characterised by radical constructivism (Posner *et al*., 1982). The conceptual changes that result may be permanent, temporary or too subtle

to detect. Georghiades (2000; 2004) prescribes metacognition as an important tool in improving conceptual change especially when conceptions are of short duration and when learners are unable to transfer their conceptions from one domain to another in overlapping subject areas. Knowledge retention and retrieval are generally enhanced when meaningful learning takes place. Dissatisfaction with a prior conception may trigger its replacement with a new concept that is "intelligible, plausible and/or fruitful" (Hewson, 1982; Hewson and Thorley, 1989; Duit and Treagust, 2003). Hewson (1981) introduced the term "conceptual capture" to indicate the assimilation of a new conception with the old, clearly indicating that the new, competing conception does not always generate dissatisfaction. Conceptual exchange or accommodation may take place when the new conception achieves higher status than the old conception. The exchange process fails should the old conception achieve higher status (Hewson and Hewson, 1984; Duit and Treagust, 2003) as determined by the student and not the teacher (Wandersee *et al*., 1994). This is in line with constructivist learning theory and the idiosyncratic nature of mental models (Von Glasersfeld, 1992).

Rapid change in the science knowledge content also presents the learner with "sudden insights" of factual detail, facilitating cognitive conflict and the need for rapid conceptual change. Vosniadou and Ioannides (1998) indicate that conceptual change is a rather gradual process where "initial conceptual structures, based on the learners' everyday interpretations of experiences, are continuously enriched and restructured" (cited by Duit and Treagust, 2003). The process depends on the learners' "metaconceptual awareness" where changes are reviewed against initial conceptual knowledge (Duit and Treagust, 2003). In accordance with Vygotskian principles, Guterman (2003) describes the effect of "metacognitive awareness guidance" as a strategy to activate and engage with the students' prior knowledge, so as to enhance reading and comprehension skills. Information is provided on the text before a reading exercise is undertaken. Students engage with this information to answer preliminary questions and formulate advanced thoughts (Ausubel, 1960; 1.2) which

assist in the understanding of the textual information. This concurs with the interactive Socratic learning approach involving dialogue between student and teacher and an emphasis on teaching by asking questions (Tomaska, 2000; Garlikov, 2003; Darby, 2005). From the above, it is clear that the design of instructional approaches can be influential in achieving conceptual change (Harrison *et al*., 1999; Duit and Treagust, 2003).Further information on this matter is provided in section 1.7.

A variety of other factors could influence conceptual change. The use of models may be beneficial to students in this regard, however some students find them intricate and challenging to explain (Garnett and Treagust, 1992a) especially when their reasoning of the abstract is poorly developed (Grosslight *et al*., 1991; Mathewson, 2005; Pata and Sarapuu, 2006). As will be shown in the present study, diagrams or symbols may also be considered as useful representations of textual information which aim to clarify and integrate concepts into mental schemata (Schönborn *et al*., 2002a; 2002b), although in some cases the opposite occurs in that a variety of difficulties with symbolism can be revealed (Chapters 4 to 7). Essentially such external representations are best understood when the mode of representation is integrated with conceptual knowledge and meaningful reasoning as shown in a model by Schönborn and Anderson (2006b). The model suggests three important factors that affect the learners' ability to interpret external representations such as diagrams, symbols, maps, shorthand structures or abridged nomenclature (Chapter 2). These are:

i. The "C factor" which denotes the students' conceptual or prior knowledge of the concepts (C factor) relevant to the external representation (ER).
ii. The "R factor" or students' reasoning ability needed to interpret the ER.
iii. The "M factor" or mode of representation or nature of the ER.

Using a Venn diagram, Schönborn and Anderson (2006b) indicate that reasoning or meaningful interpretation of an external representation (ER) is possible when a learner simultaneously engages the information about the mode of presentation (R-M) and conceptual knowledge (C-M) to interpret the overall content represented in the external representation. An integration of all three factors (C-R-M) facilitates the successful interpretation of an ER.

Relatively abstract phenomena dealing with chemical or biochemical processes could be depicted, however, the interpretation of different diagrams of the same phenomena could be a cognitively demanding task (Kirschner, 2002), generating misconceptions and incorrect ways of reasoning. Schönborn *et al*. (2002a; 2002b) demonstrate such related difficulties in the interpretation of the immunoglobulin (IgG) interaction with an antigenic epitope at the *fab* domain. In the present thesis, extensive evidence will be presented on students conceptual and reasoning difficulties with symbolism and other representations of Molecular Biology knowledge.

Animation and simulations, using video or computer technology, could also enhance conceptual change through visualization (Lowe, 2003; 2004; Hegarty, 2004; Zahn *et al*., 2004; Rieber *et al*., 2004; Ploetzner and Lowe, 2004). Similarly, some teachers advocate the use of analogies when teaching topics that appear to be complex to students (Treagust *et al*., 1992; Dagher, 1995). The inherent disadvantages of the above-mentioned techniques are revealed when misdirected learners (Venville and Treagust, 1997) show superficial understanding of concepts owing to simplification or a failure to extract vital information (Pittman, 1999; Ploetzner and Lowe, 2004). Further some learners may show a tendency to substitute elements of the analogy or simulation in scientific topics where they are irrelevant (Hegarty, 2004; Ploetzner and Lowe, 2004). Lowe (2003) draws attention to the fact that animated visual programmes may be "overwhelming" and far too dynamic, demanding high processing skills and comprehension. Conversely, visual programmes may be "underwhelming" due to aesthetic parameters such that learners do not actively

engage with the visual. However, the combined use of multimodal communications involving animated video, supplementary lectures and still visual displays, for example, could reduce the cognitive load placed on learners as they acquire strategic knowledge which promotes reasoning and conceptual change (Mayer, 2003; Parslow, 2004). Richard Mayer refers to this as multimedia learning (Mayer, 2003). Affective factors are considered important in influencing conceptual change. In this regard, learners benefit from initiatives such as establishing an environment conducive to learning and motivation on the part of both learners and teachers to promote social and group learning (Pintrich *et al*., 1993; de Jong, 2000; Duit and Treagust, 2003; Darby, 2005). Science literacy programmes, research into teaching and learning methods and classroom instructional practice should all receive attention as factors influencing conceptual development, conceptual change and difficulty prevention (de Jong, 2000; Duit and Treagust, 2003; Oh, 2005). In this study, the analysis of learning difficulties encountered by students in Molecular Biology has influenced the design and evaluation of a remediation model (Chapter 8) where conceptual change is assessed.

1.4 Assessment of the learning process

The assessment of the learning process via the use of probes (Chapter 3) is vital in determining whether or not learning has taken place and the nature of any difficulties that might emerge. It also finds use in determining the success or failure of a remedial strategy (Chapter 8). It is therefore important to understand the benefits and shortfalls of certain types of assessment of the learning process. In this study, a range of topics which illustrate symbolism in Molecular Biology are assessed in terms of understanding from both oral and written communications (1.4.3; Chapter 3). The "visual language" expressed by students must be analysed for knowledge subsumption and a clear understanding of the underlying concepts conveyed by symbolism in Molecular Biology (Kozma, 2003; Ferk *et al*., 2003; Takayama, 2005).

1.4.1 A typical learning curve

A typical learning curve, illustrated by Reynolds (1965) (cited by Atherton, 2003a), indicates that the learning process does not proceed exponentially with time. It begins gradually in terms of competence and may be characterised by short or long exponential phases, followed randomly by intermittent plateaus, a downward "negative slope" and troughs. The transition of one phase to another is characterised by energy release which manifests in the form of freedom to concentrate on other factual material. The progression of learning curves for learners will be different and characteristically reflect the pattern of knowledge subsumption and conceptualisation. It is reasonable to correlate progressive learning with the exponential phase where knowledge integration into schemata proceeds unhindered, coupled with meaningful engagement with the new information (1.3.1). Plateaus could be indicative of sustained subsumption based on knowledge frameworks which maintain a continued level of knowledge integration. Downward slopes and troughs possibly indicate difficulties such as poor conceptual capture, manifestation of misconceptions, failure to expand schemata, reasoning difficulties or poor knowledge transfer skills from related knowledge domains. In view of the above, it is necessary to constantly assess the learning process and find ways of enhancing it as shown in the remediation strategy of this study (Chapter 8).

1.4.2 Assessment strategies

Sundberg (2002) defines assessment as "a systematic method to determine if, and to what extent, student learning has occurred". The assessment may be summative which usually takes the form of an examination at the end of the coursework and grades are given to reflect the students' knowledge of the course. Both practical and theoretical components of a course may be tested this way in a written examination. However, formative assessments are generally more reliable as evaluations are made

during the course of instruction (Sundberg, 2002; Black and William, 2006). This allows for, if necessary, the transformation of content knowledge to a form that is better understood by the learners (Chen and Ennis, 1995; Black and William, 2006). Shulman (1986) refers to this transformed knowledge as "pedagogical content knowledge" (PCK). A more holistic approach is provided by Rhemtula and Rollnick (2002) who indicate that PCK entails an integrated understanding of the subject, knowledge of the students' difficulties and prior learning, knowledge of other curricula and how they affect understanding of the coursework in the programme, use of instructional media and general attributes of the teaching to do with facilities, suitably trained, competent staff and a classroom conducive to good teaching (Darby, 2005). Caldwell *et al*. (2004) describe a skills matrix for undergraduate programmes in Molecular Biology laboratory practice. Skills are taught with the aim to promote problem solving ability, critical thinking, the ability to reinforce observations and record data accurately and communicate results in written and oral form, incorporating the learning of theory and laboratory practice. These evaluations are best carried out in a formative manner to effect course transformations and remedial exercises (Caldwell *et al*., 2004; Black and William, 2006). Sundberg (2002) indicates that both formative and summative assessment strategies are respectively necessary to evaluate teaching transformations during the progression of a course and its overall success towards the end. Assessments are generally uniformly applied to a class of learners, however, exceptions may be considered for those learners who are differently able. For example, dyslectic and blind learners may be interviewed, given additional examination time, assisted with examination formats in Braille (Botha, 1991) or the use of haptic technology to get tactile feedback (Jones *et al*., 2006). Quantitative and qualitative methods of assessment are both employed, provided they yield information on specific course objectives, standards, skills, content knowledge and types of student learning difficulties (Shulman, 1986; Sundberg, 2002; Caldwell *et al*., 2004). Pre and post instruction examinations are commonly used to assess students. Where large student numbers are encountered, different learning difficulties may be determined qualitatively. More definitive, quantitative analyses, involving

statistical techniques, could be carried out to determine, *inter alia*, correlations, regressions or variance in marks in a question paper or the prevalence of learning difficulties (Snedecor and Cochran, 1980; McMillan and Schumacher, 1993). In the case of multiple choice type questions (Treagust, 1988; 1.4.3), statistics concerning item difficulty, item discrimination and distractor power may be applied to student answers to establish the usefulness and validity of the questions (Zurawsky, 1998).

Bloom (1956) (Anderson and Krathwohl, 2001; cited by Atherton, 2003b) identifies three domains of learning, namely, the cognitive, affective and psycho-motor. Different cues are provided by Bloom that could facilitate the assessment of cognitive skills which are increasingly demanding. In the cognitive domain, for example, "knowledge acquisition" is regarded as the baseline skill. "Comprehension", "application", "analysis", "synthesis" and "evaluation" are higher order, serial skills that follow in this hierarchical sequence. Progression to a higher order skill may be addressed by the teacher once competence is assessed at a lower level. A revision of cognitive skills, now expressed as verbs, regards "remembering" as a baseline skill, followed by "understanding", "applying", "analysing", "evaluating" and "creating" (Anderson and Krathwohl, 2001). The affective domain is associated with the assessment of values and perceptions while the psycho-motor domain addresses the fundamental role of imitation in skill acquisition. Both cognitive and affective learning domains are linked to the learners' predisposition to a learning environment and influence the ability to effect conceptual change (1.3.1; Pintrich *et al*., 2001; Anderson and Krathwohl, 2001).

1.4.3 Assessment as a tool to identify student difficulties

Assessments usually involve the analysis of written material including the answers to question papers, journal writes, assignments, essays, projects (Kogut, 1996; Collis *et al*., 1998; Duchovic, 1998; Sundberg, 2002) or the use of concept mapping (White

and Gunstone, 1992; Van Zele *et al*., 2004). Concept maps can be viewed as node-linked diagrams which show hierarchical levels of concepts within a knowledge domain, information clusters and cross relations (Treagust, 1988; Wallace and Mintzes, 1990; White and Gunstone, 1992). These provide an assessment of conceptual understanding, ability to differentiate facts and integrate knowledge as a framework (Pearsall *et al*., 1997; Van Zele *et al*., 2004). In this study, concept maps were not used as an assessment technique. They may be considered as a useful tool making general probes but require augmentation using more focused written probes and interviews as done in this study (Chapter 3). Treagust (1988) describes the use of two-tier multiple choice questions which can be used to assess misconceptions. The first tier involves a range of statements which may be true or false. In the second tier students are requested to select from a second set of statements that correspond to possible reasons for the choice they made in the first tier. Our Science Education Research Group (SERG), University of Kwa-Zulu Natal, Pietermaritzburg, suggests the use of a single tier of statements plus the request for an open-ended reason or justification for the choice as shown by Tamir (1989) and Amir and Tamir (1994). Gupthar (1996) indicates that false-statement type questions may be used as an alternative or accompanying questioning strategy to assess students' understanding of course material. Students are required to compare such statements with information that has been taught and learnt correctly. The success of this comparison depends on the depth of understanding of course material and therefore the confidence with which it may be used to support or refute the contention (Zohar and Nemet, 2002). In this technique, the subtlety of misinformation is varied according to the standard of course work delivered at a particular level. Tomaska (2000) describes a Socratic learning approach (1.3.1) where a teacher consciously delivers error-prone talks, in a limited way, and invites corrections from the students who have had some prior exposure to related scientific literature. A credit system motivates students to participate actively in this interactive process and they develop improved learning aptitude (Snow and Lohman, 1984; Tomaska, 2000). Students may also develop argumentation and rebuttal skills when directed to specific content knowledge in

metacognitive thinking exercises (Zohar and Nemet, 2002; Georghiades, 2004). In the present study, techniques involving the use of any misconceptions, as for example distractors in multiple choice questions, were not used as some students can show a tendency to remember undesired information (Amir and Tamir, 1994; Gupthar, 1996). "Predict-Observe-Explain" (POE) tasks probe students' understanding by requiring the following actions; (i) prediction and justification of the outcome of certain events or phenomena, (ii) provision of descriptions of what actually occurs and (iii) comparison of predictions with the observations and reconciliation of differences or aspects contributing to conflict (White and Gunstone, 1992; Liew and Treagust, 1995; Kearney 2004). "The Art of Assessing" website, created by Phil Race (1996), describes a range of assessments, *inter alia*, to do with short questions, reviews, true and false questions, "fill in blanks", "complete the statements", multiple choice questions, practical skills or analyses of a practical product. In the present study, short questions made excellent general probes (Chapter 3), revealing a range of learning difficulties in different Molecular Biology topics (Chapters 4 to 8).

The assessment of oral communications could include seminars, tactical interviews (Posner and Gertzog, 1982; White and Gunstone, 1992; Harrison *et al*., 1999), Socratic dialogue-inducing laboratories (Hake, 2008) or "think-aloud" exercises involving the combined use of annotated diagrams and interviews to elicit specific responses that relate to diagrams (de Jong, 2000; Schönborn *et al*., 2002a). Alternatively, student-generated diagrams (SGDs) may also be assessed for the understanding of underlying concepts (Kozma, 2003; Schnotz and Bannert, 2003) and the possible inclusion of idiosyncratic depiction of information (Schönborn *et al*., 2002a;b). It is imperative that assessments generate information on the achievement of course objectives, student knowledge and the nature of learning difficulties. In the present study, the writer made extensive use of interviews, SGDs and think-aloud exercises (Chapter 3) to collect information on student interpretation and understanding of symbolism, visual models and molecular processes taught in Molecular Biology.

1.5 The nature of learning difficulties in Molecular Biology and related fields

There has been extensive research done on a wide range of students' reasoning difficulties and alternative conceptions in the disciplines of Chemistry (Garnett *et al*., 1985; Garnett and Hackling, 1995; Anderson and McKenzie, 2007, CARD, website in reference list), Physics (Pfundt and Duit, 1994; Duit, 2007) and Biology (Kindfield, 1994; Fisher *et al*., 2000). In contrast, only a limited number of studies have reported on student difficulties with Biochemistry (Anderson and Grayson, 1994; Talbot, 2001; Garcia-Vallve, 2004) and Molecular Biology (Fisher, 1985; Schönborn *et al*., 2002a). Of particular interest, there are several reports on alternative conceptions or misconceptions, in the area of traditional Genetics, concerning the "gene", "allele" and "chromosome" (Rotbain *et al*., 2006). The following are examples of such difficulties extracted from the literature:

i. "Genes are characteristic cells passed from one generation to another" (Venville and Treagust, 1998).
ii. "Genes are cells that make you different" (Venville and Treagust, 1998).
iii. "Genes are small trait-bearing particles" (Lewis and Kattmann, 2004).
iv. "Genes are found in specific organs or tissues such as the reproductive system" (Lewis and Wood- Robinson, 2000).
v. "Genes are larger than chromosomes and are made up of chromosomes" (Lewis and Wood-Robinson, 2000).
vi. "Chromosome is probably a gene" (Longden, 1982).
vii. "Genes make babies" (Venville and Donovan, 2007).

From the above, it is clear that little consideration is given to the identity of a gene, its composition or relative relationship to a chromosome. Cho *et al*. (1985) report that students commonly use the term "gene" and "allele" in an interchangeable manner, providing no differentiation that the latter is one of a gene's possible forms. A failure to differentiate closely related terms may be ascribed to cognitive dissonance (Cho *et al*., 1985). This is also documented in an impressive case study involving a physics student's inability to initially tell the difference between "heat" and "temperature"

(Harrison *et al*., 1999). In another example, erroneous concept substitution for the left and right arms of a chromosome can be seen in the statement, " if a chromosome has two alleles they would be joined to each other by the centromere and each part would be an allele" (Banet and Ayuso, 2000). Basic information pertaining to the location of chromosomes can be incorrectly learnt or drawn from intuition (Garnett *et al*., 1995; McKloskey, 1983, cited by Cho *et al*., 1985) or naive beliefs (Caramazza *et al*., 1981, cited by Cho *et al*., 1985) such as "chromosomes are found in your brain" or "chromosomes are part of plant cells for photosynthesis" (Venville and Treagust, 1998). When basic protoconcepts are poorly understood, this invariably impacts negatively on the understanding of more complex related issues as explained by principles of epistemic scaffolding (Sandoval and Reiser, 2004). Lewis and Kattmann (2004) present the following excerpt from a student, "In the cells there are chromosomes which contain pigments…..for the colour of the eyes". In this case, the student shows difficulty in distinguishing between genotype and phenotype and is oblivious of a mechanism by which a gene could be expressed in the phenotype. A number of other studies have reported difficulties associated with the genetic code, gene insertion and the expression of a codified protein which may be responsible for a new trait or phenotype (Fox, 1996; Lewis and Wood-Robinson, 2000; Lewis *et al*., 2000; Marbach-Ad, 2001). Some textbooks on Genetic Engineering and Biotechnology do not provide sufficient information on basic protoconcepts which underpin more complex issues (Lock and Miles, 1993; Chen and Raffan, 1999; Hill *et al*., 2000) or present foundation knowledge in an incoherent manner (Martinez-Gracia *et al*., 2003). For example, students may show a poor understanding of the concept of a gene or biological transformants at a lower grade, yet they may be introduced to more complex strategies involving genetic manipulation of biological specimens at the ensuing higher grade (Martinez-Gracia *et al*., 2003). Recently, Hancock (2006) reported on undergraduate students' inability to define basic concepts such as phenotypic expression, variation, variance and variability in studies involving population genetics. The lack of understanding of these protoconcepts affected their interpretation of more complex aspects concerning the heritability of phenotypic traits

in different populations. Definitions and foundation material in some textbooks, prescribed for higher grades, are reported to be poorly expanded upon (Lock and Miles, 1993; Chen and Raffan, 1999; Hill *et al*., 2000; Martinez-Gracia *et al*., 2003). Simonneaux (2000) draws attention to the fact that some students may perceive viruses and bacteria as invasive agents or germs, let alone show any understanding of basic genetic manipulation or biotechnology associated with these. The following student responses are presented in that study:

i. "Viruses have sharp teeth, spikes, pitchfork features making them more dangerous than bacteria".
ii. "Viruses can eat up other cells to survive".
iii. "Once viruses feed on other cells, they become big and divide and multiply and attack other cells".
iv. "Bacteria have to feed on waste. If there is no more waste, they don't have anything. Bacteria can therefore be empty if they haven't been able to feed".

Simmoneaux (2000) also reports on linguistic confusion, presenting the following response from a student; "If M5 touches (*contacts*) a sick person, he will *contract* the disease". With regard to biotechnological processes to do with yoghurt and cheese fermentations, the following are reported (Simmoneaux 2000), respectively:

i. "It's done by fermenting milk. The bacteria attack the organic matter in milk and so they get rid of the organic matter in it".
ii. "Yes. There are bacteria on cheese. Its like a fungus actually…because there's mini-fungus which is dust, mini-fungus gets eaten with cheese".

Basic concepts such as "cloning", "genetic engineering", "biotechnology", "genetically modified foods" or "organisms" are poorly understood among school (Dawson and Schibeci, 2003; Martinez-Gracia *et al*., 2003; Seethaler and Linn, 2004) and undergraduate college students (Chapman, 2001; DebBurman, 2002), raising concerns that this will impact on socio-ethical decision-making and religious practice

(Marchant and Marchant, 1999; Sadler and Zeidler, 2004 ; 2005a; 2005b). When teaching aspects of Molecular Biology, it is natural to expect students to reflect on unique, vivid mental imagery of processes or the facts presented, however, caution should be exercised that some students can show a tendency to "go beyond the information presented" (Bruner, 1973, cited by Cooper, 2005; 1.2). For example, students who have a poor understanding of gene function may believe that transferring a gene from an animal to plant allows the plant to develop some characteristics of the animal (Chen and Raffan, 1999). Simonneaux (2001) reports on students' concerns when presenting a theoretical case study involving the propagation of transgenic salmon at a farm near a seaside village. Some students were anxious that giant, transgenic salmon would emerge from the farm and cause havoc to the ecosystem, reducing the population of smaller fish and other forms of marine life.

Molecular mechanisms such as allele segregation, chromosomal non-disjunction, gene assortment or crossing-over between homologous chromosomes are poorly understood as students are unable to define a chromosome nor differentiate it from a chromatid or simply fail to indicate the differences between mitosis, meiosis and replication (Cho *et al*., 1985; Stewart and Dale, 1989; Fisher, 1992; Kindfield, 1994). Students may draw diagrams correctly depicting the behaviour of chromosomes during cell processes but offer varying degrees of understanding when trying to explain underlying concepts (Longden, 1982; Kindfield, 1994; Kozma, 2003). Similarly, Chemistry students may be adept at balancing chemical equations yet fail to draw diagrammatic representations of the equations at a particulate or molecular level (Yarroch, 1985). Diagrams can be used as "tools to think with" (Kindfield, 1994; Pata and Sarapuu, 2006) yet in some cases, owing to poor design or complex detail, they can expose serious learning difficulties (Kirschner, 2002; Schnotz and Bannert, 2003). For example, Menger *et al*. (1998) have reported that students can be misled when micelle structures are presented in a form that resembles "spokes of a wheel" (cited by Schönborn *et al*., 2002a). Schönborn *et al*. (2002a) report on student difficulties with the interpretation of a stylized textbook diagram of immunoglobulin

G (IgG). Students referred to the antigenic epitope as "an agent trying to enter the antibody" at the $(fab)_2$ domain of the molecule, a process analogous to the way a foreign agent would "attack" or "invade" a host or body. In the same study, some students inappropriately referred to the Y-shape immunoglobulin as a DNA molecule's replicating fork. Inappropriate information transfer, surface-level reasoning and knowledge integration failure (Thorndyke and Stasz, 1985) were clearly evident in the study described by Schönborn and co-workers. Knowledge integration failure, rote learning and poor understanding were also encountered by Anderson and Grayson (1994) when teaching carbohydrate metabolism to students taking an introductory Biochemistry course. Students were found to engage in excessive memorization of different metabolic pathways, with poor conceptual understanding of the facts and inability to find linkage and inter-relationships between different pathways. Some students showed "localised reasoning" difficulties, characterised by a failure to predict the progression or non progression of a reaction series subject to the inhibition of an earlier step. In a subsequent study, dealing with students' understanding of oxidative phosphorylation, Grayson *et al*. (2001) describe how students inappropriately transfer knowledge in terms of the everyday use and meaning of the word " spontaneous" to indicate that a "spontaneous biochemical reaction" also proceeds freely, without the influence of an enzyme or activation energy. Everyday language has also been found to influence students' alternative conceptions in Chemistry. For example, the term "particle" in common everyday language would mean a small, visible piece of solid material or substance. In Chemistry, it bears reference to an atom, ion or molecule (Gilbert *et al*., 1982). A sound conceptual understanding of the particulate or submicroscopic nature of matter can be a problem as students find it confusing or fail to visualise it. Sometimes textbooks present a similar dilemma, indicating that the zwitterion or dipolar ion forms of amino acids exist only in aqueous solution yet they are present in the crystalline state and contribute to the high melting points of amino acids (Talbot, 2001). Fisher (1985) describes that some students were uncertain regarding the origin of amino acids and referred to them as proteins. Furthermore, the protein building

blocks were perceived to be products of translation. Students showed conflict that enzymes are participants in the translation process and that they originate through the same process.

There are hundreds of other examples of student difficulties which might be of interest to subject specialists. As far as possible, the nature of these difficulties should be characterised and subject to remediation. Specific difficulties in the literature will be addressed further in this work as comparison is drawn with current findings (Chapters 4 to 8).

1.6 The nature of difficulties with symbolism, nomenclature and visual models

In Science Education, symbolism may be defined as the applied use of symbols or iconic representations to depict particular information (Salomon, 2005; Gilbert, 2005). The expression or communication of information using symbols and iconic representations constitutes "symbolic language" (Salomon, 2005). The use of symbols requires a clear understanding of their meaning in specific contexts and applications as shown in diverse and specialized fields such as pharmacokinetics (Hayton, 1990), microbial growth kinetics (Barayani and Pin, 2001; Vadasz and Vadasz, 2005), human genetics (Kinnear, 1992) or bioinformatics (Leung *et al*., 2001), just to mention a few examples. In Mathematics, Vlassis (2004) reports on students' difficulties with the negative sign and its treatment in the reduction of polynomial equations. Colin (2002) reports on students' difficulties with the depiction of refractive patterns of lenses and image analyses. In Biochemistry, syntactic diagrams depicting a cycle of metabolic transitions can be interpreted differently by novices. An interesting report by Hull (2003) indicates students' perception that the citric acid cycle takes place in a circular fashion to fit the physical confines of a spheroidal cell. They indicated the transition of intermediates to take this format using arrows to depict the "circular" nature of the citric acid cycle. Perini (2005)

advocates that the information depicted by certain diagrams could be expressed linguistically, making them less complex for students to interpret. In Biology, students have presented difficulties in their reasoning of the mechanisms of mitosis and meiosis from diagrams because fundamental concepts such as the chromosome, chromatids, centromere, chromosome segregation, allele or crossing-over between homologous chromosomes were not understood (Cho *et al*., 1985; Stewart and Dale, 1989; Kindfield, 1994). Other researchers report that information transfer may be influenced by symbol structure. Generally, simple symbols rather than those perceptually rich in detail or "concreteness" can be used more effectively to facilitate the transfer of information to learners (Sloutsky *et al*., 2005). Such simple symbols may be abstract in nature but they impose a lesser burden to learners (Bechtel, 1998; Kirschner, 2002) who might be overwhelmed with structural detail of the representation, let alone explain the information that it actually depicts (Bechtel, 1998; Sloutsky *et al*., 2005). Stylianidou and Boohan (1998) describe the use of "abstract picture language" or drawings to enhance students' understanding of the Second Law of Thermodynamics concerning energy and the nature of change in the environment. Syntactic diagrams are commonly used in Biochemistry to illustrate "component features" of single-step, reversible or sequential reactions which involve changes to substrates, the interaction of co-factors, substrate-enzyme binding or product formation (Dutkiewicz, 1982; Bechtel, 1998; Perini, 2005). Thus the symbolism representing the "component features" of chemical reactions can take the form of chemical structures, formulae, reaction intermediates, names of enzymes, horizontal arrows showing forward or reversible reaction trends or curved arrows showing incoming reactants or co-factors or liberated intermediates and the formation of products (Dutkiewicz, 1982; Perini, 2005). This diversity of information can be difficult for learners to grasp (Perini, 2005) especially when students fail to correlate the molecular or submicroscopic attributes of a reaction with the symbolic representation. The submicroscopic entities such as atoms, bonds, functional groups, ions, electron transfers, electronegativity, nucleophilic attack or ionization may not be visualized by novices (Ferk *et al*., 2003; Treagust *et al*., 2003).

Experts on the other hand use a range of symbolism to depict the submicroscopic or molecular detail and expect learners to translate or make the association between such representations and the information concerning the submicroscopic. This teaching trend can be problematic to learners as the underlying information concerning such depictions is not always understood (Bowen, 1998; Kozma, 2003; Barak and Dori, 2005). In Chemistry, it has been reported that students may provide the chemical formulae of reactants and products, balance the equations but fail to explain the molecular basis of the reactions (Yarroch, 1985). Other students simply associate lines and letters to depict bonds and atoms found in listed in the periodic table (Barak and Dori, 2005). When challenged through the use of interactive computer-driven programmes to predict structures and bonding patterns, such students are compelled to revisit the theoretical information that governs bonding and structure prediction (Barnea and Dori, 1996; Bowen, 1998; Kozma, 2003; Jones *et al*., 2005; Barak and Dori, 2005). In addition, the computer-aided exercises promote visualization of molecular detail, making rather it easier for learners to comprehend concepts such as the spatial arrangement of atoms or bonding ability based on the electronic configuration of atoms (Barnea and Dori, 1996; Kozma, 2003; Jones *et al*., 2005; Barak and Dori, 2005). Atomic space-fill models may be colour-coded to enhance visualization of the spatial arrangement and distribution of atoms in 3-D models (Jones *et al*., 2005; Barak and Dori, 2005); however, students may still find this rather abstract and difficult to comprehend especially when such external representations encode molecular information which is poorly learnt (Lord, 1985; Ferk *et al*., 2003; Wu and Shah, 2004). Experts tend to reflect on theory and features of an external representation such that they "see" the theoretical attributes in such a display (Lowe, 1988; Kozma and Russell, 1997; Kozma, 2003). Their visual cognition (Gilbert, 2005) may appear to be superior but experts can be biased and tend to select certain representations to enhance or support their own knowledge of the subject under consideration. This type of "bootstrapping" (2.5.5; Cheng *et al*., 2001) is not apparent among novices who are not only inexperienced with certain forms of external representations (Schönborn and Anderson, 2006b) but have poor conceptual

knowledge on the subject. diSessa (1993) describes the knowledge acquired by novices as one composed of many small units which he refers to as "phenomenological primitives or p-prims". This is based on their superficial interpretation of representations which actually depict a wealth of information. In Physics, experts and novices have also been reported to show differences in cognitive ability and knowledge usage. Based on their deeper understanding of principles, Physics experts had structured questions in two groups, labelled "force problems" and "energy problems". Novices were reported to reflect superficially on these questions by referring to them as "pulley problems" or "plane inclination problems" (Chi *et al*., 1981; Larkin *et al*., 1980, cited by Kozma, 2003).

In eye-tracking experiments, Cook *et al*. (2006) report that novices, despite the lack of knowledge, tend to focus on salient features of DNA replication in 2D and 3D diagrams in the same way that experts do; however, their limited knowledge renders them incapable of explaining visual features concerning DNA replication. Similarly, undergraduate students may find sequential steps of DNA replication rather difficult to explain or illustrate when their knowledge of the associated enzymology is poor (Fossey and Hancock, 2005). An expression of "visual language", addressing visual subtlety, the submicroscopic or fine molecular visualization, is therefore not possible in this instance (Treagust *et al*., 2003; Jones *et al*., 2005; Yore and Treagust, 2006). Visual displays can show complex information and are also varied in their presentation of symbolism. Structural formulae, Lewis structures (Peterson and Treagust, 1989; Peterson *et al*., 1989), ball-and-stick models (Treagust *et al*., 2003; Rotbain *et al*., 2006), organic line structures or models (Fossey and Hancock, 2005), protein ribbon models (Richardson and Richardson, 2002), atomic space-fill models (Ferk *et al*., 2003; Wu and Shah, 2004), electron micrographs (Valentine and Green, 1967, cited by Schonborn and Anderson, 2006a), chromogenic enzyme-linked immunosorbent tests (Klein, 1990), computer-aided graphics (Rieber *et al*., 2004; Ploetzner and Lowe, 2004; Sins *et al*., 2005) or dynamic animations are some forms of visual displays (Hegarty, 2004; Ploetzner and Lowe, 2004). Apart from the

conceptual information that these models carry, learners may find them cognitively challenging to decipher or correlate related theory with a range of symbolic visual displays. Multimedia packages offer a form of visualization which might be overwhelming and difficult to understand. Further, mental imagery of processes can be rather idiosyncratic depending on "what the learner already knows" (Ausubel, 1968). However, multimedia-facilitated learning of symbolism can promote coherence and contiguity in terms of a mental model (Mayer, 2003) which is influenced by dual processing of verbal and pictorial representation in its content (Paivio, 1986).

1.7 Analysis and remediation of learning difficulties

A wide range of remediation strategies (Vogelezang, 1987; Geddis *et al*., 1993; Kogut, 1996; Greenbowe, 1994; de Jong, 2000; Johnson, 2000; Huddle *et al*., 2000; Stokstad, 2001; Chapman, 2001; Stokstad, 2001; DebBurman, 2002; Ozkaya, 2002; Chandler, 2004; Hegarty, 2004; Lowe, 2004; Sins *et al*., 2005; Beltramini *et al*., 2006; Orgill and Bodner, 2007) designed to address learning difficulties in various disciplines can be found in the literature. A single strategy can be specific for a difficulty that is associated with a particular discipline; however, there are many strategies which find common use and application in different disciplines (CARD database, website in list of references). In introductory courses, remediation strategies tend to simplify concepts in order to promote understanding and metacognitive thinking (Zohar and Nemet, 2002; Georghiades, 2004). For example, Vogelezang (1987) suggests that elementary school Chemistry courses should initially introduce students to the concept "substance" rather than the elusive "atoms and molecules". Students relate to this concept easily from their personal experiences. Molecules could be perceived as small pieces of a substance which may disappear in a chemical reaction or re-appear in a new form. An emphasis of the concept of "substances", their properties to undergo change were also considered as important pre-requisites to

understanding when teaching elementary Chemistry to novices (Johnson, 2000). Conceptual change or substitution with more complex knowledge occurs when such learners accept a conception as "intelligible, plausible and fruitful" (Hewson and Thorley, 1989; 1.4), based on what they already know (Salomon and Perkins, 1989) and form connections to generate more complex knowledge structures (Thorndyke and Stasz, 1985; Grayson, 1995; Duit and Treagust, 2003). Huddle *et al*. (2000) report on improved students' perception of an electrochemical reaction in a cell when representations of ions, electrons and conducting wire were made using polystyrene balls, marbles and a hosepipe, respectively. The visual impact of the model assisted students in showing that electrons move through the conducting wire and not through the salt bridge or electrolytes as originally believed. Also, students were able to show the correct movement of ions in the electrolytes and via the salt bridge. Treagust *et al*. (2003) report that the submicroscopic detail of Chemistry can be better understood when students are presented appropriate symbolic representations of the submicroscopic level, such as models or structural formulae, which could be correlated with the "macroscopic" or experimental experience they receive. Apart from the use of models, computer animation or simulation such as the "Electrochemical Cells Workbench" programme can also assist students visualise and better understand the microscopic and rather dynamic processes which take place in electrochemical cells (Greenbowe, 1994). Three dimensional models, photographs and computer-generated models of molecular structures may be more effective in enhancing students' understanding of structures than using schematic representations or stereochemical formulae (Ferk *et al*., 2003; Cook *et al*., 2006). Recently, Beltramini *et al*. (2006) reported on students' use of a precision-made, three-dimensional plastic atomic modelling kit to construct models of DNA and RNA and its enhancement of conceptual understanding of molecular mechanisms concerning DNA replication and transcription. Computer-assisted instruction (CAI) may also be used to generate cognitive conflict and conceptual change (1.3.1) among learners when presented with dynamic graphs and interactive opportunities (Rieber *et al*., 2004; Ploetzner and Lowe, 2004; Sins *et al*., 2005). This has been reported in studies

concerning learners' initial misconceptions about chemical equilibrium (Hameed *et al*., 1993) and a preconception linked to an undifferentiated weight-density concept, respectively (Kang *et al*., 2005). Ozkaya (2002) describes a conceptual change technique used to teach the chemistry of galvanic and electrolytic cells. An experimental student group would be told about common misconceptions and given the reasons as to why they are incorrect. This group would be subject to true and false statement-type questions (Gupthar, 1996), based on the misconceptions, to test their conceptual understanding rather than problem-solving ability (Taconis *et al*., 2001). This enabled a comparison with a control group of students who did not receive the same treatment. Overall, the instructional technique could be applied to all students, pending the success of the initial assessment, to overcome misconceptions. Treagust (1988) applies the true and false or multiple choice questioning strategy but confronts the students to provide a reason for their choice. This enables assessment (1.4.3) and remedial instruction on very specific issues. Predict-Observe-Explain (POE) tasks, think-aloud exercises and interviews can also be used in a similar way (1.4.3). Hand and Treagust (1988) report on a remediation strategy involving the use of conceptual conflict to challenge students with previously identified pre- or misconceptions about acids and bases. The students received lessons, followed by experimental work, worksheets incorporating the pre- or misconceptions and related questions, and then engaged in group and whole class discussions. The worksheet questions challenged the students on the correctness of the conceptions. Group discussions and interviews facilitated further confrontation of the students' conceptions, with the more able students being keen to relate inconsistencies. A close correlation between worksheet assessments and active participation at group discussions was noted. In addition, subsequent class tests revealed an improvement in the understanding of some of the misconceptions.

The need to establish pedagogical content knowledge is an important pre-requisite to both teaching and the adoption of a remediation strategy (Geddis *et al*., 1993; de Jong, 2000). As shown by Geddis *et al*. (1993), students are required to understand

the basic concept of isotopes and measurements of their atomic masses in order to extend on the teaching of related atomic theory. Prior knowledge, effective teaching, the ability to understand alternative representations of the subject matter and importance of the topic to the overall curriculum constitutes "curricular saliency" (Geddis *et al*., 1993; Schnotz and Bannert, 2003). Grayson (1995) draws attention to the fact that "a student may be said to know what speed is if he or she can state how fast an object is moving. Such knowledge is declarative knowledge". In contrast, procedural knowledge requires a higher level of cognition where the student needs to know more than the definition in order to calculate speed, taking in effect the distance travelled divided by time. Prior knowledge may be established by interviewing students (Hand and Treagust, 1988; Harrison *et al*., 1999) or conducting a pre-test (Grayson, 1995), followed by class discussions, debates and argumentation which may be guided carefully by the teacher to promote remediation (Zohar and Nemet, 2002). In this way, students may discard or exchange facts or beliefs which are untrue or expand on knowledge schemata which are encouraged (1.3.1). The integration of knowledge may be fostered even further once learners are subject to think critically (Kogut, 1996; Huitt, 1998), analytically and quantitatively when presented with problem-solving exercises (Taconis *et al*., 2001; Wood, 2002). Much emphasis has been placed on the strategy "understanding by doing" (Georghiades, 2000; 2004). Novak (1984) draws attention to the fact students may record observations from laboratory exercises yet not conscientiously consider appropriate principles and theory to explain their results or observations. Novak (1984) advocates the use of Gowin's heuristic Vee device to assist in interpreting laboratory work. Typical Vee maps, constructed by students, incorporate conceptual parameters on the left arm of the Vee including appropriate theory, principles and concepts while methodological parameters are indicated flanking the right arm. This includes a record of claims and laboratory findings. The apex of the Vee records selected objects or events to view. The trough of the Vee presents a focus question which relates to the aim of the exercise. Thus the heuristic device provides a holistic approach to "understanding by doing". Concept maps are particularly useful in assessing the understanding of

domain-specific knowledge, cross relations, information clusters and a hierarchy of conceptual knowledge structures (1.4.3) and could be used in combination with Gowin's heuristic Vee device to test for knowledge integration (Novak, 1984).

Tutorials generally place an emphasis on improving students' understanding of selected aspects of course curricula through active, inquiry-based learning (Stokstad, 2001). Remedial exercises may take the form of assistance with problem-solving, encouraging discussion among students in study groups, directing questions that encourage critical thinking and setting of assignments and essay questions which promote understanding and communication (Kogut, 1996; Stokstad, 2001; Chapman, 2001; DebBurman, 2002). Further "examination repechage" should be encouraged by tutors or instructors where students are given a second chance ("repechage", the French, meaning "second chance") to provide solutions to problems and discuss this further with the instructor (Duchovic, 1998). This opportunity may follow any test or examination where students consult textbooks or notes or seek assistance freely from other students (Kearney, 2004) or the instructor (Oh, 2005). In the current study, consultation with the instructor and "examination repechage" are important features of a remediation model and will be discussed in Chapter 8.

Diagrams, animations, audio-visual facilities and interactive computer programmes (1.6) may be used during tutorial or remedial exercises to promote mental "imagery" (Chandler, 2004; Hegarty, 2004; Lowe, 2004) and haptic responses (Jones *et al*., 2006), augment course content (Kindfield, 1994; Henderson, 1999) and yield desired outcomes such as metavisual cognition (Gilbert, 2005) as far as it is possible. Particular modes of visual representation have been shown to enhance different aspects of cognitive engagement and knowledge processing among learners. For example, interactive diagrams have been shown to promote understanding more so than animations (Kindfield, 1994; Tversky, 2003). Although there is no shortage of visual resources, extensive problems exist relating to students' ability to understand and use these effectively, critically and constructively (Mathewson, 2005).

1.8 Summary and conclusion

The selected literature reviewed in Chapter 1 informs this researcher that learning and teaching can be challenging processes. Learning comes with a range of difficulties which must be carefully assessed and understood by all researchers. The learning difficulties reported in this study (Chapters 4 to 8) can be best explained knowing the diversity and nature of difficulties that exist in allied disciplines of Science. The offer of remediation to assist with learning is vital and its success depends on the researcher's understanding of knowledge expression from different mental schemata. An appraisal of symbolism and the visual imagery related to Molecular Biology requires conceptualization (Hewson and Hewson, 1984; Hill *et al*., 2000), reasoning (Sins *et al*., 2005; Pata and Sarapuu, 2006) and an understanding (Shulman, 1986) of the mode or context of information presentation (Mathewson, 2005; Takayama, 2005; Schönborn and Anderson, 2006b). The correct expression of symbolic language characterizes "visual literacy" and the understanding of the underlying concepts conveyed by symbolism (Christopherson, 1997, cited by Takayama, 2005; Kozma, 2003; Ferk *et al*., 2003; Takayama, 2005). There is a critical need to identify and study the diverse forms of symbolism in Molecular Biology and the learning challenges they impose. This is the major focus of Chapter 2.

CHAPTER 2

THE NATURE AND DIVERSITY OF SYMBOLIC LANGUAGE USED IN MOLECULAR BIOLOGY

2.1 Introduction

As a working definition for this thesis, the author regards the term "symbol" as a form of shorthand notation, structure or entity representing particular information. The use of symbols to represent particular information is "symbolism (Salomon, 2005). Symbolic representations convey a wealth of information and should ideally serve as an aid to enhance the understanding of information (Treagust *et al*., 2003; Ferk *et al*., 2003; Kozma, 2003; Gilbert, 2005). Evidence does exist that diverse forms of symbolism have been used for information dissemination in a range of disciplines (1.6). They may be considered as convenient teaching tools which aid mental imagery and lessen the cognitive demands of a learning task (Kirschner, 2002; Kozma, 2003). Contrary to this view, researchers have reported on a range of learning difficulties (1.6) that students experience when presented with symbolic representations. Given that a diverse range of symbolism use also exists in Molecular Biology, as will be exposed in this Chapter, it is reasonable to expect that similar learning difficulties (1.6) might emerge when symbolic representations are used in the teaching of this discipline. Molecular Biology presents with thousands of entities affected by diverse symbolic representation (Conn *et al*., 1987; Kornberg and Baker, 1992; Stryer, 1995; Zubay *et al*., 1995; Horton *et al*., 1996; Karp, 1996; Weaver, 1999; Lodish *et al*., 2004; Cooper and Hausman, 2004; Voet and Voet, 2004; Elliot and Elliot, 2005; Garrett and Grisham, 2005). In this chapter, the author therefore exposes the nature and diversity of symbolism that could be used to teach Molecular Biology to undergraduate students. Based on teaching experience, the author also presents learning difficulties that the various forms of symbolism could present.

Single or multiple forms of symbolic representations are known to impact on visual cognition (Gilbert, 2005), the ability to interpret and communicate factual information on the same

entity or phenomenon (Kozma, 2003; Seufert, 2003). For example, Schönborn and Anderson (2006b) showed that a simple interaction involving homologous antigen and antibody may be illustrated using an electron micrograph, a graphical plot, a chromogenic ELISA test or a space-fill model of antigenic epitope and the fragment antigen binding domains of the antibody. Thus meaningful learning is essential for students to interpret multiple representations or symbolic forms of various single entities, processes or phenomena (Seufert, 2003; Tsui and Treagust, 2003). Learning from symbolic representations can be a cognitively demanding task (1.6) as the underlying meaning (Kozma, 2003; Gilbert, 2005) of depictions is generally poorly understood by some students. The depth of information conveyed by symbolic representations makes molecular visualization and the symbolic language of communication exceptionally complex (Christopherson, 1997; Treagust *et al*., 2003; Ferk *et al*., 2003; Mathewson, 2005; Takayama, 2005; Gilbert, 2005).

In addition, the construction of propositional knowledge (Chapter 3) by this researcher required a thorough understanding of the information represented by symbolism in Molecular Biology. This chapter presents a content analysis (3.3) of entities encountered in this study to illustrate how each is affected by the nature and diversity of symbolic language communication. Other forms of symbolism which appear in the syllabi presented to undergraduate students (3.1) are also described in this Chapter, exposing their meaning and use in Molecular Biology.

2.2 Content analysis briefing and aim

Hypothesis: A diverse range of symbolism does exist in Molecular Biology and the interpretation of the symbolic language or meaning associated with this form of representation may pose learning difficulties among students. Therefore, an appraisal of different categories of symbolism describing various examples of associated symbolic language is presented in this chapter, following an acceptable protocol (3.3) for such an analysis (Cohen *et al*., 2000). In addition, the exercise serves to address the primary research

question of this study, “What is the nature and diversity of the symbolic language used in Molecular Biology’? (See Introduction and aim of study).

The aim of the content analysis is to expose the nature of symbolic language in Molecular Biology so that similar representations could be analysed in terms of the cognitive and visual challenge (Kirshner, 2002; Kozma, 2003; Gilbert, 2005) they may impose to some students during learning. The offer of remedial assistance (Chapter 8) to such students is possible once the learning difficulties associated with symbolic language are carefully assessed and characterised as described (1.4).

2.3 Content analysis, selection of textbooks and documentation of supporting literature

The content analysis protocol (3.3) was performed on symbolism and symbolic language in a conveniently selected or limited sample (Staver and Lumpe, 1993) of twelve textbooks on Molecular Biology and Biochemistry (Table 2.1) that are currently used in South Africa. The study focused on the following selected entities: (i) single molecular components represented by symbols, (ii) genes and symbolic shorthand notations, (iii) oligonucleotide shorthand structural forms, (iv) plasmid nomenclature, structural maps and their functional domains, (v) structural, conformational and functional domains of DNA, (vi) nomenclature, structural and functional domains of RNA, (vii) the DNA replication process and structural intermediates and (viii) cloning and gene expression terminology.

The majority of the selected textbooks were also recommended to students at their time of publication and later use, for reading purposes and supplementation of lecture content where needed. The books were published between 1987 and 2005 (Table 2.1) and were written largely by American authors (Conn *et al*., 1987; Kornberg and Baker, 1992; Stryer, 1995; Zubay *et al*., 1995; Horton *et al*., 1996; Karp, 1996; Weaver, 1999; Lodish *et al*., 2004; Cooper and Hausman, 2004; Voet and Voet, 2004; Elliot and Elliot, 2005; Garrett and Grisham, 2005). The information in the selected books could be placed in various categories

and sub-categories, each representing the use of particular forms of symbolism to convey specific information in the field of Molecular Biology.

2.4 Prevalence of symbolism in the textbooks

The prevalence of each form of symbolism (Table 2.1) in the various textbooks is expressed as a percentage of those containing the symbolic form. In addition, supporting scientific literature is discussed to show usage of the various forms of symbolism especially those investigated in this study. Based on the teaching experience of the author, symbolism possibly contributing to learning difficulties or common misconceptions, as shown in this study (Chapters 4 to 8), was also recorded from the sources of information. Table 2.1 shows the prevalence or absence of some of the types of symbolism that are encountered in Molecular Biology. With regard to the use of symbols to represent single components, the symbol "σ" was commonly used to express a transcription initiation factor. On examination of the literature, the same symbol is used to describe a constituent of RNA polymerase (2.5.1) which is involved in transcription initiation, making it potentially confusing to the reader. Further, different types of transcription initiation factors are differentiated using superscripts, for example "σ^{70}", "σ^{54}" (Voet and Voet, 2004; Lodish *et al*., 2004), "σ^{gp28}" and "σ^{gp34}" (Voet and Voet, 2004). It is clear that single symbols can be used in different contexts and they pose a challenge in terms of differentiation and analysis of factual detail. In this study, the symbols "Δ" and "**::**" were investigated in specific contexts to do with gene deletion, transposition and in tandem gene fusions as shown (Chapter 3). However, their prevalence in modern textbooks was uncommon (Table 2.1) although there appears to sufficient alternate literature which illustrates their use in Molecular Biology (2.5.2). Hence, the dissemination of information on such uncommon symbols to students must ensure clear definition and use in specific contexts. The schematic presentation of oligonucleotides (Figure 3.1; Chapter 3) appeared in 11 of the 12 books used in the content analyses. Its defined format showing deoxyribose sugar, the phosphodiester link and synthesis intermediates characterised by the phosphite function (Chapter 3) is clearly defined in this study. Students are expected to provide the molecular detail that this symbolic representation

depicts (Chapter 4). The use of the symbolic "cccDNA" nomenclature to illustrate the relaxed covalently closed circular form of a plasmid could not be found in all 12 textbooks used in the content analysis. Clearly, the differentiation of conformational plasmid forms (2.5.4) is illustrated elsewhere using symbolic language such as "supercoiled" DNA, "cccDNA" for the relaxed covalently closed circular form, "ncDNA" for the topoisomerase I-nicked circular form or "ocDNA" to describe the open circular form (2.5.4; Kornberg and Baker, 1992; Turner *et al*., 1997; Virtual Lab Book website, 2005). Plasmid nomenclature was illustrated in 8 of the 12 textbooks while a functional domain such as the ori site showed 83% prevalence (Table 2.1). All 12 textbooks showed circular maps of plasmids; however, only three reflected on linear plasmid maps. In this study, plasmid conformation, linear and circular maps as well as functional plasmid domains (Chapter 5) receive special attention as forms of symbolism in Molecular Biology. Structural and symbolic forms of nucleic acid representation included most commonly the organic or molecular structures, schematic diagrams, line drawings and atomic space fill models, and to a lesser extent, symbolic forms illustrated using outputs from X-ray diffraction and crystallography (Table 2.1). Thus the identification of molecular detail in these multiple structural forms requires meaningful learning and interpretation as described (2.5.5; 2.5.6). Restriction maps of DNA illustrated the placement of restriction sites; using several phage ori sites (Kornberg and Baker, 1992), arbitrary fragments (Conn *et al*., 1987; Weaver, 1999;Voet and Voet, 2004; Garrett and Grisham, 2005), phage λ genome (Horton *et al*., 1996; Cooper and Hausman, 2004), polyoma virus genome (Karp, 1996) or SV40 DNA (Zubay *et al*., 1995; Stryer, 1995; Voet and Voet, 2004) but none of the 12 books showed an example of a plasmid restriction map as done in this study (Chapter 3; Figure 3.3). The functional DNA domain "TATA box" appeared in all 12 books compared with the "TATAAT" sequence which received coverage in 8 textbooks. The mechanism of DNA replication which is investigated in this study was illustrated largely by line drawings and 2-D diagrams to show sequential steps and labels of structural DNA features, *inter alia*, such as the replication bubble or eye, replication fork, Okazaki Fragment, nick and primer (Table 2.1). The "rolling circle" mechanism of replication was illustrated in 4 of the 12 books. A single textbook showed a 3-D diagram of DNA polymerase and its interaction with DNA (Horton *et al*., 1996). Illustrations of plastic model constructs of DNA replication intermediates as shown by Beltramini *et al*. (2006) did

Table 2.1 Prevalence (+) or absence (-) of symbolic language per total pages in selected Molecular Biology and Biochemistry textbooks

Selected entity influenced by symbolic language	Conn *et al.* (1987)	Kornberg and Baker (1992)	Stryer (1995)	Zubay *et al.* (1995)	Horton *et al.* (1996)	Karp (1996)	Weaver (1999)	Voet and Voet (2004)	Lodish *et al.* (2004)	Cooper and Hausman (2004)	Elliot and Elliot (2005)	Garrett and Grisham (2005)	Prevalence (%)
Book number	1	2	3	4	5	6	7	8	9	10	11	12	
Component labels e.g. Sigma Factor (σ)	+	+	+	+	+	+	+	+	+	+	+	+	100
Gene symbols													
Δ - deletion	-	+	-	-	-	-	-	-	-	-	+	-	17
:: - integration	-	-	-	-	-	-	-	-	-	-	-	-	0
- fusion	-	-	-	-	-	-	-	-	-	-	-	-	0
Oligonucleotides													
structure schematic e.g. Fig.3.1	+	+	+	+	+	+	+	+	+	+	-	+	92
shorthand e.g.dT_2	+	+	+	-	+	+	+	-	-	-	-	-	50
Plasmids													
nomenclature shorthand e.g. pBR322	+	+	+	+	+	-	+	+	-	-	-	+	67
conformation shorthand e.g.cccDNA	-	-	-	-	-	-	-	-	-	-	-	-	0
maps - linearized	-	-	-	-	+	-	+	-	-	-	-	+	25
- circular	+	+	+	+	+	+	+	+	+	+	+	+	100
functional domains shorthand e.g. ori site	+	+	+	+	+	-	+	+	+	+	-	+	83
DNA													
structure													
organic/molecular	+	+	+	+	+	-	+	+	+	+	+	+	92
line drawings	+	+	+	+	+	+	+	+	+	+	+	+	100
schematic	+	+	+	+	+	+	+	+	+	+	+	+	100

	1	2	3	4	5	6	7	8	9	10	11	12	
atomic space fill	+	+	+	+	+	+	+	+	+	+	+	+	100
3-D stick model	-	+	+	-	+	-	-	+	-	+	-	+	50
diffraction pattern	-	-	+	-	-	-	+	+	-	-	-	-	25
electron micrograph	-	+	+	+	+	+	+	+	+	-	+	+	83
restriction map	+	+	+	+	+	+	+	+	-	+	-	+	83
conformation atomic space fill	+	+	+	+	+	-	-	+	+	+	+	+	83
3-D stick model	+	-	+	+	+	-	-	+	-	-	-	+	50
functional domain shorthand													
e.g. "TATA box"	+	+	+	+	+	+	+	+	+	+	+	+	100
"TATAAT box"	+	+	+	+	+	+	-	+	-	-	+	-	67
RNA nomenclature													
prefix code and abbreviation	+	+	+	+	+	+	+	+	+	+	+	+	100
structure													
organic / molecular	+	-	+	+	+	+	+	+	+	+	+	+	92
schematic	+	+	+	+	+	+	+	+	+	+	+	+	100
atomic space fill	-	-	-	-	-	-	-	-	-	+	+	+	25
line drawings	+	+	+	+	+	+	+	+	+	+	+	+	100
X-ray crystallography	-	-	-	-	-	-	-	+	-	-	-	-	8
DNA replication representation													
Line drawings	+	+	+	+	+	+	+	+	+	+	+	+	100
2D model	+	+	+	+	+	+	+	+	+	+	+	+	100
3D model	-	-	-	-	+	-	-	-	-	-	-	-	8
Model - plastic constructs	-	-	-	-	-	-	-	-	-	-	-	-	0
structure													
replication eye / bubble	+	+	-	+	+	-	+	+	+	-	-	-	58
replication fork	+	+	+	+	+	+	+	+	+	+	+	+	100

	1	2	3	4	5	6	7	8	9	10	11	12	
rolling circle	-	+	-	-	-	-	-	+	+	+	-	-	33
Okazaki Fragment	+	+	+	+	+	+	+	+	+	+	+	+	100
nick	-	+	-	-	+	+	+	+	+	+	+	+	75
primer	+	+	+	+	+	+	+	+	+	+	+	+	100
Cloning													
Symbolic language													
vector	-	+	+	+	+	+	+	+	+	+	+	+	92
chimaeric molecule	-	+	+	-	-	-	-	+	+	+	-	+	50
sticky end	+	+	+	-	+	+	+	+	+	-	+	+	83
blunt end	+	+	+	-	+	-	+	+	+	-	+	+	75
downstream	+	-	-	-	+	-	+	+	+	+	-	+	58
adaptor	-	-	-	-	-	-	-	+	-	-	-	-	8
cassette	-	-	+	-	-	-	-	+	-	-	-	+	25
genomic library	+	+	+	+	+	+	+	+	+	+	+	+	100
Gene Expression													
Symbolic language													
probe	+	+	+	+	+	+	+	+	+	+	+	+	100
chromosome walking	-	+	+	+	+	+	-	+	-	-	-	-	50
Southern blot/transfer	+	+	+	+	-	+	+	+	+	+	+	+	92
sequencing	-	+	+	+	+	+	+	+	+	+	+	+	92
footprint	-	+	-	-	-	-	+	+	+	+	-	+	50
fingerprint	-	+	+	+	-	+	+	+	+	-	+	+	75
DNA chips	-		+	-	-	-	-	+	+	-	+	+	42
% Prevalence of selected entities	57	72	72	59	70	54	67	83	65	59	57	74	

not appear in any of the textbooks. With regard to the use of symbolic language, those chosen in the field of gene cloning and expression reflect a prevalence ranging from 8 to 100 % (Table 2.1). The terms "cassette" and "adaptor" may be judged as uncommon; however, there appears to be literature supporting their use in recombinant DNA technology (Table 2.1; Bahl *et al*., 1978; Barr *et al*., 1989; Horton *et al*., 1996).Of interest, the percentage representation of selected 54 symbolic entities in the university level books selected ranged from 57 to 83 (Table 2.1).

2.5 Content analysis of symbolism affecting entities of Molecular Biology

The content analysis yielded a diversity of symbolic representations describing components, properties of various entities, processes or dynamic changes affecting entities. These are described in detail to show their diversity, representation format and potential learning difficulties that they could impose to students.

2.5.1 Component labels and use of Greek symbolism

Symbols are a form of shorthand notation, convenient to use and easily indicated to represent a parameter or component in formulae or equations. In Biochemistry or Molecular Biology, symbols such as Greek letters are commonly used to name components of enzymes or polypeptides making up a protein. For example, subunits of DNA polymerase III holoenzyme are designated τ, ɵ, α, β, γ, χ , ψ ε, δ and δ′ (Kornberg and Baker, 1992). RNA polymerase from *E.coli* has five subunits of four types, namely α_2, β, β′ and σ (Kornberg and Baker, 1992; Horton *et al*., 1996). Greek letter equivalents are also assigned to the nomenclature of immunoglobulin chain components. The heavy chains of immunoglobulins IgM, IgG, IgA, IgD and IgE are named μ, γ, α, δ and ε, respectively. There are two types of immunoglobulin light chains, designated κ and λ. The chain constitution for immunoglobulins varies, for example, the pentameric form of IgM is indicated as $(\mu_2\kappa_2)_5$ or $(\mu_2\lambda_2)_5$ and that of IgG is $\gamma_2\kappa_2$ or $\gamma_2\lambda_2$ (Klein, 1990). Sometimes the choice of symbol bears

significance in a particular nomenclature. For example, a DNA replication intermediate has been isolated from *E.coli*, analysed by autoradiography, to show a closed circle with an internal loop (Cairns, 1963a). Such forms of DNA are referred to as theta structures because of their resemblance to the Greek letter, ɵ (Cairns, 1963a; cited by Stryer, 1995). Kornberg and Baker (1992) refer to sigma (σ) DNA, a lariat-shaped DNA intermediate encountered in the rolling circle mechanism of replication (3.5.8). Sigma factor (σ) and Rho protein (ρ), play a role in the initiation and termination of prokaryotic transcription, respectively (Losick and Pero, 1981; Kornberg and Baker, 1992) while other protein factors such as EF-Tu and EFG are implicated in protein synthesis. EF-1α and EF-1β are other examples of symbol use to describe protein factors involved in protein synthesis of eukaryotic cells (Zubay *et al*., 1995; Stryer, 1995; Horton *et al*., 1996). Since symbols are polysemic, they provide a challenge for students to differentiate in terms of their use and meaning in different contexts. It could be a daunting task for students to reflect on a range of information that accompanies the use of a single symbol in different contexts. For example, common Greek symbols, α and β, have been used to name molecular components such enzyme subunits or polypeptide components, carbon atoms in unsaturated fatty acids (Nes and McKean, 1977) and glycosidic bonds (Horton *et al*., 1996; Elliot and Elliot, 2005; Garrett and Grisham, 2005).

2.5.2 Gene symbol representation

Gene symbols are designated by three italicized letters, usually an abbreviation for a pathway or phenotype, in accordance with the nomenclature system proposed by Demerec *et al*. (1966) who did pioneering work on genetic nomenclature at Cold Spring Harbor laboratories, New York. Dominant and recessive genes are denoted by italicized upper and lower case letters respectively, followed by the allele number, gene designation if necessary which may take the form of a single capital letter, followed by a hyphen and then the locus number. A plus sign or capital letter may follow the locus number to indicate wild type gene or gene cluster, respectively (Demerec *et al*., 1966; cited by Sherman and Lawrence, 1974; Malloy, 2001). Phenotypic gene markers may begin with a capital, italicized letter, followed by two lower case italicized letters and superscripts such as "+", "-", "s" or "r" to indicate

prototrophy or expression of a trait, auxotrophy or negative expression of a trait, sensitivity or resistance, respectively. In this study, phenotypic gene markers, associated with the somatic hybridoma fusion technology, are indicated as $Hgprt^{+}$ Tk^{-}, $Hgprt^{-}$ Tk^{+} or $Hgprt^{+}$ Tk^{+} to denote production (+) or non production (-) of the enzymes hypoxanthine guanine phosphoribosyltransferase and thymidine kinase, respectively (Köhler and Milstein, 1975; Staines, 1983; Chapter 3). Phenotypic plasmid markers (Chapter 3), used in this study, include *Lac(z),* Amp^{r}, Amp^{s}, Tet^{r} and Tet^{s} to represent β-galactosidase production, resistance to ampicillin, sensitivity to ampicillin, resistance to tetracycline and sensitivity to tetracycline, respectively in accordance with the gene designations provided by Demerec *et al*. (1966; 1968).

Peculiar traits will require special definition of selected symbols in accordance with recognized gene bank websites or journal nomenclature guidelines (Malloy, 2001). In human genetics, the pattern of gene symbols can vary. For example, ABO denotes the gene that controls ABO blood type and F8C denotes the gene that controls the ability to express factor 8 which is involved in blood clotting (Kinnear, 1992). These symbols do not indicate the specific chromosomal locus associated with a specific function. Alleles are represented by symbols such as variants of a letter, namely B or b, or superscripts to one letter, for example, 1^{A} or 1^{B}. When the allelic form of factor 8 is expressed by F8CH, normal clotting results whereas haemophilia results when the allelic form of factor 8 is not present or expressed by F8Ch (Kinnear, 1992). As can be seen from the symbolism associated with gene nomenclature, students are exposed to a wealth of descriptive information which can be difficult to learn as they are also presented in rather diverse formats.

In this study, the use of "Δ" and "**::**" as symbols representing gene deletion and gene insertion or fusion, respectively receives particular attention (Chapter 3). According to Jon Beckwith, pioneering genetic engineer, of the Harvard Medical School, USA, it is reasonable to assume that the symbol Δ was first chosen to represent gene deletion because "it is simply the Greek delta for D for deletion!" (Beckwith, 2005, pers.comm.). Should delta precede a gene written in lower case italics, this implies gene deletion. For example, Δ *tok1* and Δ *chs 1* are gene markers or symbols showing gene deletion which adversely affects potassium ion

channel function in *Saccharomyces cerevisiae* (Ahmed *et al*., 1999) and chitin synthase I activity in *Kluyveromyces lactis* (Jablonowski *et al*., 2001), respectively. Gene deletions may be induced by UV light or use of a chemical mutagen such as N-methyl-N´-nitro-N-nitrosoguanidine, resulting in lesions of various kinds which affect reading frames, replication and transcription (Turner *et al*., 1997).

The first use of "Δ" and "**::**" as symbols representing gene deletion and gene insertion, respectively cannot be confirmed. Thomas Silhavy, leading biotechnologist and genetic engineer at Princeton University, USA, indicates that it might be at Cold Spring Harbour in the laboratory of Milislav Demerec, where early work, post World War II, was done on bacterial and fruit fly mutations and "shifting genes" (now known as transposons) in maize (McClintock, 1951, 1956; Demerec *et al*., 1966; Silhavy, 2005, pers.comm.). In support of this fact, the National Centre for Biotechnology Education, UK, posts " The Transformer Protocol" (2001) Version 1, Teachers' Guide to symbol use, in which reference is made to the Demerec System of Symbol Use (Demerec *et al*., 1966; 1968) that includes both Δ and **::** as indicators of gene deletion and gene insertion, respectively. The Demerec System of Symbol Use, concerning "Δ" and "**::**" as indicated, is also adopted by several gene banks or genetic stock centres, for example, the *E.coli* Genetic Stock Centre, *Salmonella* Genetic Stock Centre and *Drosophila* Flybase, websites of which may be found on the World Wide Web.

The symbol "**::**" was originally used to indicate gene insertion (McClintock, 1951, 1956) and currently remains the accepted way to show insertion of one gene within the domain of another, resulting in the inactivation of the disrupted gene. For example, Lambrechts *et al*. (1996) describe the "ligation" of the *URA3* (uracil biosynthesis gene) into the domain of the *MUC1* mucin biosynthesis gene, resulting in the construct *muc1***::***URA3*. The insertion results in the inactivation of mucin-encoding gene. Similarly, Kadonaga *et al*. (1987) describe the insertion of transposable element Tn5 within the domain of the β-galactosidase-encoding gene, *Lac(z)* to result in *lac(z)***::***Tn*5, thereby destroying *lac(z)*'s ability to express β-galactosidase.

Unfortunately, there is currently a lack of standardization in the choice of symbol to indicate in tandem gene fusions (Chapter 3) which generate an open reading frame that can be transcribed to mRNA. The single translation product would express the functions of both fused genes. Crenshaw III *et al*. (1987) describe the expression of a metallothionein-calcitonin fusion gene (MT-CAL) in transgenic mice, yielding a single product with dual function or properties of both metallothionein and calcitonin. This gene fusion is indicated by a hyphen as shown above. The journal, "The Plant Cell" prescribes in its instructions to authors that a single colon or hyphen be used to indicate gene fusion (Website in list of references). The FlyBase Reference Manual C (2005) and Gene Proforma Version 33 database (2003) recommend the use of the double colon to indicate in tandem gene fusion and that this be clearly defined as a fusion which generates an open reading frame, leading to the transcription and translation of a single translation product expressing the functions of both fused genes (Websites in list of references). Verstrepen *et al*. (2001) describe a fusion between the heat shock promoter, HSP30p and flocculation gene *FLO1* to result in the construct, HSP30p**::***FLO1* where *FLO1* is brought under transcriptional control of the HSP30p promoter.

The lack of standardization of symbol use can be potentially confusing to students who read on these from different sources of literature. Some students can show a tendency to reflect superficially and incorrectly (Chi *et al*., 1981; Schönborn *et al*., 2002a) on information which appears in various sources of literature where symbol format is applied differently. It is imperative that very specific definition be emphasized by the teacher (Duchovic, 1998; Darby, 2005) when using gene symbols in a particular context. Unfortunately, experts present a range of symbolism in Biochemistry textbooks (Table 2.1) which may lack standardization in presentation format. Such symbolism is interpreted with relative ease by experts who are familiar with their own creation of symbolic forms and the underlying information which accompanies symbolic or iconic representations. The understanding of the underlying information depicted by symbolism is of paramount importance as learners with poor mental schemata (Nurrenbern, 2001) are prone to misinterpret or fail to analyse certain forms of symbolism (Bechtel, 1998; Kozma, 2003; Bodemer *et al*., 2004). This difficulty

characterizes “intrinsic cognitive load” where prior learning is inadequate or absent to facilitate the understanding of the external representations (Sweller and Chandler, 1994).

2.5.3 0ligonucleotide representation

The synthesis of oligonucleotides, via the phosphite triester approach, exposes students to specialized Chemistry (Letsinger *et al*., 1975; Letsinger and Lunsford, 1976; Tanaka and Letsinger, 1982) and symbolic forms representing structural detail of these molecules. A solid support such as silica gel 60 may be subject to derivatization in order to synthesize a ligand (Figure 3.1; Chapter 3), from which the oligonucleotides could be synthesized (Letsinger *et al*., 1975; Letsinger and Lunsford, 1976; Tanaka and Letsinger, 1982). The synthesis generates oligonucleotide intermediates which could be shown by a schematic representation (Figure 3.1; Chapter 3), commonly used in Biochemistry textbooks (Conn *et al*., 1987; Stryer, 1995; Kornberg and Baker, 1992; Weaver, 1999; Garrett and Grisham, 1995; 2005). In an older textbook, White *et al*. (1959) used similar schematic representations of oligonucleotides, minus the solid support and ligand, to show cleavage points by bacterial nucleases and a snake venom diesterase. With reference to Figure 3.1, the vertical line leading to either a purine or pyrimidine base, thymine in this example, represents the β-(D)-2- deoxy-ribose sugar, with C-1′ and C-5′ at the top and bottom, respectively (Conn *et al*., 1987). “P” denotes the phosphorus atom in the trivalent (A) and pentavalent (B) form. In diagram C, “P” represents the 3,5 phosphodiester link between the nucleotides, following chemical treatment as follows. The oxidation of the trivalent phosphite form of the oligonucleotide (A) to the phosphotriester form in which phosphorus is pentavalent (B) may be achieved by using dilute I_2 in a tetrahydrofuran-pyridine-H_2O solution (Letsinger *et al*., 1975; Letsinger and Lunsford, 1976; Pon, 1987). The removal of both the methyl protecting group on the 3’ phosphoryl oxygen and diparamethoxytrityl (DPMT) protection group, as well as cleavage of the oligonucleotide from the solid support, can be achieved by various chemical methods (Letsinger *et al*.,1975; Letsinger and Lunsford, 1976; Tanaka and Letsinger, 1982). Diagram C (Figure 3.1; Chapter 3) presents a schematic representation of the 3,5 phosphodiester bond in which phosphorus is pentavalent. From the above, the

differentiation of phosphite and phosphodiester groups can be clearly done on the basis of bonding structure or valency (Chapter 3).

Oligonucleotides may constitute probes or primer fragments. These have been shown using schematic line drawings (Horton *et al*., 1996; Weaver, 1999) which may include either a phosphate or hydroxyl group at the 3′ or 5′ ends. This symbolic form represents the sugar-phosphate backbone of nucleotides linked by the 3,5 phosphodiester bond. The shorthand notation of the structural form of oligonucleotides is also a form of symbolic language expressing more descriptive knowledge in an abridged format, for example dT_2 or $^{5'}TpT^{3'}$ (Table 2.1). This denotes a dimeric deoxyoligonucleotide comprising of two units of thymidine linked by a 3,5 phosphodiester bond (Gupthar, 1983; Kornberg and Baker, 1992).

Alternate shorthand notation of molecular structure can be potentially confusing to students who are unfamiliar with different symbolic formats representing such a structure (Ferk *et al*., 2003). Novices present with great difficulty trying to learn from such an array of symbolic representation. Schönborn and Anderson (2006b) draw attention to the fact that the disulphide link in certain protein structures can be seen as " –S-S-" (Stryer, 1995), as a yellow bar (Garrett and Grisham, 1995) or straight black line (Bohinski, 1987). The lack of standardization in such a depiction or "visual language" poses challenges to learners who seek information from different Biochemistry textbooks. In a recent textbook by Garrett and Grisham (2005) phosphodiester bonds between RNA nucleotides are indicated in some diagrams in expanded structural form to show that the phosphorous atom is pentavalent yet other diagrams show this group by an encircled "P" with two bond linkages between nucleotides, deceptively indicating to a learner at first glance that the valency could also be two. Variation in the illustration of external representations is not always interpreted correctly by certain learners.

2.5.4 Plasmid representation

Plasmid nomenclature presents a form of symbolism as hundreds of plasmids are each named differently. An example of a plasmid is pBR 322 wherein the letter "p" denotes a plasmid, "BR" refers to the eponymous naming after the plasmid engineers, Bolivar and Rodrigues who differentiated "322" from other constructs "325", "327" and "329" in their laboratory (Bolivar *et al*., 1977). Messing and Vieira (1982) developed different plasmids from the phage M13 genome and named these M13mp8 and M13 mp9. Yeast plasmids may fit one of the following categories, namely YIp, YEp,YCp, YRp or YAC (Gardner *et al*., 1984; Brown, 1986; Stryer, 1995; Horton *et al*., 1996; Elliot and Elliot, 2005; Garrett and Grisham, 2005) . These symbols convey a wealth of information in terms of the plasmid properties (Barr *et al*., 1989). YIp vectors are termed "yeast integration plasmids" as they integrate genes into the genome by homologous recombination. The symbolic nomenclature of the others also indicates important information. YEp vectors are "yeast episomal plasmids" capable of autonomous replication because they possess the 2μ ori site derived from the *S.cerevisiae* 2μ plasmid (Armstrong *et al*., 1989). YRp vectors are "yeast replication plasmids" which produce high copy numbers as an ARS sequence in their construct ensures autonomous replication. YCp vectors are "yeast centromere plasmids" which contain a centromere sequence that promotes stability of the plasmid during mitosis. YACs are "yeast artificial chromosomes" or plasmids which incorporate into the nucleus as a chromosome owing to the presence of the centromere and telomere sequences in their construct. These four types of yeast vectors may be termed "shuttle vectors" (Gardner *et al*., 1984; Brown, 1986) as they may be transferred routinely from yeast to bacterial cells (*E.coli*) provided they contain an appropriate origin of replication which is active in both microbes. Symbolic forms of plasmid nomenclature actually convey important information on their construction and biological properties, posing potential learning difficulties to students who might be exposed to varied information on plasmids of different origin. Such variation can be rather subtle, making the differentiation of plasmids somewhat difficult and confusing to the learner.

DNA plasmids from microbial sources are largely double-stranded and circular. Various shorthand notations are used to describe the conformational form of plasmids which can be

difficult for students to visualize or differentiate. In bacteria, they are found be supercoiled in the native state but can be isolated as a nicked linear form, relaxed covalently closed circular (cccDNA) form, topoisomerase I-nicked circular form (ncDNA) or open circular form (ocDNA, nicked on one strand) or as plasmid multimers due to joining of two or more molecules. In bacteria, an enzyme such as the gyrase (topoisomerase II) is known to twist plasmid and genomic DNA, inducing superhelical tension which assists in processes such as replication and transcription (Weaver, 1999). Topoisomerase I nicks one strand of superhelical DNA so that DNA polymerase has access to the DNA for replication (Kornberg and Baker, 1992; Turner *et al*., 1997; Virtual Lab Book website, 2005). An understanding of the molecular information which differentiates each conformational or symbolic form certainly enhances the students' ability to visualize the structural features of plasmids and explain the representation of symbolic language (Henderson, 1999; Schönborn and Anderson, 2006a) such as "cccDNA", "ncDNA", "ocDNA" or "supercoiled" DNA.

Circular double-stranded expression plasmids may be restricted to a linear form as shown by plasmid maps (Figure 3.2). Functional domains of plasmids are generally indicated in shorthand notation to fit clearly demarcated regions within the plasmid structure. These regions may be shaded in different colours (Zubay *et al*.,1995; Stryer, 1995; Horton *et al*., 1996; Garrett and Grisham, 2005) and labelled in shorthand notation to show, *inter alia*, the regulatory sequences, "o", the operator and "p", the promoter, "ts", the terminator sequences, "ori site", the origin of replication, "ARS", the autonomous replicating sequence, "*LEU* ", a dominant leucine biosynthesis marker gene or "*Amp*^{r}", an ampicillin resistance marker gene (Chapter 3). Also indicated are restriction sites, for example, *Pvu*II, *Bam*HI, *Hin*dIII or *Pst*I which can be cleaved by site-specific endonucleases that bear the same names as the respective site they cleave. The restriction enzymes *Bam*HI, *Hin*dIII and *E.co*RI are named in shorthand italicized form to reflect the genus and species of bacterial strains from which they were first isolated (Horton *et al*., 1996). These were *Bacillus amyloliquefaciens, Haemophilus influenzae and Escherichia coli*, respectively. Over 650 of such enzymes (Boehringer-Mannheim, FRG: 1987 Chart of recognition sequences of restriction endonucleases and methylases) are named from different bacterial isolates and each carries a shorthand name in italics to reflect the genus and species, followed by letters which indicate

strain and / or Roman numeral(s) to indicate the order of discovery of enzyme in that strain (Horton *et al*., 1996, Table 20.3 p.582). It is imperative that this form of symbolic representation be understood as functional domains of a plasmid (Gardner *et al*., 1984; Brown, 1986; Kornberg and Baker, 1992; Weaver, 1999) can influence the mechanics of gene cloning, analysis of gene expression and transformant selection in particular ways (Chapter 3). Symbolism, associated with structural form and functional domains, requires learning of rather specific information associated with each. An appraisal of various structural forms might be demanding in terms of visual cognition (Mathewson, 2005; Gilbert, 2005) where learners find difficulty in associating the cognate symbolic depiction and descriptions of the plasmid conformation. Likewise a reflection on a variety of functional domains can be an imposing task for some learners.

2.5.5 DNA representation

Double-stranded DNA (Watson and Crick, 1953) is composed of two anti-parallel strands which are complementary in terms of base sequence and run 5→3′ in opposite directions. The structure of double-stranded DNA may be illustrated in numerous ways. Simple line drawings, resembling a ladder, actually depict the two anti-parallel strands of sugar-phosphate backbone and, the rungs represent complementary base-pairing by hydrogen bonding (Conn *et al*., 1987). Schematic computer-generated diagrams of double-stranded DNA may take different formats. For example, shaded thick lines may be used to show helical twists of the sugar-phosphate backbone, coupled with differently coloured rectangular strips to represent each base, the complementary base-pairing shown by interlocking of these rectangular strips and stacking of each base pair within the helix (Horton *et al*., 1996). Others use lines and letters to show linkage of the sugar (S)-phosphate (P) backbone and the respective aromatic bases adenine (A), thymine (T), guanine (G) and cytosine (C) to “S”. Using lines, two and three hydrogen bonds respectively are also shown between complementary bases as follows, A═T and G≡C (Elliot and Elliot, 2005). The 5 to 3′ direction of the two strands in this type of representation is shown labelled at the ends of the sugar-phosphate backbone (Elliot and Elliot, 2005) or using arrows in opposite directions of

the “S-P” (sugar-phosphate) backbone (Conn *et al*., 1987). Diagrams are certainly a form of static symbolism which may carry a wealth of information as the “external representation” to the learner. They may depict model constructs which are useful when learners not only learn to interpret such external representations but reflect meaningfully on scientific information that has been learnt on the subject (Cheng *et al*., 2001; Seufert, 2003; Beltramini *et al*., 2006; Cook *et al*., 2006) and then attempt to correlate this information with the static visual representations (Kozma, 2003). The unaided exercise, entailing the coupling of these initiatives, may be regarded as “bootstrapping” (Cheng *et al*., 2001). Unfortunately, different diagrams illustrating the same entity or subject may be constructed in various ways (Schönborn and Anderson, 2006b), posing learning challenges with visual competence (Takayama, 2005) and the understanding of varied symbolism (Wheeler and Hill, 1990; Henderson, 1999; Kozma, 2003; Schönborn and Anderson, 2006b). This imposes “extraneous cognitive load” to the learner (Scaife and Rogers, 1996; Lewalter, 2003; Chandler, 2004; Lowe, 2004). Some learners present with “germane cognitive load”, which is generated by uneasy and complex processing of knowledge in the LTM (1.3) (Carey, 1986; Chandler, 2004). The mode of presentation of such “external representations” should promote conceptual reasoning and expression of accurate detail about it (Schönborn and Anderson, 2006b) and not be overwhelming or too complex (Lowe, 1988; 2003; 2004). Diagram properties to do with shape, colour, size, labels, captions, depth of information, spatial arrangement or artistic features tend to make diagrams rather complex entities or “tools to think with” (Kindfield, 1994). The visual cue that diagrams provide must be accurate, unambiguous and should correctly facilitate reasoning and expression of the symbolic representation (Wheeler and Hill, 1990; Kindfield, 1994; Bodemer *et al*., 2004; Takayama, 2005). Illustrations of atomic space-fill models, 3-D stick models and X-ray diffraction patterns show the spatial arrangement of atoms (Seeman, 2004). Conformational variants of DNA, for example A-, B- and Z-DNA and differentiating features of the helical twist in the respective forms, are commonly shown using atomic space fill (Horton *et al*., 1996; Garrett and Grisham, 2005; Elliot and Elliot, 2005), 2-D schematic line and 3-D stick models (Richardson and Richardson, 2002; Patrick *et al*., 2005; Garrett and Grisham, 2005; Cook *et al*., 2006). A more realistic depiction of structural form is an electron micrograph of DNA (Horton *et al*., 1996); however, this shows a magnified visual of the physical form

rather than the individual atoms and their spatial arrangement which constitutes the entire macromolecule. The visual form of an atomic space fill model is a symbolic representation of a realistic entity. It requires learning aptitude and skill to interpret a "visuo-spatial" feature (Lord, 1985; Wu and Shah, 2004; Seeman, 2004) or structural nucleoprotein motifs (Garrett and Grisham, 2005; Elliot and Elliot, 2005; Takayama, 2005) such as the "Zinc Finger" or "Leucine Zipper" (Garrett and Grisham, 2005; Elliot and Elliot, 2005). An understanding of the molecular information (Treagust *et al*., 2003; Schönborn and Anderson, 2006a) depicted by a "finger" or "zip"-like feature is vital to the interpretation of models depicting such motifs. Molecular organic structures of DNA tend to show intricate features such as the bonding of atoms, the planar ring form of purine and pyrimidine bases, base pairing, structural features of deoxyribose sugar and the 3,5 phosphodiester link between nucleotides. This type of information can be difficult to learn but must accompany interpretations of all symbolic representations of DNA which may be characterised by an assembly of plastic beads, synthetic balls and sticks, lines or plastic components resembling the geometric shape of the furan deoxyribose ring, six-sided pyrimidine ring or five-sided imidazole component of a purine double ring structure (Rotbain *et al*., 2006; Beltramini *et al*., 2006).

Functional domains of double-stranded DNA may also be affected by the use of symbolic language. The two strands are labelled either "coding" or "template", depending on their respective function (3.5.7; Figure 3.4). Transcription initiation sites, generally rich in adenine and thymine bases, can be found on both the coding and template strands of eukaryotic and prokaryotic DNA and are referred to as "TATA box" or Hogness box (also called Hogness-Goldberg box, after the discoverers) (Lifton *et al*., 1978) and "TATAAT box" or Pribnow box (also called Pribnow-Schaller box, after the discoverers), respectively (Pribnow, 1975; Schaller *et al*., 1975). "TATA" and "TATAAT" constitute the 5→3′ consensus base sequence where initiation factors bind to initiate the transcription process. These sequences lie "upstream" of the site where transcription begins (position 1) and are given base or nucleotide positions or domains, for example -24 to -35 would be a typical "TATA box" domain and -10 to -35, a typical "TATAAT box" domain (Gardner *et al*., 1984; Horton *et al*., 1996; Karp, 1996; Garrett and Grisham, 2005). Symbolically, "box" could imply "confined area" or domain where initiation factors assemble together with the DNA dependent-RNA

polymerase. This enzyme "depends" on the template strand sequence to synthesize messenger RNA that is complementary in sequence to that of the template DNA strand but similar to that comprising the DNA coding strand (Figure 3.4). Symbolism may be presented in a static mode as seen printed on a page or physically constructed such as a ball-and-stick or atomic space-fill model which depicts the molecular arrangement of atoms. Alternatively, dynamic animations may be used to follow molecular processes or movement of submicroscopic elements to enhance the process of visualization (Hegarty, 2004; Ploetzner and Lowe, 2004). RasMol, Chime, UMass OWL and MolView are some tradename examples of an extensive range of multimedia packages currently used as teaching supplements in Molecular Biology or Biochemistry. CD-ROMS are also currently supplied as textbook supplements to facilitate visualization of the molecular and submicroscopic element (Rieber *et al*., 2004; Ploetzner and Lowe, 2004; Sins *et al*., 2005).

2.5.6 RNA representation

RNA molecules are differentiated into diverse types, namely; messenger, transfer, ribosomal, small nuclear, small interfering, non coding, small temporal, small nucleolar or micro RNA. Using a coded prefix, the symbolic format concerning nomenclature of these RNA types may be designated mRNA, tRNA, rRNA, snRNA, siRNA, ncRNA, stRNA, snoRNA and miRNA, respectively (Garrett and Grisham, 2005). The coded prefix gives some indication of function or location of specific RNA molecules. For example, tRNA transfers or carries specific amino acids to the ribosomes for protein synthesis while mRNA triplet codons determine the amino acid sequence and composition in a polypeptide, following interaction with the anticodon sequence of tRNA. The double-stranded small interfering RNA is known to induce translation silencing of mRNA in combination with the Argonaute protein complex to form the RNA-induced silencing complex (RISC). RISC facilitates unwinding of siRNA and the degradation of mRNA at midpoint of the mRNA- antisense siRNA strand complex (Zamore, 2001; Ariatti, 2007, pers.comm.). Acccording to ancient Greek mythology, "an Argonaut" was the term used to describe heroic sailors who achieved great feats (Wikipedia Encyclopedia, 2007a, website in list of references). Symbolically, could the Argonaute

protein-assisted silencing mechanism save a cell by suppressing the expression of certain protein such as those expressed by viruses in infected cells? There is currently much evidence to support this view (Dykxhoorn and Lieberman, 2006; Ariatti, 2007, pers.comm.), disregarding the author's personal interpretation of the symbolism associated with "Argonaute". The nomenclature, small nucleolar RNA (snoRNA) and ribosomal (rRNA) indicate the presence of these species in the nucleolus and ribosome, respectively. The structure of RNA molecules may be indicated using atomic space fill 3-D models, schematic diagrams, line drawings, ribonucleotide-linked shorthand notation resembling the format shown in Figure 3.1, molecular organic structures, X-ray crystallography for 3-D structures or conformational energy analysis and programmes such as RNAdraw and mfold for secondary structures showing loops and stems in convoluted rRNA or "clover"-shaped tRNA molecules, "hammerhead" or "hairpin"-shaped RNA molecules with catalytic activity (ribozymes) (Horton *et al*., 1996; Du, 1999; Elliot and Elliot, 2005; Garrett and Grisham, 2005). The functional domains of RNA molecules are also characterised by the use of unique symbolism. For example, mRNA molecules have a cap structure which promotes its binding to the ribosomes. The cap may be illustrated as " $^{m7}G^{5'}ppp^{5'}A^{2'OCH}{}_3pN$-" which denotes the nucleotide 7-methylguanosine 5′→5′triphosphate-linked to 2′-O – methyladenosine followed by a 3,5 phosphodiester- linked nucleotide (A, U, G or C) (Horton *et al*., 1996). "AUG" or "GUG" are generally the initiator triplet codons which interact with the anticodon of tRNA carrying formylmethionine or methionine to the ribosome for protein synthesis. Precursor mRNA transcripts show alternate domains of intervening sequences or introns which are excised during splicing as exons are linked to generate the translatable cistronic region comprising triplet codons which are specific for anticodon base pairing. Diagrams of precursor mRNA molecules generally show "introns" and "exons" as labels or differentiate these using alternate colours to fit a line drawing (Stryer, 1995; Horton *et al*., 1996; Karp, 1996; Garrett and Grisham, 2005; Elliot and Elliot, 2005). Position 3 or the "wobble position" of the triplet codons (running 5 to 3′) of the transcript tends to show variation in the type of base found, making the code "degenerate" in terms of "coding fidelity" and variation of amino acid composition in a peptide. The "stop codons", UGA (opal), UAG (amber) and UAA (ochre) signal the stop in the translation process. The codon UAG was named after Harris Bernstein ("Bernstein" meaning amber in German), a friend of the codon discoverers

Dick Epstein and Charles Steinberg (Wikipedia Enclopedia, 2007b, website in list of references). The other two alternate stop codon names "opal" and "ochre" simply follow a colour theme approach. The poly (A) tail on 3′ end of mRNA is so-called because it comprises at least 150-200 units of adenosine which promotes mRNA migration from the nucleus to cytoplasm (Conn *et al*., 1987; Stryer, 1995; Horton *et al*., 1996; Karp, 1996; Garrett and Grisham, 2005; Elliot and Elliot, 2005). Functional domains of tRNA are termed dihydrouracil loop, TΨC loop, anticodon loop and "CCA acceptor" stem. The symbolic language or nomenclature associated with these indicate the presence of unique bases such as dihydrouracil in one loop; bases thymine, pseudouracil (Ψ) and cytosine in that conserved sequence in one loop; the anticodons found at the base of one loop and the bases, cytosine (x2) and adenine making up the CCA acceptor stem which binds amino acids to the tRNA after an activation step. From the above, it is clear that a range of symbolic language is based on the functional and structural domains of RNA molecules. Learning of the underlying information depicted by symbolic representations of RNA could be challenging to students as these molecules differ in various structural forms and are associated with a range of molecular processes.

2.5.7 Representation of selected molecular processes

Symbolism and visual language also affect molecular processes such as replication or gene cloning which are investigated in this study (Chapter 3) and therefore given special attention in this survey.

2.5.7.1 Replication

Replicative DNA synthesis is a complex process (Chapter 3) which yields many structural changes to DNA. Visual representations of such structural features can be found in line diagrams (Bohinski,1987; Kornberg and Baker, 1992; Novick, 1998; Fossey and Hancock, 2005), 2-D and 3-D models (Rotbain *et al*., 2006; Cook *et al*., 2006; Beltramini *et al*., 2006)

and scientific programmes illustrated by video or software packages (Cook *et al*., 2006; Beltramini *et al*., 2006) to promote an understanding (Ametller and Pinto, 2002; Ferk *et al*., 2003; Fossey and Hancock, 2005) of the process. It is clear that various depictions of structural DNA intermediates actually constitute a form of symbolism and an explanation of the visual form of each derivative or intermediate (Losick and Shapiro, 1998; Lemon and Grossman, 1998) is in fact the "visual language" that one associates with the representation (Ferk *et al*., 2003; Yore and Treagust, 2006). In this study, emphasis is placed on two replication models, namely, the *E. coli* replication "bubble" model (Cairns, 1963a;b), adapted and presented in diagram format by Bohinski (1987) and the ΦX174 double-stranded RF I→ RF II DNA replication model (RF denoting "replicative form"), adapted from Goulian and Kornberg (1967), Dressler and Wolfson (1967) and Gilbert and Dressler (1968). This model is illustrated (Gupthar, unpublished; Chapter 3) using the "rolling circle" mechanism and guidelines from Kornberg and Baker (1992). The replication "bubble" model presents with structural features such the "eye" or "bubble", an unravelled form of duplex DNA at the site where replication begins, namely the "origin" or ori site (Oka *et al*., 1980; Funnell *et al*., 1987). It also presents the "replication fork", a bifurcated point in the structure of DNA which is unwound by a helicase, extending the "template" strands of DNA for bi-directional synthesis (Baker *et al*., 1986) that takes place 5 to 3′ in a continuous and discontinuous manner (Chapter 3). RNA strands act as "primers" for DNA synthesis and the insertion of primers or priming process is facilitated by a complex of proteins making up the "primosome" complex (Funnell *et al*., 1987). Discontinuous DNA fragments or "Okazaki Fragments" are named after the discoverer, Reiji Okazaki (Okazaki *et al*., 1968). These are separated by "nicks". "Gaps" result when RNA primers may be removed by the exonuclease action of pol I (Chapter 3). This form of nomenclature or symbolic language accompanies transition features characterising DNA replication. The ΦX174 double-stranded RF I→ RF II DNA replication model presents the "rolling circle" mechanism which depicts the unwinding of the "+" strand template from the "–" strand by Rep protein. The endonuclease "gpA protein" is so-named as it facilitates endonucleolytic cleavage between GA bases at positions 4305 and 4306 respectively within the ori site (van Mansfield *et al*., 1986). The "gpA" endonuclease may differentiated using this symbolic format from the shorthand notation for the dinucleotide, "GpA". Replication presents a wide range of symbolism, including

diagrams which indicate sequential steps of the process. These could pose learning difficulties in terms of nomenclature, an understanding of the structure and function of DNA replication intermediates and complex mechanisms associated with the process itself (Chapter 3).

2.5.7.2 Cloning and gene expression

Diverse symbolic language finds application in cloning and the analyses of gene expression. Plasmids which "carry" genes of interest are termed "vectors". This bears similarity to the definition given to "carriers of disease or infection" (The Concise Oxford Dictionary, 1982) The ligation of a gene to a vector generates a recombinant vector or chimaeric (also spelt chimeric) molecule (The Concise Oxford Dictionary, 1982; Garrett and Grisham, 2005). Symbolically, a "chimera" in Greek mythology represented a hybrid monster with a lion's head, goat's body and serpent's tail (The Concise Oxford Dictionary, 1982). Ligation may be facilitated using DNA ligase and ATP provided the gene and vector have "cohesive" or "sticky" ends, or linkage may be achieved using "blunt"-ended DNA. Cohesive ends are generated using certain restriction endonucleases which cleave specific and identical recognition sequences in both vector and gene of interest (Horton *et al*., 1996; Weaver, 1999), without compromising gene function, such that a section of the cleaved, single-stranded molecules have complementary base types which can base pair (Watson and Crick, 1953) during ligation. Blunt-end ligation may be facilitated between double-stranded DNA molecules in tandem. Where vector and gene of interest are cleaved with different restriction enzymes, oligonucleotides may serve to facilitate linkage as "adaptor" molecules (Bahl *et al*., 1978). A gene of interest is usually ligated "downstream" of the regulatory operator-promoter regions and "upstream" of the terminator sequences. These functional domains determine the direction of transcription or gene expression which is indicated downstream of the promoter by an arrow up to the start of the terminator sequences. Depending on the location of functional plasmid domains (2.5.4), the direction of transcription may be anticlockwise or clockwise in circular plasmid (Weaver, 1999). This is relevant for a static diagram although a variety of DNA configurations may form *in vivo* as molecules move in

three dimensions (Weaver, 1999). A gene flanked by regulatory and terminator sequences may be referred to as an "insert" or "cassette" (Horton *et al*., 1996). A cassette may be constructed within a transposon to facilitate its transfer from plasmid and integration into genomic DNA (Barr *et al*., 1989). Cloning within a functional plasmid domain such as a marker gene results in "insertional inactivation" of that marker while the replacement of a marker gene with a gene of interest is referred to as "directional" cloning. These examples illustrate the use of symbolic language to define the mechanism of cloning (Horton *et al*., 1996; Weaver, 1999).

An entire genome may be subject to restriction digestion and each fragment cloned separately into a plasmid. The recombinant plasmid collection constitutes a "genomic library" where a specific gene may be identified from a clone by hybridization with radioactive probe DNA or RNA that is complementary in sequence. Hybridization can be analysed using a combination of agarose gel-electrophoresis, "Southern blotting" (named after the E.M Southern who described the method of transferring DNA from gel onto nitrocellulose paper) and X-ray sheet overlay to detect radioactivity at the precise probe-DNA complex (Southern, 1975; Horton *et al*., 1996; Weaver, 1999). "Northern" (RNA) and "Western" (protein) blots may be used to analyse gene expression at the transcriptional and translational levels, respectively. This may involve the analyses of mRNA-cDNA hybridization or Western blot-ELISA (ELISA: enzyme-linked immunosorbent assay) using antibodies specific for the blotted protein. Symbolically, these analogous techniques are named after a "direction" to differentiate them from the first DNA blot described by E.M. Southern. Recently, DNA microarrays (fluorescent "DNA chips") have been used to detect specific genes using robotic technology to synthesize and immobilize oligonucleotides as probes (Elliot and Elliot, 2005; Garrett and Grisham, 2005; Takayama, 2005). Terminally labelled radioactive cDNA probes may also be used to detect specific genes by "chromosome walking" where the hybridization of probes covers the entire length of chromosomal DNA (Horton *et al*., 1996; Weaver, 1999). In addition, cloned genes may be subject to sequencing where the order and composition of purine and pyrimidine bases are determined from the 5′ to 3′ end (also symbolically indicated as the 5′→ 3′ sequence) in one strand and illustrated from left to right (Weaver, 1999; Takayama, 2005). The complementary sequence of the

other strand in double-stranded DNA is easily determined from base-pairing pattern of DNA (Watson and Crick, 1953). Techniques such as DNase footprinting, DNA fingerprinting or restriction fragment length polymorphism (RFLP) analyses (Horton *et al*., 1996; Weaver, 1999) reveal the unique identity of DNA samples based on nucleotide sequence. Symbolically, a "footprint" or "fingerprint" reveals the unique identity of an individual by providing distinctive impressions on most surfaces. Using these techniques, unique banding pattern gives the definitive characteristic or identification property of a DNA sample. The symbolism associated with gene cloning and the analysis of gene expression presents with unique and wide-ranging terminology, technical procedures and molecular mechanisms associated with expression. The learning of such information can be an imposing task as various forms of symbolism carry a depth of information on these topics.

2.6 Summary and conclusion

The analysis of the diversity and prevalence of symbolic language in this Chapter exposes this researcher to unique patterns of information dissemination. As can be seen in the literature, various forms of symbolism or representations communicate information related to a range of entities encountered in Molecular Biology. The symbolic representations must be understood in terms of their underlying meaning and it is imperative that learners become "visually literate" by providing correct interpretations of symbolic representations in Molecular Biology. In response to the research question, "What is the nature and diversity of the symbolic language used in Molecular Biology?", the author indicates that there is sufficient evidence that symbolic representations in Molecular Biology are not only diverse but complex in terms of the information they depict (2.5). Not only is the diversity of symbolic language challenging, but there is an implicit hierarchical structure of conceptual information that is linked to the symbolic representation. These "hidden transformations" are also not provided to the students and can be equally prohibitive for learning. Diverse symbolic representations of the same conceptual information clearly challenges conceptual development and the epistemological profile of learners. Learning from such representations can be rather challenging and confusing to students. The information presented in this

chapter constitutes a strong motivation for the current investigation into the nature of learning difficulties with symbolism and visual language used in Molecular Biology. There is currently limited published research on the learning difficulties that symbolism presents in Molecular Biology. Using specific probes (Chapter 3) to conduct the investigation, the author presents complex learning difficulties associated with symbolism and visual language in Molecular Biology (Chapters 4 to 8).

CHAPTER 3

METHODS

3.1 Selection of participants

This study was conducted over the period 2001 to 2004 on undergraduate students (groups n_{1-4}) who had enrolled for the Biochemistry major at the former University of Durban-Westville, SA, currently known as the University of KwaZulu-Natal. The student enrolment between 2001 and 2004 was 54 (n_1), 35 (n_2), 31 (n_3) and 44(n_4), respectively. In 2005, a further 30 students (n_5) contributed to the evaluation of a remediation model (Chapter 8). The participants were of Afro-Asian extraction of the age group 19 to 24. English is widely spoken by the participants; however, varying degrees of fluency may be encountered as it is not regarded as a native, domestic language.

3.1.1 Instructional objectives and learning experiences

All participants had passed first year (level 1) modular course components in Chemistry; Physics or Mathematics; Zoology or Botany or Biology; Introductory Biochemistry and Microbiology; and, English Language Development or Changing Society: Culture, Ideas and Values. Second year (level 2) credits comprised modular components making up combinations of Biochemistry, Microbiology and either Zoology or Botany or a combination of Biochemistry, Microbiology and Chemistry. In the third year of study, the Biochemistry major was generally taken in combination with either Chemistry or Microbiology. Molecular Biology features strongly in various undergraduate modules of Microbiology and Biochemistry. Courses are taught by dissemination of lectures, tutorials and practical experimental work related to the theoretical components of their respective syllabi. All components are assessed by written tests during the coursework and examination at the end of a 13 week semester. In Biochemistry, the course outcomes are focused on developing an understanding of Biochemistry and Molecular Biology topics. The molecular attributes of processes, regulation, control of chemical processes

and related laboratory training receive strong emphasis in the teaching process. The author induces Socratic dialogue in some of his lectures, based on the dissemination of reading material prior to lectures and a revision of previously taught lecture content (Guterman, 2003; Hake, 2008). The standard and depth of information, for different levels of study, is regulated through a system involving moderation, external examination and a comparison of content information with those taught at other South African Universities. This enables revision of content and inclusion of current information that might be of interest.

3.1.2 Ethical considerations

The participants were clearly informed that their involvement in this study was voluntary and purely to assist this researcher investigate learning difficulties. They were also informed that this initiative would assist in remediation and help improve the teaching of Molecular Biology. Following a verbal agreement from the participants, the researcher developed probes to investigate learning difficulties. For the sake of confidentiality, the participants were informed that their identity or names will not be revealed in this work and that interview recordings will be converted into written transcripts. Name codes, such as S1, S2, S3 etcetera, are used instead (Chapters 4 to 8). This project commenced on the basis of an ethical clearance agreement between the former University of Natal (now University of KwaZulu-Natal) and the Science Education Research Group led by Professor Trevor R. Anderson.

3.2 Determination of research focus areas

The research focus areas were chosen by the author on the basis of his observation of certain recurring difficulties shown by his students that he encountered over a 15 year period as a teacher of Molecular Biology. In addition, the content analysis (Chapter 2) revealed complex and diverse symbolism associated with the focus areas of this study. Certainly an investigation into potential difficulties that such symbolism presents will be

beneficial to researchers, teachers and students in terms of remedial assistance. The author anticipated that similar difficulties might emerge among the selected participants (3.1) of this 5-year study. The current research focuses on:

i. Symbols, with specific reference to "Δ" and "**::**" denoting gene deletion and gene fusion or transposition, respectively.
ii. Shorthand notation of oligonucleotides, with specific reference to the identification of phosphite and phosphodiester groups.
iii. Symbolic representation of plasmid conformation, function and restriction mapping.
iv. Gene markers: the visual impact, their role in cell fusion technology and the selection of transformants.
v. Nomenclature and visual representation of the function of nucleic acid templates.
vi. Schematic diagrams: DNA replication intermediates and related models.

The research approach or paradigm (Kuhn, 1962, cited by Carey, 1986; de Jong, 2002; Matthews, 2004), which is adopted for the areas of research, is presented in section 3.4 of this Chapter. In addition, the aim of the work presented in this chapter was to synthesize propositional knowledge and appropriate content information (Cohen *et al*., 2000) on each of the above six focuses, based on the nature of recurring student difficulties (Chapters 4 to 8) and the writer's experience as a teacher. Such statements and information would constitute scientifically "correct" controls which would enable identification of any difficulties through a comparison (Cohen *et al*., 2000) with information, gathered from oral and written student communications, in response to interviews and probes, respectively (Chapters 3 to 8). New or unexpected students' responses, showing evidence of possible student difficulties, are given special attention in terms of the research framework presented by Grayson *et al*. (2001) (3.4). The propositional knowledge was constructed on specific aspects relating to two of four modular components of the Biochemistry major (3.1), namely, Advanced DNA Chemistry and Advanced Protein Chemistry and Dynamics. Cell fusion hybridoma technology (3.5.6.2) was taught within the framework of Immunochemistry in the latter

module. These modules comprised theoretical and practical components and were both taught over a thirteen-week semester. Further, the propositional knowledge was ratified as correct by Professor Mario Ariatti at the University of KwaZulu-Natal, Durban, South Africa. Two other experts in Molecular Biology also assisted in validating the propositional knowledge.

3.3 Content analysis of the nature of symbolism in Molecular Biology textbooks

According to Cohen *et al*. (2000) content analysis is a method used to derive quantitative data from qualitative sources of information including visual forms of communication (Bodemer *et al*., 2004; Mathewson, 2005; Schönborn and Anderson, 2006a). In this study, content analysis served as a briefing to establish and understand the nature and diversity of symbolism (Chapter 2) that is found in a selection of Molecular Biology and Biochemistry textbooks. The content analysis was initiated with a hypothesis in anticipation that diverse information would be found during this exercise. The selection of books was based on those read by the writer over the past 15 years. The majority of these were recommended to students as reading texts to supplement the information received at lectures where necessary. The books have been written by renown experts in the field of Molecular Biology. An initial preview of symbolism from the books and related scientific literature revealed diverse forms of symbolism which required placement in various categories, followed by further refinement into subcategories. The categories included single component labels, geometric or structurally defined symbols, for example, squares, triangles, circles, shaded dots or arrows, abbreviated nomenclature, drawings, models, maps, graphs, radioactive transmission on X-ray sheet, machine-generated symbolism such as computer-generated images, electrophoretograms, fluorescence scans, diffraction patterns or X-ray crystallography images and symbolic language itself. The list of such categories is far more extensive than described here and depicts thousands of examples of symbolism in Molecular Biology.

In view of the above, an understanding of the use of symbolism was therefore sought in a defined range of Molecular Biology topics or entities (Chapter 2; Table 2.1) which formed the basis of investigations in the current study (3.2). Other examples of interest which fitted the content of the undergraduate Biochemistry curriculum (3.1) also received coverage (Chapter 2). Content analysis was used to establish the diversity and pattern of information communication, coupled with an evaluation of standardisation and prevalence in the textbooks. Not all examples, fitting the various categories, could be reported in this analysis as "culling" was inevitable to allow for typical representation fitting each category. Symbolic representations fitting the category of "structure" required clear differentiation for placement in various sub-categories (Table 2.1). Based on this researcher's judgement, subjectivity was minimized and inter-rater reliability was ensured by comparing the data presentation from the various textbooks with information from published literature (Chapter 2) and data collected from students' written communications (3.4). To ensure constancy of this exercise, a re-evaluation of the search for information was attempted after a time lapse of two weeks.

3.4 Adoption of a methodological framework for data analysis

The four-level methodological framework of Grayson *et al*. (2001) was used to investigate several level 2 difficulties that were suspected from several years of teaching Molecular Biology to final-year Biochemistry students at the University of Durban-Westville. In essence, the four-level framework classifies student difficulties according to the extent of information and understanding researchers have about the nature of each difficulty. The criteria used to classify difficulties at each level are as follows. Difficulties that have been systematically investigated and found to be well-established in different contexts, (e.g. different courses, institutions, different student populations and educational settings) and for which there is a stable description are classified at level 4 as "established". Difficulties that are known to researchers but not extensively explored are classified at level 3 as "partially established". Level 2 difficulties are those that are "suspected" by the researcher on the basis of learning or teaching experience as described

in this study. Difficulties which emerge unexpectedly from the analyses of the data are classified at Level 1 as "unanticipated". As far as the researcher knows, no systematic investigations have been carried out on the prevalence and nature of the Molecular Biology-related difficulties reported in this thesis.

Data on student difficulties were obtained from student responses to free-response type written probes that had been specifically designed to focus on the taught subject matter. All the participants (n_{1-4}) (3.1) received the same free-response probes during the course of a thirteen-week semester. The responses were compared with the respective propositional knowledge (3.5) and sorted by inductive analysis (McMillan and Schumacher, 1993) into various categories or groups according to the nature of difficulty displayed. Re-grouping of difficulties was necessary when some responses were re-analysed because of obscurity or unintelligible expression. In this way, various categories of difficulties were allowed to emerge (Lincoln and Guba, 1985) and the incidence of each could be calculated. At this stage, the assignment of an initial classification level (Grayson *et al*., 2001) of 1 or 2 for unexpected and anticipated difficulties was possible, respectively.

3.4.1 Validation and expansion of research findings

The free-response probes were designed so as not to lead the students into particular ideas but to allow the responses to emerge naturally and without inhibition (Lincoln and Guba, 1985). The initial analyses of data (3.4) prompted further investigations into the nature of various difficulties. This involved the subsequent dissemination of focused written probes to collect data from the students (n_{1-4}) (3.1), further confirming that the probes did in fact search the nature of the initial recorded difficulty. Focused probes were also designed to collect student-generated diagrams (SGDs), where applicable, and their annotated notes to explain the meaning of scientific phenomena or processes (Barlex and Carre, 1985; Kindfield, 1994; Henderson, 1999). The incidence of all difficulties was recorded (Chapters 4 to 8). A content analysis (Cohen *et al*., 2000; 3.3) of certain sources of

literature was also used to assess factors possibly contributing to students' reasoning and conceptual difficulties. As more information was collected from the students and literature, this prompted a re-classification of some of the difficulties in accordance with the four-level framework provided by Grayson *et al*. (2001).

Greater insight into certain difficulties could be recorded from tactful interviews (Posner and Gertzog, 1982) involving selected students and respondents who had volunteered to engage in this process. Interviews were initiated in a relaxed and friendly manner, establishing rapport with the interviewee where the initial discussion focused on university life, happenings over the weekend or current sport. The interviewees appeared relaxed when it was made known that the interview was simply a discussion to assist with the current research rather than a clinical analysis of answers. It was further explained that the interviewer was not assessing the correctness of answers but simply keen to establish their understanding of the certain issues. The recording of interviews served to transcribe information verbatim (Posner and Gertzog, 1982; Cohen *et al*., 2000). The transcription included recordings of hesitation, silence, laughs, sighs, a pause etcetera including physical gestures such as nodding, pointing, showing or drawing of diagrams. Questions were focused on the subject at hand and every attempt was made not to lead the students into desired responses but to allow responses emerge freely and naturally (Lincoln and Guba, 1985). Questions encouraged further discussion to delve deeper into understanding where candidates were encouraged to say more about their subject. At closure all interviewees were thanked for their input. Both questions from the interviewer (I) and responses from the students (name codes S1, S2, S3 etcetera) are given in italics, respectively. Where indicated, the interview was varied to facilitate the interactive "think aloud" exercise (de Jong, 2000; Schönborn *et al*., 2002a) which allowed responses to emerge freely and naturally (Lincoln and Guba, 1985) as students interpreted diagrams given to them (3.5.8; Kindfield, 1994).

In summary, qualitative data collection (Chapters 4 to 8) was achieved through a system of triangulation involving a free-response probe, focused probe (3.5) and interview (3.4.1) to ensure cross validation of findings (McMillan and Schumacher, 1993). The

written probe work was conducted as tests during a thirteen-week semester to ensure continuous assessment, coupled with a semester-end examination of total course content. The course of interviews, for each section investigated (3.5), is presented in Chapters 4 to 8.

3.5 Written probes and corresponding propositional knowledge

In response to the probes, the propositional knowledge presented in this study may exceed the depth of information expected from undergraduate students who reflect largely on content knowledge taught during coursework. The compilation of broader propositional knowledge, which exceeds such content knowledge, certainly assisted this researcher evaluate answers from those students who read beyond the prescribed syllabus or extracted information from diverse sources or framed this incorrectly against a written or oral probe. The references, supporting the propositional knowledge in the sections that follow, are not expected from the students but simply a source of knowledge extraction by this researcher. Further, a broader search for the correct use of factual information in a specific context by many other reporters validates and enhances acceptance of the knowledge itself. In addition, the author also provides concise propositional knowledge statements (Garnett *et al*., 1985) on various forms of symbolism which enable quick reference to definitions and a comparison of students' responses to the probes. All propositional knowledge, relevant to each Molecular Biology topic investigated, (Chapters 4 to 8) finds use in analysing the students' response to interview questions and think aloud exercises (3.4.1).

.

3.5.1 The symbol "Δ"

The use of the symbol delta in the various scientific contexts concerning temperature, optical density, free energy change, proton gradient, lipid chemistry and gene deletion, as

described below, have been introduced to students in their undergraduate Biochemistry training.

The initial free-response written probe, given to students, comprised the following:

Explain the meaning of "Δ" in a scientific context.

The following propositional knowledge is constructed to assess the students' responses on the use of delta in a scientific context. The sign is widely accepted as the chemical sign of heat in many sites on the World Wide Web. In laboratories, it is found as a label on hazardous chemical bottles. Scientists have different and specific uses for delta. These include difference in the measurement of a parameter, for example, $\Delta T\ (^{O}C) = T_2 - T_1$, ($\Delta T$ = temperature difference between two readings) (Mammino, 2001), ΔG which bears reference to the change of free energy of a reaction (Feinman, 2004), ΔH^{+} as reference to proton gradient (Garcia-Vallve, 2004) or Δ^{5} which bears reference to the placement of a double bond between carbon 5 and carbon 6 in a steroid or unsaturated lipid structure (Nes and McKean, 1977). Should delta precede a gene written in lower case italics, this implies gene deletion (2.5.2).

Subsequently, a more focused probe was designed to invoke specific written responses and to challenge the respondents that the symbol "Δ" could be used differently in specific contexts. This probe read as follows.

Give the meaning of "Δ" in the following contexts:

1. ΔOD
2. ds DNA $\xrightarrow{\Delta}$ ssDNA
3. Δ *leu*
4. $\Delta T\ (^{O}C)$.

The following propositional knowledge enables the assessment of students' responses to the focused probe. In the specific contexts above, the symbol "Δ" denotes in; (1) the difference in optical density measurements ($OD_2 - OD_1$), (2) the conversion of double-stranded DNA to single-stranded DNA by heat, (3) deletion of a leucine biosynthesis gene and (4) the difference in temperature measurements ($T_2 - T_1$) in degree Celsius. Concise propositional knowledge statements concerning the use of delta also appear in Table 3.1.

3.5.2 The symbol ": :"

With regard to the symbol "**: :**", students were presented the following free-response probe:

Discuss the use of the symbol "**: :**" in genetic engineering.

The following propositional knowledge is presented to assess the students' responses on the use of the double colon in genetic engineering. The symbol "**::**" was originally used to indicate gene insertion (McClintock, 1951, 1956; Demerec *et al*., 1966;1968; 2.5.2) and currently remains the accepted way to show insertion of one gene within the domain of another, resulting in the inactivation of the disrupted gene (Table 3.1). The FlyBase Reference Manual C (2005) and Gene Proforma Version 33 database (2003) also recommend the use of the double colon to indicate in tandem gene fusion (2.5.2) and that this be clearly defined as a fusion which generates an open reading frame, leading to the transcription and translation of a single translation product expressing the functions of both fused genes (Websites in list of references). The accompanying definition is necessary to differentiate it from gene insertion or transposition. Transposon Tn5 has been shown to insert within the domain of *Lac(z)* to result in *lac(z)::Tn5* thereby destroying *lac(z)*'s ability to express β-galactosidase (2.5.2; Table 3.1).

Table 3.1 Propositional knowledge statements associated with the use of the symbols delta and double colon and shorthand structures for phosphite and phosphodiester groups

Symbolism / Symbolic language	**Propositional knowledge statements**
ΔT (^{O}C)	Delta reflects temperature difference between two readings in degrees Celcius e.g. T_2 - T_1.
ΔOD	Delta reflects the difference in optical density measurements e.g. $OD_2 - OD_1$.
ds DNA $\xrightarrow{\Delta}$ ssDNA	Delta reflects heat which converts double-stranded DNA to single-stranded DNA.
Δ *leu*	Delta reflects deletion of a leucine biosynthesis gene.
*lac (z)***: :**Tn5	The double colon implies transposition of the transposon Tn5 into the domain of gene *lac (z)* thereby inactivating it.
GENE1**: :**GENE 2	The double colon indicates in tandem gene fusion. The genes may be flanked by regulatory sequences of an expression plasmid.
valency	The ability of an atom to bond with another.
phosphodiester form of P	Chemical group showing the pentavalent form of phosphorus.
phosphite form of P	Chemical group showing the trivalent form of phosphorus.

The focused probe comprised the following:

Give the meaning of the symbol "**: :**" in the following gene constructs;

1. *lac (z)* **: :** Tn5
2. 5′-operator-promoter-GENE1**: :** GENE 2-terminator sequence-3′ (part of plasmid).

The propositional knowledge concerning example (1) of the focused probe has been described above as is the information concerning gene fusion from The FlyBase Reference Manual C (2005) and Gene Proforma Version 33 database (2003) which apply to example (2). Concise definitions are also provided in Table 3.1.

3.5.3 Phosphodiester and phosphite groups in oligonucleotide shorthand notations

Phosphodiester and phosphite groups may be differentiated in oligonucleotide shorthand notations on the basis of valency of the phosphorus atom. Hence, the following free-response probe established whether or not students understood the meaning of valency:

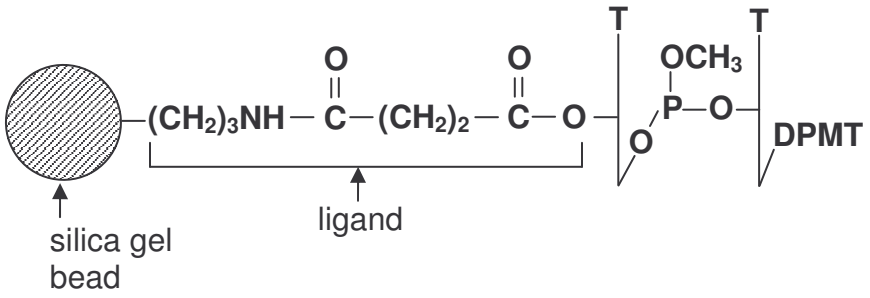
Explain the meaning of "valency of an atom".

A propositional statement (Table 3.1) such as "Valency simply means the ability of an atom to bond with another" (Partington, 1951) would be acceptable in response to the free-response probe.

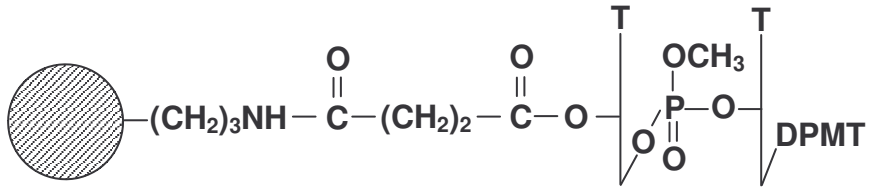
A focused probe presented schematic diagrams of oligonucleotide intermediates A, B and C as shown in Figure 3.1. Students were asked to comment on the valency of the P atom in each and differentiate the phosphodiester and phosphite groups (2.5.3; Table 3.1) encountered in the chemical synthesis of the oligonucleotides (Letsinger *et al*., 1975; Letsinger and Lunsford, 1976; Tanaka and Letsinger, 1982; Pon, 1987). The following propositional knowledge on the Valence-Shell-Electron-Pair-Repulsion (VSEPR) theory

is presented to enable an assessment of the structural form of the phosphodiester and phosphite groups .The Valence-Shell-Electron-Pair-Repulsion (VSEPR) theory indicates that the arrangement of bonds around an atom is determined by the number of electron pairs around an atom and the size and shape of the orbitals in which these electrons are housed. Based on the electron configuration of phosphorus (P), namely, $1s^2\ 2s^2\ 2p^6\ 3s^2\ 3p^3$ or $1s^2\ 2s^2\ 2p^6\ 3s^1\ 3p^3\ 3d^1$, the valency of this atom is either 3 or 5 as shown in the bond structure of the phosphite and phosphodiester groups, respectively (Partington, 1951). The energy sub-shells of phosphorus, namely s, p and d can accommodate a maximum of 2, 6 and 2 electrons, respectively. Excluding bonds to hydrogen, the Octet Rule specifies that the combination of atoms yielding a bond is such that each atom is surrounded by an octet of electrons (Peterson and Treagust, 1989; Peterson *et al*., 1989). In the phosphite form, P will accommodate a single electron from each of three oxygen atoms. The electron configuration of the oxygen atom is $1s^2\ 2s^2 2p^4$, sub-shell 2p being able to accommodate a total of 6 electrons. In the phosphodiester form, P will accommodate a single electron each from 3 oxygen atoms, forming 3 single bonds with oxygen. In the double bond involving the 4^{th} oxygen, four electrons, namely a pair each from phosphorus and oxygen, are shared between these atoms (Partington, 1951).

A



B



C

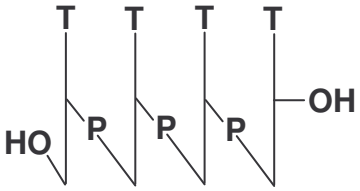


Figure 3.1 Schematic diagrams of oligonucleotide intermediates generated by the phosphite triester synthesis technique (Gupthar, 1983).

3.5.4 Plasmid conformation and function

Maps of a fictitious recombinant plasmid pBCH 301 of 4.8 kb, as shown in Figure 3.2, were presented to students. Supplementary information was also provided as follows. "Gene X, encoding an amylase, is cloned downstream of the promoter, following the *Bam* HI-*Hin*d III excision of a *Tet*^{r} marker gene. The abbreviations O, P, ts and Z(α) denote the operator, promoter, terminator sequence and β-galactosidase gene, respectively".

As more focused probes investigated the students' understanding of functional domains of the plasmid, they did not receive further supplementary information on; the ori site, autonomous replicating sequence (ARS), an amylase gene fused in tandem (**::**) with the *Lac(z)*gene encoding β-galactosidase, *LEU*, dominant leucine biosynthesis gene (excluding any amendments with reference to gene designation, allele or locus (Demerec *et al.,*1968) in order to simplify the format for students), *Amp*^{r}, the ampicillin resistance gene and restriction endonuclease sites *Pvu*II*, Bam*HI*, Hin*dIII and *Pst*I.

Students were presented the following free-response probe:

> Is it possible to express the genes of interest X and Z (α) in a linearized plasmid or must the plasmid be "circular and covalently closed" for this to occur?

The following propositional knowledge is presented to enable an assessment of the students' responses on plasmid conformation and gene expression. Cleavage of a plasmid could also affect the expression of the cloned genes adversely. Cleavage affects plasmid conformations (2.5.4) which might be preferred for transcription of the cloned genes (Brown, 1986). With regard to the *Pvu* II- linearised form of the hypothetical plasmid pBCH 301 (Figure 3.2), none of the plasmid gene sequences are uninterrupted thus expression is a possibility (3.5.4) although a prediction of the level of expression would

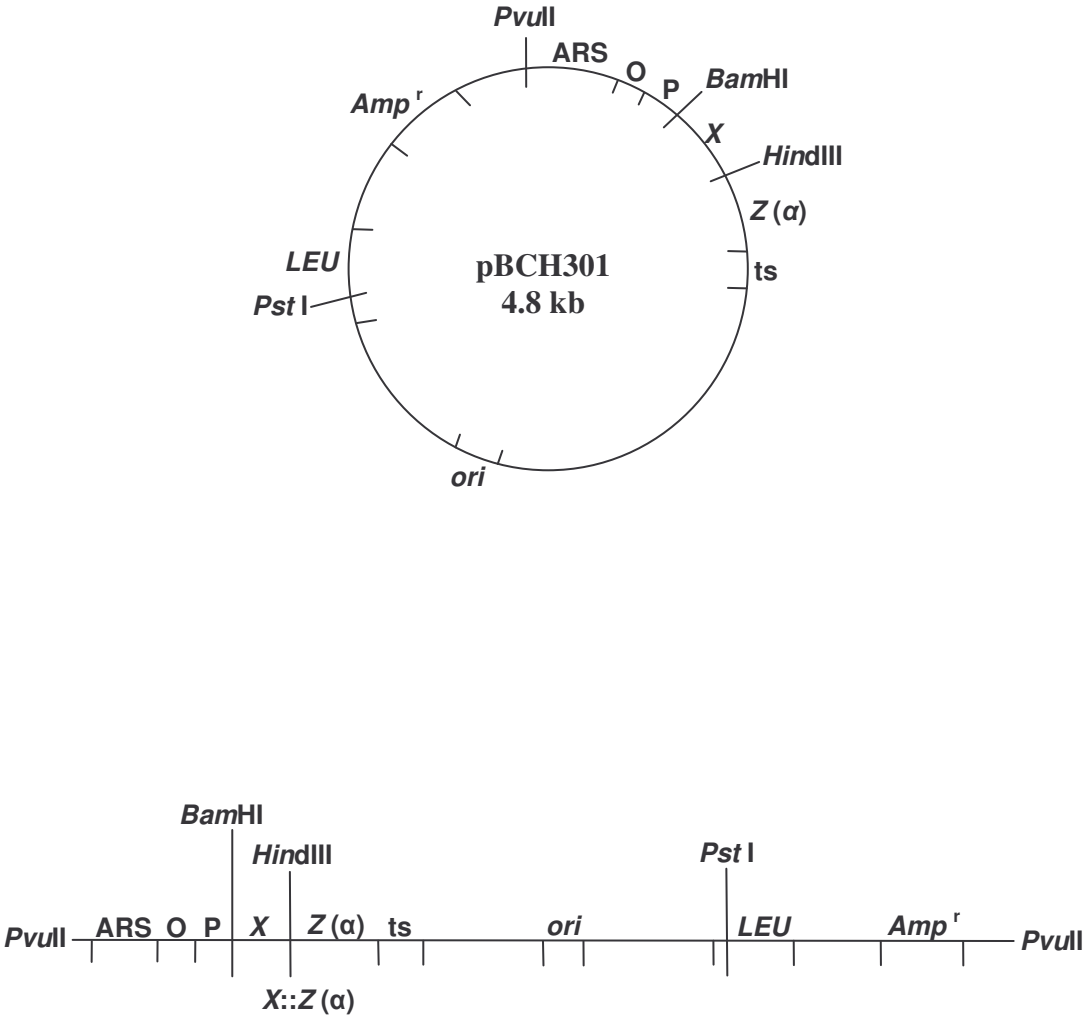


Figure 3.2 Circular and *Pvu*II- linearized map of plasmid pBCH 301.

be speculative. The *Pvu* II-cohesive ends of the plasmid indicate that recircularization and supercoiling are likely to take place *in vivo* (Gardiner *et al*., 1984; Brown, 1986). This is the conformation associated with gene expression by most plasmids (3.5.4; Brown, 1986). Plasmid forms may differ in cells. They are generally supercoiled (Kornberg and Baker, 1992; Turner *et al*., 1997; Virtual Lab Book website, 2005) in the cell or native state. Perhaps this is the conformation best suited for expression of the cloned genes rather than the "circular and covalently closed" conformation. However, certain yeasts harbour linear double-stranded plasmids which are associated with the expression of a killer phenotype (Ligon *et al*., 1989; McCracken *et al*.,1994).

In addition, students received the following focused probes:

1. Explain the notation "**: :**" in Gene X **: :** Z (α) and the consequences of expressing such a construct using pBCH 301?
2. How may we prevent re-circularization of the *Pvu* II – cleaved plasmid pBCH 301?
3. Give the role of ARS, ori, operator and promoter sequences in pBCH 301.

An extension of this work involving pBCH 301 and the selection of transformants are presented in section 3.5.6.3. The following propositional knowledge is presented to enable an assessment of the students' responses to the focused probes. In addition, Table 3.2 provides concise propositional knowledge statements on structural domains of the plasmid and their related functions.

In response to part (1) of the focused probe, students are expected to indicate that the notation refers to in tandem gene fusion. Gene fusion generates an open reading frame (orf), expressing both amylase and β-galactosidase as a single translation product as explained earlier (3.5.2).

Table 3.2 Propositional knowledge statements on symbolism associated with structural domains of a plasmid and their functions

Symbolism / Symbolic language	**Propositional knowledge statements**
operator (o)	The operator is a DNA sequence which controls the rate of mRNA formation.
promoter (p)	The promoter is the sequence at which DNA dependent-RNA polymerase binds to facilitate transcription.
terminator sequence (ts)	The terminator sequences signal the stop of transcription.
ARS	ARS is the autonomous replication sequence which ensures autonomous replication of plasmid DNA.
LEU	*LEU* denotes a dominant leucine biosynthesis gene.
*Pvu*II, *Bam*HI, *Hin*dIII and *Pst*I	*Pvu*II, *Bam*HI, *Hin*dIII and *Pst*I are restriction enzyme sites where the plasmid could be cleaved. The restriction endonucleases which cleave the respective sites carry the same nomenclature.
ori site	The ori site is the DNA sequence where replication begins.
Amp^r	Amp^r is the gene encoding resistance to ampicillin.
Tet^r	Tet^r is the gene encoding resistance to tetracycline.
Lac(z)	*Lac(z)* denotes the gene encoding β-galactosidase.
open reading frame (orf)	The open reading frame refers to a DNA sequence that is transcribed to mRNA.
downstream	Downstream refers to the position or location of DNA sequences that follow the operator and promoter sequences.

Following cleavage at a single site, the re-circularization of circular double-stranded DNA plasmids may be prevented using alkaline phosphatase to cleave the phosphate function at 3′ or 5′ positions of the deoxyribose sugar. This prevents re-ligation of the cleaved ends (Brown, 1986).

The ARS is associated with the autonomous replication of the plasmid. The operator controls the rate of mRNA formation from the gene(s) cloned downstream (Figure 3.2) of the regulatory sequences, namely operator and promoter. The promoter is the sequence at which DNA dependent-RNA polymerase binds to facilitate transcription of the cloned gene(s) which yield an open reading frame in the example presented. The terminator sequences signal the stop of transcription while the ori site is a region where plasmid replication begins. Propositional knowledge concerning functional domains of the hypothetical plasmid was extracted from Gardner *et al*. (1984), Brown (1986), Kornberg and Baker (1992) and Weaver (1999).

3.5.5 Restriction mapping

The following free-response probe was presented to students:

What do you understand by the term “restriction mapping”?

The following propositional knowledge can be used to assess students’ responses to the above probe. Restriction mapping is the placement of restriction endonuclease cleavage sites on DNA (Szeberényi, 2002; Table 3.2). Using a combination of “double-cutters” (two restriction endonucleases which cut at two specific nucleotide sequence sites) and single restriction endonucleases, fragments of various lengths may be obtained from circular double-stranded DNA. As first shown with SV40 DNA and commonly sited in textbooks (Table 2.1; 2.5), double-stranded circular DNA may be cleaved by a restriction endonuclease to linearize the molecule (Danna and Nathans, 1971). Cleavage of circular DNA, by different site-specific restriction endonucleases, generates fragments which

could be electrophoresed on polyacrylamide (Danna *et al*., 1973) or agarose gels containing ethidium bromide which facilitates their visualization under $UV_{254\ nm}$ light (Sharp *et al*., 1973; Szeberényi, 2002). Cleavage sites may be labelled on a diagram matching the electrophoretic pattern of the restriction digest between cathode and anode.

The following focused probe provided further information on the understanding of restriction mapping. Restriction endonuclease digestion of a plasmid (circular, double-stranded DNA) yields the following fragments (Figure 3.3):

*Hin*d III			4.1 kb		
*Hin*d III	+	*Pst* I	2.3	+	1.8 kb
*Hin*d III	+	*Pvu* I	0.4	+	3.7 kb
*Hin*d III	+	*Bam* H I	0.9	+	3.2 kb
Pst I	+	*Pvu* I	2.7	+	1.4 kb
Pst I	+	*Bam* H I	0.9	+	3.2 kb

1. Predict the electrophoretic separation of the fragments using agarose gels.
2. The plasmid (undigested) produces 2 bands of *ca.* 4.0 and 4.1kb when subject to agarose electrophoresis. Why should this occur?
3. Using data from the restriction digest, construct the plasmid map.

The following propositional knowledge can be used to assess students' responses to the focused probes on restriction mapping. In addition, concise propositional knowledge statements (Table 3.3) may also be used to assess the students' responses.

The predicted electrophoretic separation of larger retarded fragments from the smaller fragments is shown in Figure 3.3, using a relative kilobase scale and an illustration of an electrophoretogram. The retarded migration of larger fragments, relative to the smaller ones, will be seen towards the negatively charged cathode. Smaller fragments migrate rapidly towards the positively charged anode.

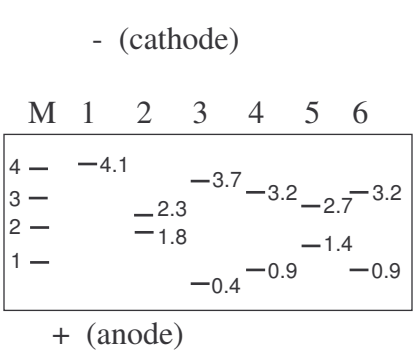
Table 3.3 Propositional knowledge statements on symbolism associated with plasmid conformation and restriction mapping

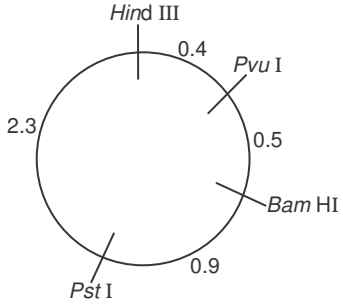
Symbolism / Symbolic language	**Propositional knowledge statements**
Restriction mapping	Restriction mapping is the identification and placement of restriction endonuclease cleavage sites on DNA.
Supercoiled DNA	DNA such as circular double-stranded plasmids that display a highly folded conformation.
cccDNA	cccDNA could refer to plasmid DNA which is circular, uninterrupted in sequence and displays a relaxed circular conformation.
ncDNA	ncDNA is nicked circular DNA as shown by DNA plasmids nicked on one strand by the enzyme topoisomerase I.
ocDNA	ocDNA is open circular DNA, an alternate symbolic nomenclature to ncDNA, used to describe plasmid DNA that is nicked on one strand by the enzyme topoisomerase I.

Restriction endonuclease digestion of a circular, double-stranded DNA plasmid yields the following fragments:

1.	*Hin*d III			4.1kb		
2.	*Hin*d III	+	*Pst* I	2.3kb	+	1.8kb
3.	*Hin*d III	+	*Pvu* I	0.4kb	+	3.7kb
4.	*Hin*d III	+	*Bam* H I	0.9kb	+	3.2kb
5.	*Pst* I	+	*Pvu* I	2.7kb	+	1.4kb
6.	*Pst* I	+	*Bam* H I	0.9kb	+	3.2kb

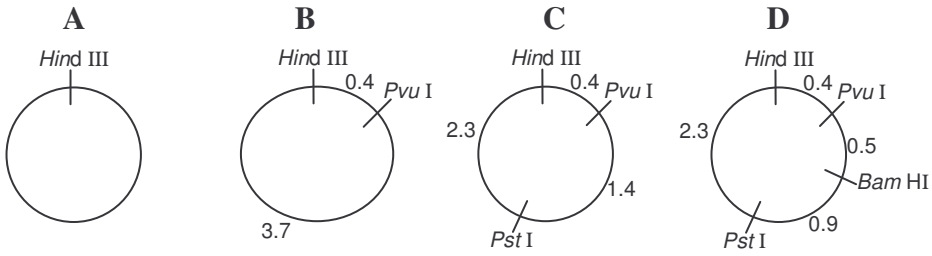
The following are the predicted electrophoretic separation of fragments:(arbitrary kb scale relative to an appropriate marker, M) and plasmid map ;





Plasmid map *[***E***]*

Rough work:



*The single Hin*d *III cut [* **A** *] yields a fragment of 4.1kb, indicating that the plasmid is 4.1kb. The sum of fragment sizes in all other cuts confirms the plasmid size. The Hin*d III-*PvuI digest yields 2 fragments of 0.4 and 3.7 kb [* **B** *].The Hin*d*III-PstI cut yields a 2.3 and 1.8kb fragment. The PstI cleavage site would therefore be between Hin*d*III and PvuI to give a 2.3kb fragment [***C****]. This means that the residual plasmid section of 1.8kb yields the 1.4kb PstI-PvuI kb fragment and 0.4kb Hin*d*III-PvuI fragment [* **D***].Relative to the PstI site, Bam HI must be 0.5kb away from PvuI[* **E***, refer to plasmid map] to generate PstI-Bam HI fragments of 0.9 and 3.2 kb.*

Figure 3.3 Construction of a plasmid map using restriction endonuclease digestion and fragment overlap.

An undigested plasmid would yield a rapidly migrating supercoiled form at 4.0 kb and possibly the ncDNA or cccDNA form at 4.1kb. The electrophoretic separation profile of plasmid conformers on agarose gels indicates that the compact supercoiled form migrates most rapidly, followed very closely by ncDNA (nicked circular DNA) or ocDNA (open circular DNA), the linear form and cccDNA (covalently closed circular DNA) (2.5.4) which is retarded towards the cathodic end (Brown, 1986; Turner *et al*., 1997; Virtual Lab Book website, 2005).

Based on the overlap generated by fragment cleavage patterns, it is possible to construct a restriction map of a circular plasmid (Figure 3.3) (Sharp *et al*., 1973; Weaver, 1999; Szeberényi, 2002). The plasmid map construction using restriction fragment banding and analysis of overlap is shown in Figure 3.3.

3.5.6 Gene markers and phenotypic expression

3.5.6.1 Probing the understanding of gene markers

The following free-response probe was given to students:

What are gene markers?

The following propositional knowledge is presented to enable an assessment of students' responses to the free-response probe. Gene markers are a form of shorthand notation which explains the phenotypic expression of genes (2.5.2). Genes expressing the phenotype are given the shorthand notation as nomenclature. An example would be the Tet^S gene which encodes the property or phenotype "sensitivity to tetracycline" (2.5.2) or $Hgprt^+$ Tk^+ characteristic of hybridoma cells which express the "salvage" enzymes hypoxanthine guanine phosphoribosyl transferase (*Hgprt*) and thymidine kinase (*Tk*), respectively (3.5.6.2).

Table 3.4 Propositional knowledge statements on symbolism associated with gene markers and phenotypic expression

Symbolism / Symbolic language	**Propositional knowledge statements**
Gene markers	A form of shorthand notation which explains the phenotypic expression of genes.
Phenotypic expression	The expression of the property or trait that a gene encodes is referred to as the phenotype.
Hgprt^{+} or *Hgprt*$^{-}$	Gene marker notation or symbolism indicating the expression (superscript +) or non expression (superscript -) of the enzyme hypoxanthine guanine phosphoribosyl transferase.
Tk^{+} or *Tk*$^{-}$	Gene marker notation or symbolism indicating the expression (superscript +) or non expression (superscript -) of the enzyme thymidine kinase.
Tet^{S}	*Tet*^{S} is the gene encoding sensitivity to tetracycline.
Amp^{S}	*Amp*^{S} is the gene encoding sensitivity to ampicillin.
leu^{-}	Gene marker symbolism indicating auxotrophy for leucine.
Lac(z)$^{-}$	Gene marker symbolism indicating non expression of β-galactosidase

More focused probes delved further into the students' understanding (Chapter 5 and 6) of gene markers.

These were:

1. Give examples of gene markers and present information regarding their function.
2. What do you understand by the term "phenotypic expression"?
3. Can gene markers be visualized or seen?

The following propositional knowledge is presented to enable an assessment of students' responses to the focused probe. Table 3.4 also lists concise definitions for various forms of symbolism used in this aspect of research. There are numerous examples of gene markers as indicated in Chapter 2. For example, Tet^S and Amp^S indicate sensitivity to tetracycline and ampicillin, respectively (2.5.2) while leu^- and $Lac(z)^-$ indicate auxotrophy for leucine and non expression of β-galactosidase, respectively.

The expression of the property or trait that a gene encodes is referred to as the phenotype.

Gene markers cannot be seen but they could be identified by a property or trait they encode (e.g. the phenotypic gene marker Tet^S encodes sensitivity to tetracycline). If such a (phenotypic) marker is isolated as a DNA fragment or gene, and subjected to agarose gel-electrophoresis and ethidium bromide staining (Sharp *et al*., 1973), it could be visualized as a fluorescent band of DNA. This makes it similar to a fluorescent band of molecular weight marker DNA which is subject to the same electrophoresis. Further discussion on student difficulties differentiating gene markers (phenotypic gene markers) from marker DNA (molecular weight markers) is presented in Chapter 5.

3.5.6.2 Selection of hybridomas

The following free-response probes were initially used to ascertain whether or not students understood the concept of a hybridoma and its selection (Chapter 6):

What are hybridomas? How are they produced and selected?

The following propositional knowledge is presented to enable an assessment of students' responses to the free-response probe. Hybridomas are antibody-secreting cells generated by the fusion of a sensitized population of murine B cells and myeloma cells (Köhler and Milstein, 1975). They are selected in HAT medium and can be propagated indefinitely using this medium.

A focused probe was designed to extract finer detail on the role of gene markers and hybridoma selection. It read:

> Discuss the basis for hybridoma *Hgprt*$^{+}$ *Tk*$^{+}$ selection using HAT medium. Why is the same medium detrimental to hybridoma progenitor cells of the phenotype, *Hgprt*$^{+}$ *Tk*$^{-}$ / *Hgprt*$^{-}$ *Tk*$^{+}$?

The following propositional knowledge is presented to enable the assessment of students' responses on gene markers and hybridoma selection. In addition, concise propositional knowledge statements are also included on the symbolism associated with the hybridoma cell fusion technique (Table 3.5). The hybridoma cell fusion technique (Köhler and Milstein, 1975) involves the fusion of a sensitized population of both murine B cells and myeloma cells to generate hybridomas. The hybridoma acquires the ability to secrete antibodies of defined specificity from the B cell progenitor. In addition, it can be propagated to be immortal *in vitro* using HAT medium, a property characteristic of the myeloma cells. The parental progenitors of this somatic cell fusion may be subject to mutation using UV light or a chemical agent such as N-methyl-N′-nitro-N-nitrosoguanidine to induce complementary phenotypic markers, for example, sensitized B cell *Hgprt* $^{+}$ *Tk* $^{-}$ and myeloma *Hgprt*$^{-}$ *Tk*$^{+}$ (2.5.2), where the superscripts (+) or (-) denote the expression or non expression of the "salvage" enzymes hypoxanthine guanine phosphoribosyl transferase (*Hgprt*) and thymidine kinase (*Tk*), respectively (Köhler and Milstein, 1975; Staines, 1983; 2.5.2). The fusant or hybridoma *Hgprt* $^{+}$ *Tk* $^{+}$ is selected using HAT medium, essentially a mineral salts medium (Dulbecco's Modified Eagles Medium) containing the components hypoxanthine, aminopterin and thymidine

Table 3.5 Propositional knowledge statements on terms and symbolism associated with hybridoma and transformant selection

Symbolism / Symbolic language	Propositional knowledge statements
Hybridomas	Hybridomas are an antibody-secreting cells generated by the fusion of a sensitized population of murine B cells and myeloma cells.
Sensitized Population	Population of cells, usually B cells, derived from an animal or individual whose humoral immune system has been challenged by an antigen.
HAT medium	HAT medium is a shorthand notation for Dulbecco's Modified Eagles' Medium containing the supplements hypoxanthine, aminopterin and thymidine.
"Salvage" enzymes	Salvage enzymes are those produced to facilitate steps in nucleotide metabolism especially when the *de novo* synthesis of certain nucleotide metabolites is blocked by an inhibitor owing to inhibition of *de novo* pathway enzymes.
dUMP	dUMP is the abbreviated form indicating deoxyuridine monophosphate.
dTMP	dTMP is the abbreviated form indicating deoxythymidine monophosphate.

Table 3.5 continued.

NADPH	NADPH is the abbreviated form indicating nicotinamide adenine dinucleotide phosphate in the reduced state.
lMP	lMP is the abbreviated form indicating inosine monophosphate.
HX	HX is the abbreviated form indicating hypoxanthine.
AMP	AMP is the abbreviated form indicating adenosine monophosphate.
GMP	GMP is the abbreviated form indicating guanosine monophosphate.
N^5, N^{10}	This form of symbolism indicates a nitrogen atom at positions 5 and 10 of the molecule, the numbering of atoms following rules set out by the IUPAC (International Union of Pure and Applied Chemistry).
Transformant	A transformant is a cell which receives a plasmid.
Xgal	Xgal is the shorthand notation for 5-bromo-4-chloro-3-indolyl-galactopyranoside.
LEU$^+$	*LEU*$^+$ is gene marker symbolism which denotes leucine prototrophy.

(deoxynucleotide). Aminopterin blocks the *de novo* synthesis of purines and inhibits the conversion of dUMP to dTMP via thymidylate synthase. In addition, the compound inhibits dihydrofolate reductase which reduces dihydrofolate to tetrahydrofolate using NADPH as the reductant. Tetrahydrofolate which converts to N^5, N^{10} - methylene tetrahydrofolate, using serine or glycine, acts as methyl group donor in the conversion of dUMP to dTMP which is facilitated by thymidylate synthase (Köhler and Milstein, 1975; Staines, 1983; Kornberg and Baker, 1992; Stryer, 1995). To counteract the inhibitory effects of aminopterin on *de novo* purine biosynthesis, sensitized B cells $Hgprt^+$ Tk^- express the salvage enzyme hypoxanthine guanine phosphoribosyl transferase which uses endogenous phosphoribosylpyrophosphate to convert the free base guanine or hypoxanthine from the medium to GMP and IMP, respectively. IMP is an important precursor for the biosynthesis of purine nucleotides, AMP and GMP. However, the B cells are unable to counteract the inhibitory effect of aminopterin on the conversion of dUMP to dTMP via thymidylate synthase. The salvage enzyme, thymidine kinase is not expressed; hence the nucleoside thymidine from the medium cannot be phosphorylated to the nucleotide, dTMP. B cells of the phenotype $Hgprt^+$ Tk^- do not survive in HAT medium owing to the inhibition of synthesis of the pyrimidine nucleotide, dTMP.

Myeloma cells of the phenotype $Hgprt^-$ Tk^+ express the salvage enzyme thymidine kinase which converts thymidine from the HAT medium to dTMP thus overcoming the inhibitory effect of aminopterin on thymidylate synthase and dihydrofolate reductase which would be implicated in the conversion of dUMP to dTMP as described. However, these cells do not survive in HAT medium as they are unable to counteract the inhibitory effects of aminopterin on *de novo* purine biosynthesis. The salvage enzyme, hypoxanthine guanine phosphoribosyl transferase is not expressed hence the extraneous supply of hypoxanthine would not be converted to the purine precursor nucleotide, IMP as described above. The hybridoma $Hgprt^+$ Tk^+ survives in HAT medium as it overcomes the inhibitory effects of aminopterin on purine and pyrimidine biosynthesis by expressing the function of both salvage enzymes as described.

3.5.6.3 Selection of transformants

The free-response probe read:

> What do you understand by the term "transformant"? In general, how are transformants selected?

A transformant is a cell which receives a plasmid or one which takes up a fragment of DNA. They may be selected on the basis of new properties that they acquire through expression of plasmid-borne genes.

Students were also given the following focused probe involving transformant phenotypes and media components.

> Given the map of *Pvu* II - linearised recombinant plasmid pBCH301 (Figure 3.2) and the supplementary information above (3.5.4), students were asked to explain the possible selection or non selection of a transformant *Lac (z)*$^{+}$ *X*$^{+}$, using the host *E.coli Tet*^{s} *Lac (z)*$^{-}$ *leu*$^{-}$ *Amp*^{s} *X*$^{-}$ and minimal nutrient agar medium containing combinations of the following components.
>
> i. Xgal plus leucine plus tetracycline
> ii. Xgal plus leucine plus ampicillin
> iii. Starch plus leucine plus ampicillin
> iv. Starch minus leucine.

The following proposition knowledge enables an assessment of student responses to the focused probe. Table 3.5 presents concise propositional knowledge statements on symbolism associated with the transformant selection. An understanding of the phenotypic traits of both host and plasmid is a prerequisite to any prediction regarding the possible selection of a transformant on any one of the media listed above. *Tet*^{S} and *Amp*^{S} indicate that the host is sensitive to both tetracycline and ampicillin, respectively (2.5.2). The markers, *leu*$^{-}$ and *Lac(z)*$^{-}$, indicate that the host is auxotrophic for leucine and does not express β-galactosidase, respectively. In addition, the host does not express amylase as indicated by *X* $^{-}$ in the probe. The host would revert to a transformant expressing the following plasmid-encoded traits, namely, the production of both amylase and β-

galactosidase, ampicillin resistance and leucine prototrophy. The respective plasmid markers are indicated as X^+ for amylase production, $Lac\ (z)^+$ or Z^+ for β- galactosidase production, Amp^r for ampicillin resistance and LEU^+ (2.5.2) for leucine prototrophy or ability to synthesize leucine. A Tet^r marker gene (2.5.2) is excised from the plasmid (3.5.4) during the cloning of the amylase gene, "X".

A minimal nutrient agar (MNA) medium containing Xgal, leucine and tetracycline would be inhibitory to the transformant. Although the transformant would be prototrophic for leucine (LEU^+), it could assimilate the amino acid from the medium. In addition, the transformant would express β-galactosidase, an enzyme capable of breaking the chromogenic substrate Xgal (5-bromo-4-chloro-3-indolyl-galactopyranoside) (Vieira and Messing, 1982; Srinivasan, 1999) to galactoside and a 5-bromo-4-chloro-3-indolyl dye. Such a dye could impart a blue colouration to the transformant colonies. However, the transformant is likely to be inhibited in its growth by tetracycline.

On MNA medium containing Xgal, leucine and ampicillin, blue, leucine-assimilating colonies would emerge. They could break Xgal as described, assimilate leucine despite their prototrophy and overcome the inhibitory effects of ampicillin as the ampicillin resistance gene is expressed by the plasmid.

On MNA medium containing starch, leucine and ampicillin, the transformant is likely to survive. Being able to express amylase, the transformant could break starch to glucose and utilize the monosaccharide. Leucine assimilation is possible despite the transformant's acquisition of prototrophy (LEU^+) from the plasmid and the effects of ampicillin will be overcome by the plasmid-borne ampicillin resistance gene. On MNA medium, containing starch but minus leucine, the transformant would utilize glucose from starch as described and synthesize its own leucine.

3.5.7 Nomenclature and function of nucleic acid templates

The following free-response probes were used to collect data on the nomenclature and function of nucleic acid templates.

The following is representative of double-stranded (ds) DNA;

A

5′_____3′

_____

3′ B 5′

1. Name strands A and B and explain why?
2. (a) Which strand(s) is/are implicated in;
   i. Replication
   ii. Transcription?
   (b) Explain why in each case.
3. Comment on the mobility or lack of mobility of DNA templates and enzyme complexes during replication.

The following propositional knowledge enables the analysis of student responses to the probes. Further concise propositional statements on symbolism associated with DNA-strand nomenclature and function are presented in Table 3.6. Double-stranded DNA (Watson and Crick, 1953) is composed of two anti-parallel strands which are complementary in terms of base sequence and run 5→3′ in opposite directions as indicated in the diagram above. The two strands are labelled either "coding" or "template", depending on their respective function. The Biochemistry Student Companion (Scism, 1996) defines the coding strand as "the strand of DNA within a gene whose nucleotide sequence is identical to that of the transcribed RNA with the replacement of T by U in RNA". The template is defined as "the strand of DNA within a gene whose nucleotide sequence is complementary to that of the transcribed RNA". During transcription, RNA polymerase binds to, and moves along, the template in a 3→5′direction, catalysing the synthesis of RNA in a 5→3′ direction (Figure 3.4; Garrett and Grisham, 1995). An arbitrary binding position of the enzyme is illustrated in a static

diagram at the 3′ end of the template (Figure 3.4). Depending on strand labelling, an mRNA sequence will differ in complementarity should the labels of coding and template strands be swapped around. Complementarity in the sequence of mRNA is read from a straight chain depiction of the template (Figure 3.4). As template conformations in three-dimensional space are speculative, owing to flipping or rotation, students were neither challenged to provide comment on this matter nor to predict the sequence of transcripts that may result from altered template conformations.

When teaching DNA replication, it is customary to refer to both constituent strands of double-stranded DNA as template strands for semi-conservative replication (Figure 3.4). Illustrations of this generally indicate that each template strand is associated with a nascent DNA strand whose nucleotide sequence is complementary to that of the template strand (Figure 3.4; Meselson and Stahl, 1958).

In *E.coli*, DNA synthesis occurs in a 5→3′ direction where both template strands are encircled by the dimeric β_2 subunits of DNA polymerase III holoenzyme (Kong *et al*., 1992). The β_2 subunits of the dimeric core (Maki *et al*., 1988) act as a sliding clamp, facilitating rapid sliding and turning of both DNA and polymerase during replication, allowing at least 1000 nucleotides to be polymerized per second (Kong *et al*., 1992; Stryer,1995). Studies have been conducted of the movement of the holoenzyme on single-stranded linear and circular templates, primed at different known positions with a synthetic 15-mer. The enzyme was found to move in the direction of DNA synthesis, namely 5→3′, exploiting the 3′OH function or structure of the primer to effect the synthesis of DNA, complementary in sequence to that of the template (O′Donnell and Kornberg, 1985; Kornberg and Baker, 1992).

A focused probe, requiring student-generated diagrams (SGDs), was structured as follows:

> Give labelled diagrams to illustrate the transcription process from double-stranded (ds) DNA. Indicate DNA strand labels and their specific functions as well as the

Table 3.6 Propositional knowledge statements on symbolism associated with DNA-strand nomenclature and function

Symbolism / Symbolic language	Propositional knowledge statements
Replication	Replication refers to the synthesis or duplication of DNA using the parental DNA strands as templates.
Transcription	Transcription refers to RNA synthesis from a template strand of DNA.
Coding strand	The coding strand is the strand of DNA within a gene whose nucleotide sequence is identical to that of the transcribed RNA with the replacement of T by U in RNA.
Template strand	The template strand of DNA is one within a gene whose nucleotide sequence is complementary to that of the RNA transcribed from it.
Semi-conservative	This refers to a replication mechanism where each parental DNA template strand becomes associated with a nascent DNA strand whose nucleotide sequence is complementary to that of the template strand.
Translation	Translation refers to protein synthesis using the mRNA translatable sequence which determines the sequence and composition of amino acids in a polypeptide chain.
cDNA	cDNA is a shorthand notation for complementary DNA which has a nucleotide sequence complementary to that found in the template strand of DNA used in its synthesis.
mRNA transcript	mRNA transcript refers to mRNA that is synthesized from a template strand of DNA.

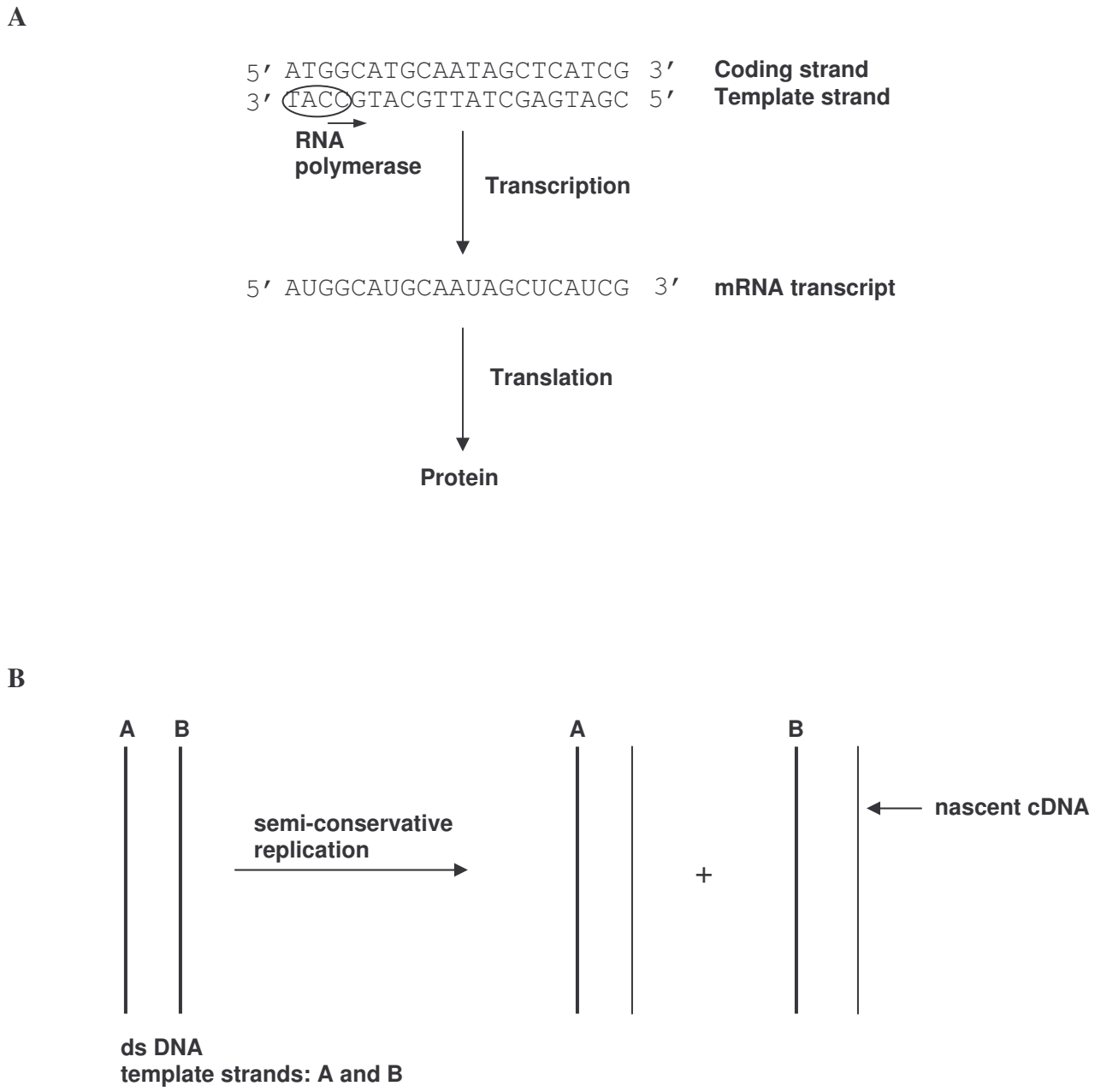


Figure 3.4 Illustrations of DNA-strand nomenclature and function in (A) transcription and (B) replication, adapted from Garrett and Grisham (1995) and Meselson and Stahl (1958), respectively.

end result of the transcription process. How do dsDNA strand function and labelling differ when referring to the replication process?

The propositional knowledge presented above, the concise propositional knowledge statements of Table 3.6 and Figure 3.4 proved adequate to analyse students' responses (Chapter 7) to this focused probe.

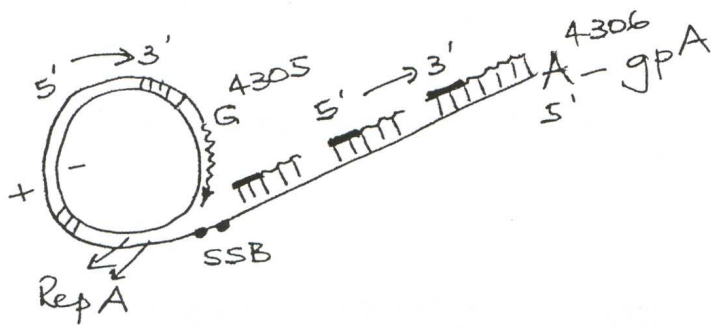
3.5.8 DNA replication intermediates and related models

The following free-response probes were used to assess students' difficulties with the mechanism of DNA replication, symbolic forms of structural representations and the factual information explaining the visual effect of changing structural DNA replication intermediates:

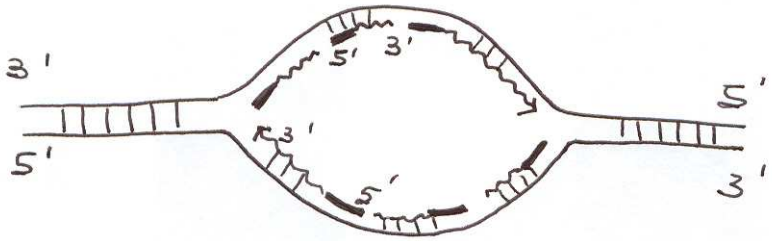
1. Give an illustrated account of the structural features of replicating DNA molecules and mention how each is derived.
2. Provide a well-labelled replication structural intermediate of your choice and write short notes about it involving the effect of enzymes.

The following focused probes were used to extract further information on students' difficulties with the mechanism of DNA replication:

1. Give an illustrated account of the structural features of ΦX 174 RF DNA and mention how each is derived.



2. What does the above illustration represent? Write short notes on all the structural features of the above and mention precisely how each is formed?



3. What does the above illustration represent? Write short notes on all the structural features of the above and mention precisely how each is formed?

Furthermore, the illustrations shown in focused probes 2 and 3 above (3.5.8) were also labelled to facilitate interviews and the accompanying "think aloud" exercises (3.4.1) involving randomly selected candidates. Labelled structural features (Figure 3.5) could be pointed to the student for identification and further comment as the interview progressed. Figure 3.5 (B) carries an error, deliberately incorporated to facilitate identification and correction during the course of interviews. Students were exposed to replication models of two types, namely:

i. An *E. coli* replication "bubble" model (Cairns, 1963a;b), adapted and presented by Bohinski (1987).

ii. A ΦX174 double-stranded RF I → RF II DNA replication model (RF denoting "replicative form") adapted from Goulian and Kornberg (1967), Dressler and Wolfson (1967) and Gilbert and Dressler (1968). This model is illustrated by the author (Figure 3.7) using the "rolling circle" mechanism and guidelines from Kornberg and Baker (1992).

The depiction of structural replication intermediates in the models represents a form of symbolism (2.5.7.1). The following propositional knowledge not only reflects the depth of the coursework and content presented to the students but a presentation of facts that decode the type of symbolism (2.5.7.1) associated with DNA replication. The description of consecutive changes to structural DNA replication intermediates, which are influenced by enzymes and auxiliary proteins, can be found in Bohinski (1987), Kornberg and Baker (1992) and Stryer (1995).

Replication bubble model

The following propositional knowledge may be used to evaluate students' responses on the *E. coli* replication bubble model. Specific propositional knowledge statements are also provided in Table 3.7 to assist with the definition of a range of symbolism encountered in DNA replication studies.

E. coli DNA replication begins at a site called the origin or ori C site (Figure 3.6). The sequence of the origin spans 245 bp in length (Oka *et al*., 1980). It is rich in palindromes and inverted repeat sequences which might play a role as recognition sites for various proteins implicated in the replication process (Oka *et al*., 1980; Funnell *et al*., 1987; Kornberg and Baker, 1992). Auxiliary proteins, designated dnaA and protein n′(PriA), bind to specific sites within the origin which must be negatively supercoiled by a gyrase for this to occur. Two other proteins, dnaB and dnaC join dnaA and protein n′ in opening

Table 3.7 Propositional knowledge statements on symbolism associated with DNA replication

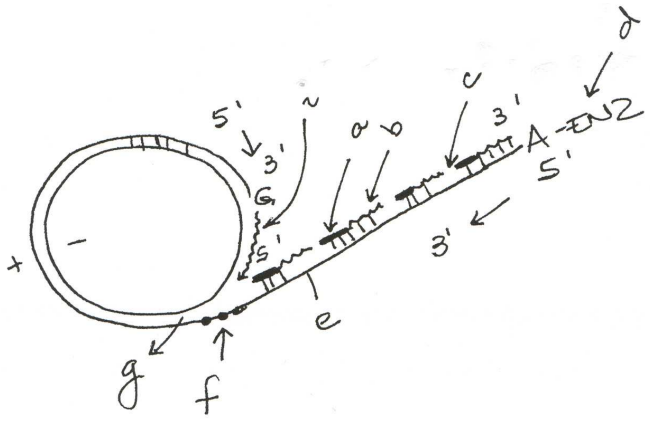
Symbolism / Symbolic language	Propositional knowledge statements
Origin	DNA sequence or site where replication begins.
"Eye" or "bubble"	Unravelling of the origin to give DNA the conformation resembling a bubble or shape of an eye.
Palindrome	Palindromes are sequences of bases that read the same in both directions on opposite strands of double-stranded DNA.
Inverted repeat sequence	This refers to the exact base sequence in opposite directions in duplex DNA.
Gyrase	A gyrase or topisomerase II is an enzyme which unwinds double-stranded DNA to result in a right-handed orientation or negative supercoil.
SSB protein	SSB protein refers to a protein which binds single-stranded template DNA preventing re-annealing of parental templates.
dnaA protein	DnaA protein guides the binding of the primosome complex to the replication eye or bubble where it also promotes opening of the double-stranded DNA.
dnaB protein	DnaB protein is the alternate nomenclature for a helicase which requires ATP to unwind DNA.
DNA primase	DNA primase is the enzyme which synthesizes RNA primers using the DNA template strand to determine the complementary RNA nucleotide sequence.
Protein n´(PriA), Protein n (PriB), Protein i(dnaT) and Protein n´´(PriC)	Protein n´(PriA), protein n (PriB), protein i (dnaT) and protein n´´(PriC) are protein components of the primosome complex which facilitates RNA primer synthesis. . .

dnaC protein	A protein which complexes dnaB protein and delivers dnaB protein to DNA at the origin.
Primosome complex	A complex of several proteins which facilitate the synthesis of RNA primer at various points along a DNA template.
Replication fork	Bifurcated or Y- shape feature in helical DNA produced during helicase action, separating the template strands of double-stranded DNA for replicative DNA synthesis.
Helicase	Helicase is an enzyme which unwinds DNA using ATP to do so, exposing the DNA template strand for RNA primer and nascent cDNA synthesis.
Re-annealing	Rehybridization of single-stranded nucleic acid strands based on nucleic acid base complementarity.
Negative supertwist or negative supercoil	Negative supertwist or negative supercoil refers to a relaxed right-handed twist in the conformation of double-stranded DNA induced by the action of gyrase or topoisomerase II.
Unravelling	Unravelling refers to the unwinding of the helical DNA structure at the origin to form an open structure resembling the conformation of an eye or bubble.
Positive supercoil or positive supertwist	Positive supercoil or positive supertwist refers to a tight left-handed twist in the conformation of double-stranded DNA induced by the action of topoisomerase I.
ATP	ATP is the abbreviation which refers to adenosine 5´-triphosphate.
rNTP	rNTP is the abbreviation which refers to ribonucleoside triphosphate.
RNA primer	RNA primer refers to a short template-bound RNA fragment which serves as a growth point for cDNA synthesis.
dNTP	dNTP is the abbreviation which refers to deoxyribonucleoside triphosphate.

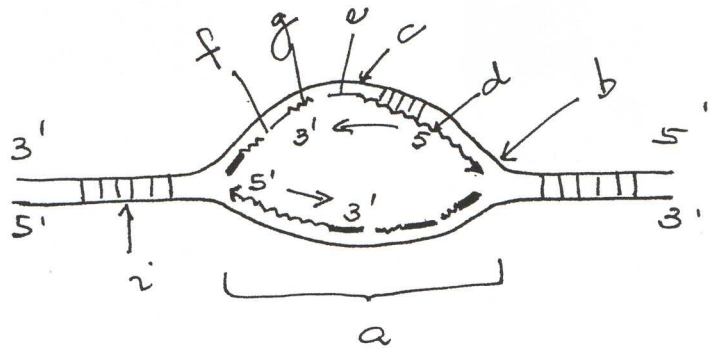
Leading strand	The leading strand refers to nascent DNA which is continuous in its synthesis.
Lagging strand	The lagging strand refers to nascent DNA which is discontinuous in its synthesis.
Okazaki Fragments	Okazaki Fragments are short discontinuous single-stranded DNA fragments synthesized from the 3′ ends of RNA primers and they make up the lagging strand of DNA.
Rnase H	Rnase H is the abbreviation for ribonuclease H.
3′OH function or structure	The 3′OH function or structure refers to the hydroxyl attached at atom position 3 usually of a ribose or deoxyribose sugar.
Endonucleolytic property	Endonucleolytic property refers to that of an enzyme capable of cleaving DNA at an internal position or nucleotide sequence.
Ligase	Ligase refers to an enzyme capable of joining a nick in DNA
5→3′	5→3′ is the shorthand notation which indicates the nucleotide sequence in a DNA or RNA strand or direction of synthesis of nucleic acid strands such as nascent cDNA and RNA primer.
Nick	A nick is a feature showing breakage or unjoined 3,5 phosphodiester structure in a strand of DNA or RNA.
Gap	A gap is a feature characterized by the removal of nucleotides usually by exonuclease action from one strand of double-stranded DNA or RNA.

or unravelling the origin into an “eye” or “bubble” (Figure 3.6). These proteins constitute a pre-priming complex (Funnell *et al*., 1987). DnaB protein is a helicase, requiring ATP to unwind the DNA bi-directionally (Baker *et al*., 1986). The unwound DNA is stabilized by the interaction of SSB protein (single-stranded binding protein). Unravelling of the DNA generates a positive supercoil which is relaxed to a negative (right-handed) supercoil by DNA gyrase. This allows replication to continue. The process of priming and DNA synthesis are followed once a replication eye is formed. The interaction of dnaB protein and DNA primase with proteins n (PriB), i(dnaT), n′′(PriC) and dnaC generates a mobile, multi-subunit primosome complex which is responsible for priming both continuous and discontinuous replication (Figure 3.6). The binding of the primosome complex at opposite ends of the replication eye is guided by protein n′(PriA). As dnaB protein unwinds the helix at the replication fork, an expanded replication eye develops, characterised by single-stranded regions which are stabilized by SSB protein (Baker *et al*., 1986; 1987). A swivel or negative supertwist that prevents re-annealing of the parental DNA strands (Baker and Kornberg, 1988) is formed by the gyrase. These single-stranded regions are the template to which RNA primers are synthesized by rNTP-requiring DNA primase. DNA polymerase III (pol III) uses dNTPs and Mg^{2+} to synthesize cDNA 5→3′ from the 3′OH function of the RNA primers, releasing pyrophosphate (PPi). DNA synthesis occurs in opposite directions. As the replication forks advance (Cairns, 1963a; b), newly exposed template requires priming by the primosome complex, followed by discontinuous DNA synthesis or generation of the lagging strand. In each direction, the original RNA primer forms the growth point for continued cDNA synthesis, generating the leading strand. This is the characteristic pattern of semi-discontinuous DNA replication. The short discontinuous DNA fragments, synthesized from RNA primers of the lagging strand, are called Okazaki Fragments (Okazaki *et al*., 1968). RNA primers may be removed by the exonuclease action of pol I, and digestion by Rnase H, followed by copy synthesis by the same polymerase to fill in the gaps, stopping short of the 5′ end of the next fragment. Using ATP, DNA ligase facilitates the closure of the nicks between 3′ and 5′ ends of adjacent fragments (Funnell *et al*., 1986). It is clear that the parental strand templates are each associated with a complementary

A



B



C

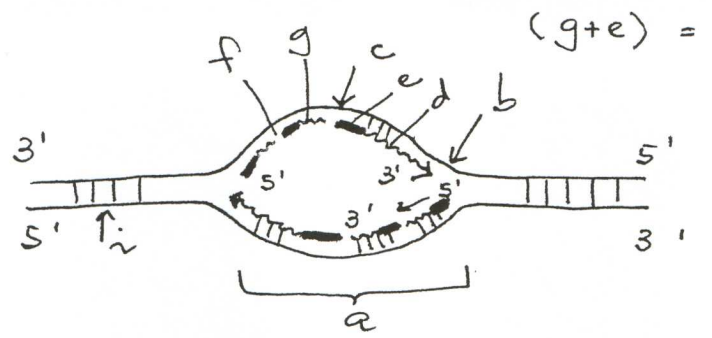
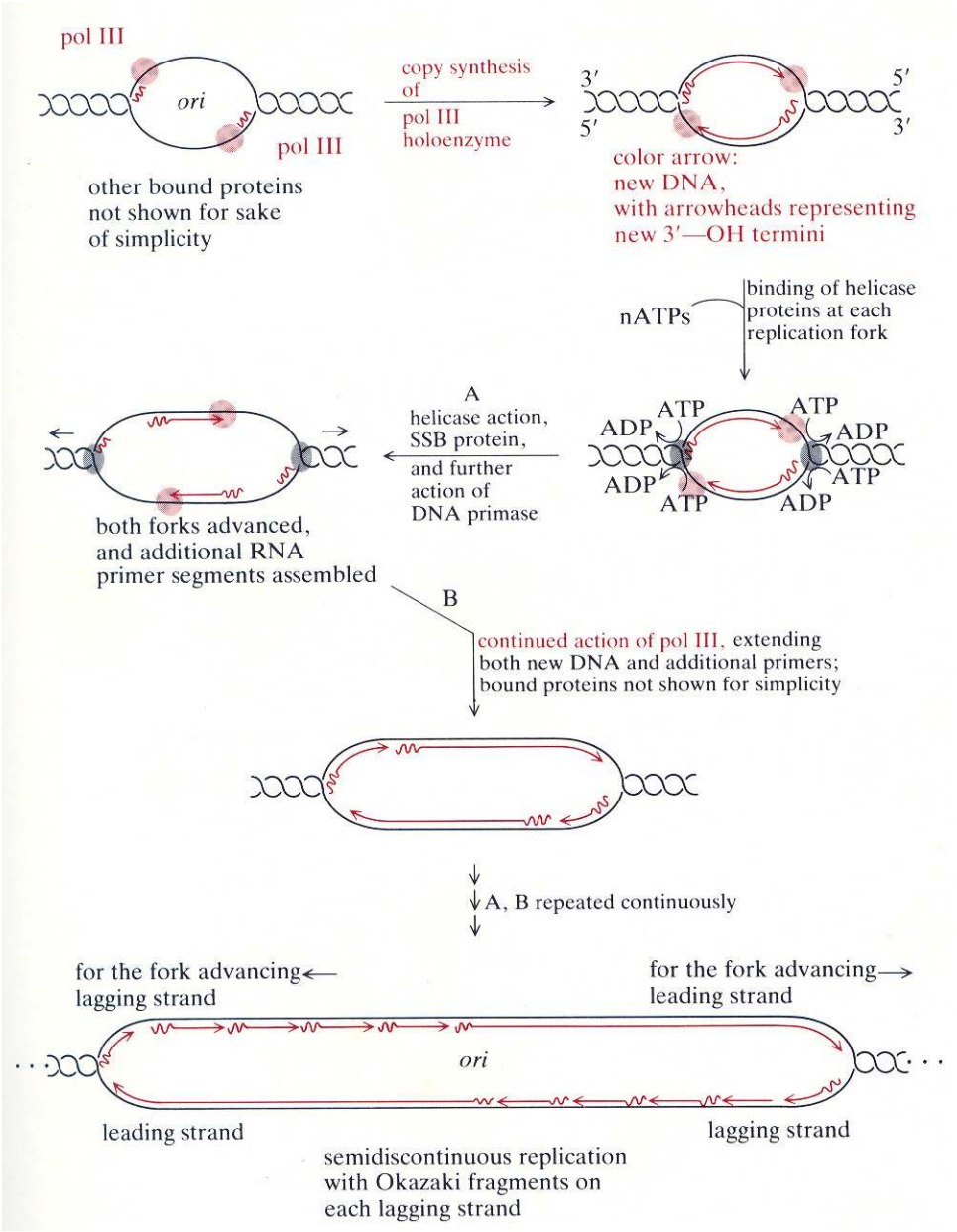


Figure 3.5 Labelled structural DNA replication intermediates.



Replication bubble model continued overleaf

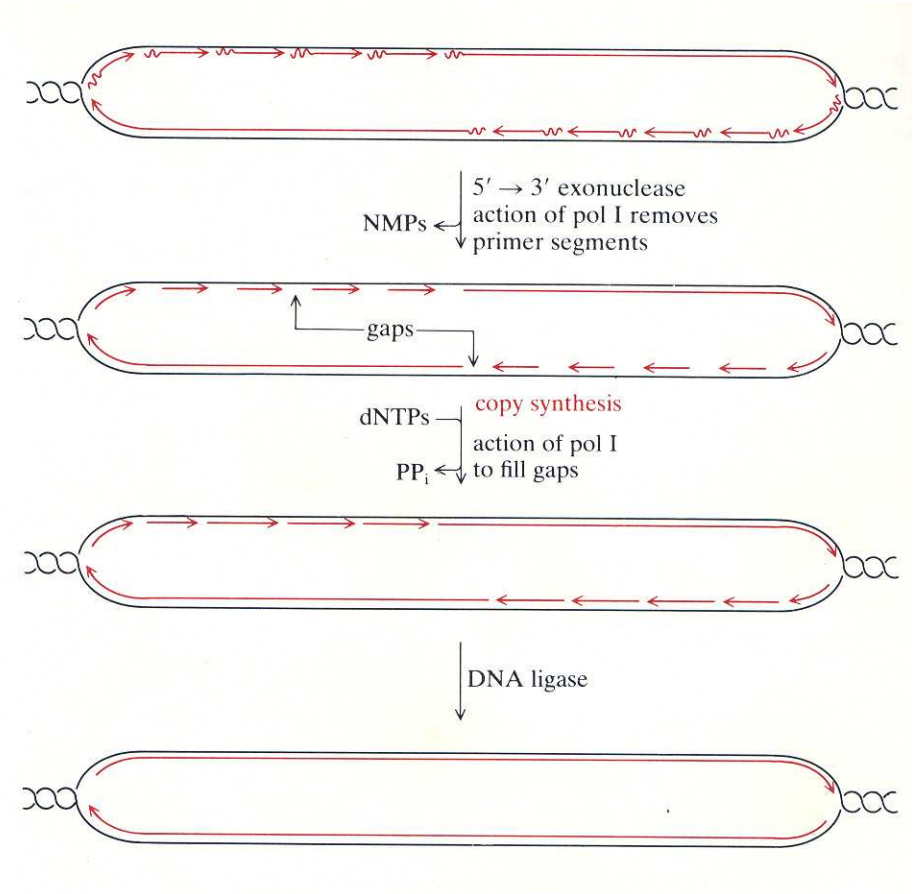


Figure 3.6 Flow diagram of a replication bubble model (Bohinski, 1987).

Reproduced with permission from Pearson Education Inc., Upper Saddle River, NJ, USA.

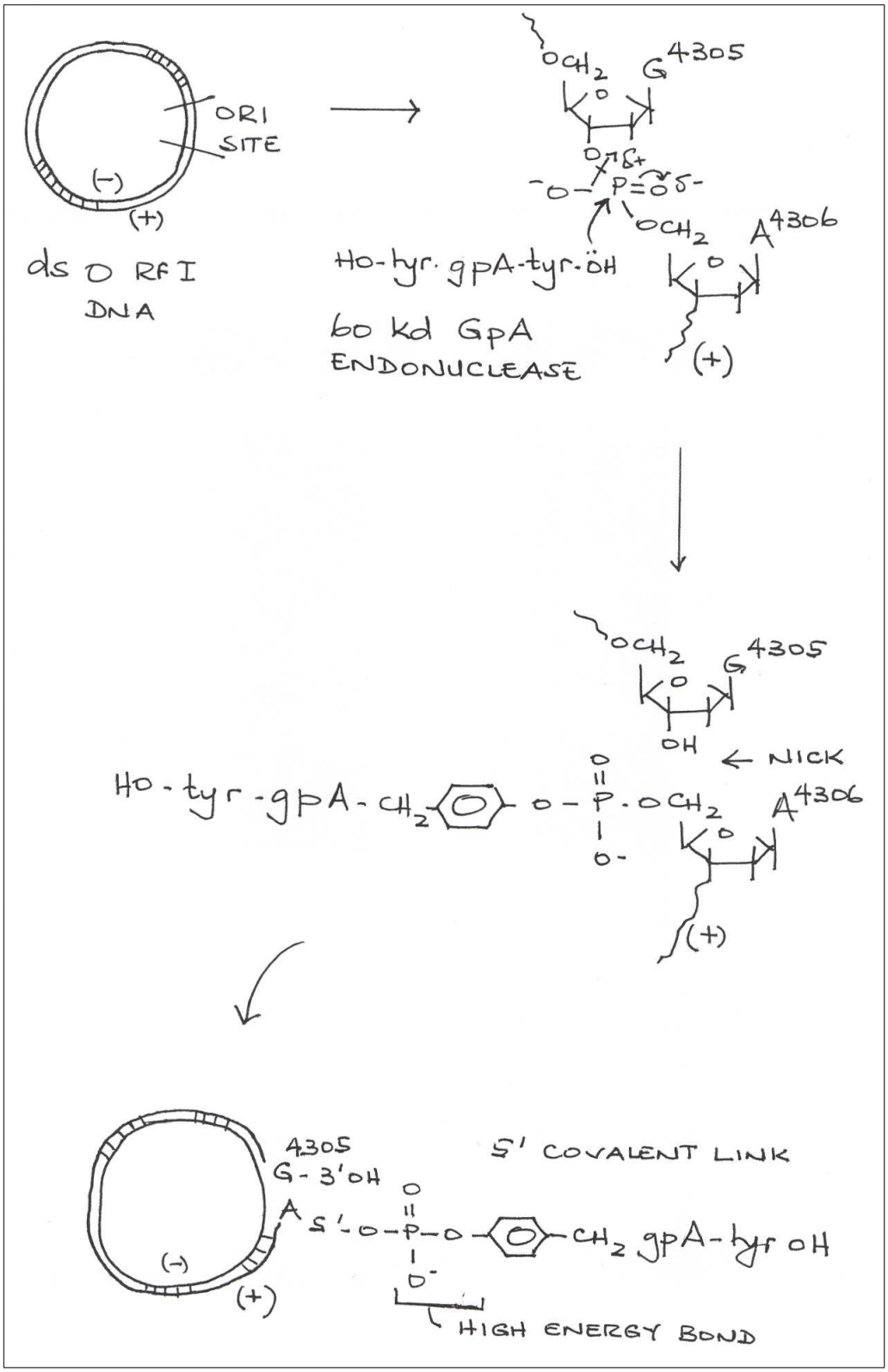
nascent strand of DNA, depicting the semi-conservative nature of DNA replication (Figure 3.4).

ΦX174 ds RF I→RF II replication model

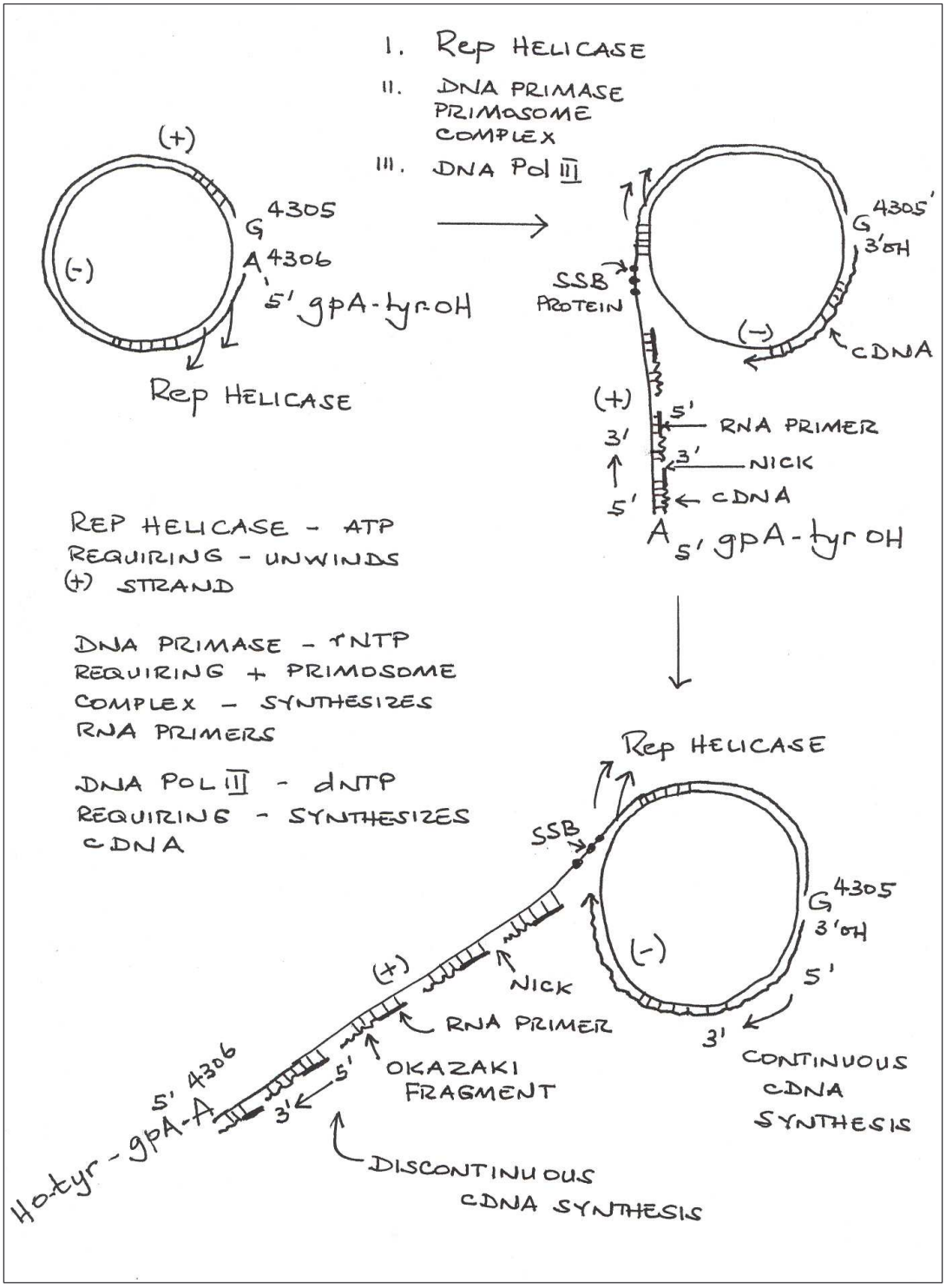
The following propositional knowledge may be used to evaluate students' responses on the ΦX174 ds RF I→RF II replication model. Specific propositional knowledge statements are also provided in Table 3.7 to assist with the definition of a range of symbolism encountered in DNA replication studies.

The native state genome of phage ΦX174 is single-stranded (ss) circular DNA of 5386 nucleotides (Sanger *et al*., 1978). This strand is called the "+" strand which replicates to form the " - " strand in bacterial hosts. Both strands generate the replicative double-stranded (ds) circular form (RF I) DNA. RF I molecules can replicate to form ds circular RFII molecules which will be the emphasis of this section (Figure 3.7). RF II DNA molecules can give rise to the single-stranded "+" DNA which is required for phage assembly (van Mansfield *et al*., 1986; Kornberg and Baker, 1992).

Phage ΦX174 gpA protein of 60kd is probably the only phage-encoded enzyme implicated in the replication of ΦX174 DNA. Enzymes, of the bacterial host *E.coli*, facilitate several steps in replication of the phage DNA. GpA protein has both endonucleolytic and ligase properties (van Mansfield *et al*., 1986). At the ori site, GpA protein introduces a nick on the "+" strand of negatively supercoiled ds RF I DNA. The nick results in a 3′OH function at a G residue at position 4305 and the high energy, covalent attachment of gpA via an active tyrosyl group to the 5′ phosphate moiety of an A residue at position 4306 (Eisenberg *et al*., 1977; Ikeda *et al*., 1979). The 3′OH function acts as a growth point for continuous 5→3′ cDNA synthesis facilitated by *E.coli* DNA pol III using the "–" strand as template (Figure 3.7). No RNA primer is required in this case (Ikeda *et al*., 1976; Kornberg and Baker, 1992). As the "+" strand is unwound by *E.coli* Rep helicase, SSB protein binds to it, preventing its reassociation with the "–"



Rolling circle model continued overleaf



Rolling circle model continued overleaf

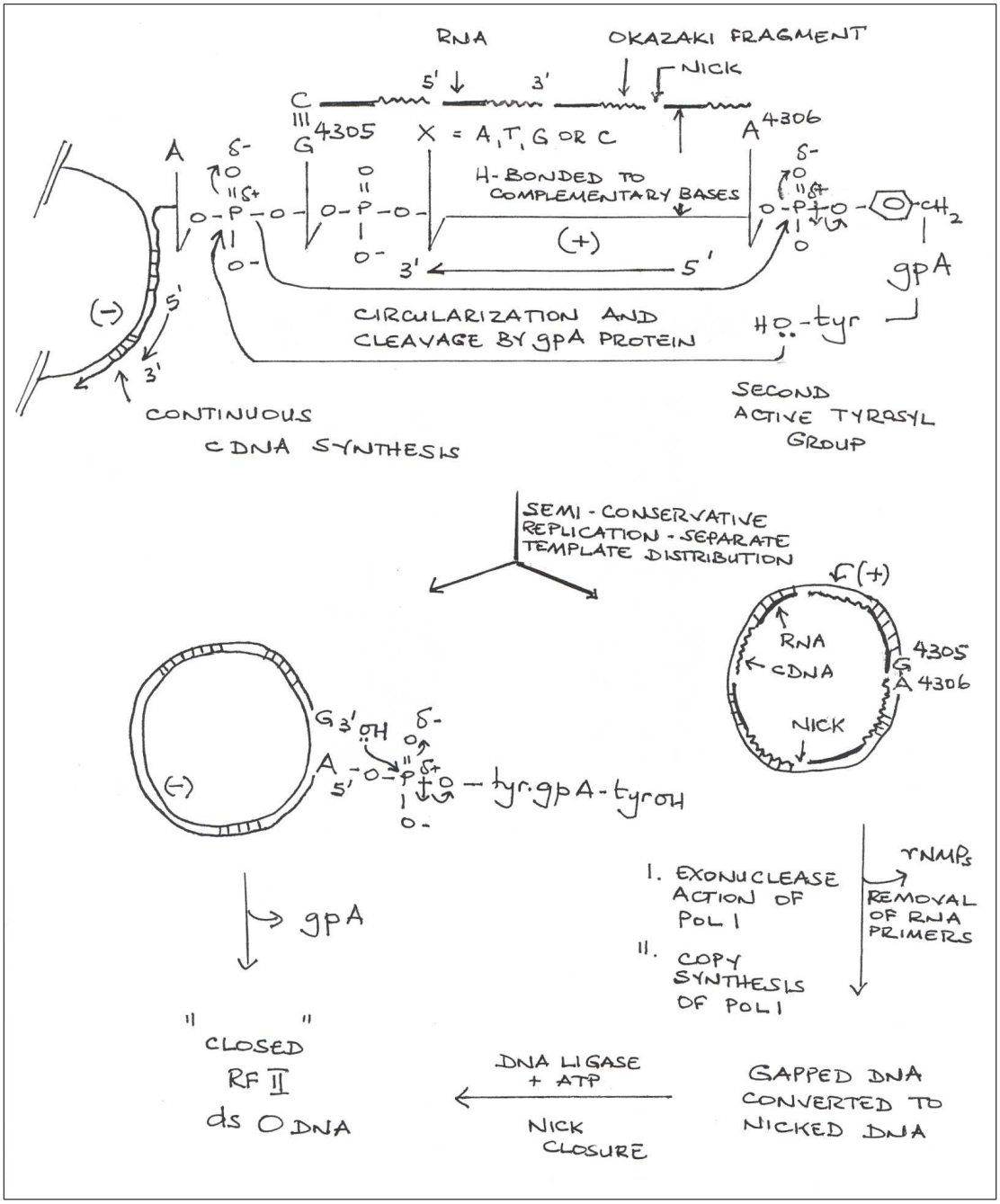


Figure 3.7 Flow diagram of a rolling circle model illustrating the replication of ΦX 174 RF I→ RF II DNA (Gupthar, unpublished original illustration).

strand template. *E. coli* Rep helicase requires two molecules of ATP for each base pair melted (Kornberg *et al*., 1978). The action of the helicase gives the effect of the "rolling circle" (Dressler and Wolfson, 1967; Gilbert and Dressler, 1968) as the "+" strand is unwound from the inner "–" strand circular template (Figure 3.7). As the length of the unwound plus strand increases, it requires priming by the *E.coli* primosome complex as described. Short RNA fragments are required for 5→3′ discontinuous cDNA synthesis opposite the plus strand template, generating Okazaki Fragments which are separated by nicks (Figure 3.7). Following synthesis of the nascent "+" strand around the "-" strand template, the original linearized "+" template circularizes and the second active tyrosyl residue of the 5′attached gpA protein cleaves the replication origin as shown (Figure 3.7) to release circular intermediates A and B. The energy generated by the cleavage is stored in the gpA-DNA intermediate A and is used to sustain a ligation between 3′OH function and the 5'phoshate group, displacing gpA protein (Figure 3.7) (van Mansfields *et al*., 1986; Kornberg and Baker, 1992). The RNA primer in intermediate B is excised by the template exonuclease action of Pol I, the gaps filled by Pol I - synthesis of DNA and nicks closed by DNA ligase as described in the replication bubble model. The original strands are each associated with nascent DNA in separate molecules A and B, a characteristic feature of semi-conservative DNA replication (Meselson and Stahl, 1958).

3.6 Summary and conclusions

The author provides propositional knowledge and concise tabulated statements (Garnett *et al*., 1985) to assist with the interpretation of students' responses to the various probes. The nature of student difficulties with symbolism will be assessed (Chapters 4 to 8) taking into account relevant detail from this compilation. The tabulated concise statements find use in assessing specific difficulties with symbolism. The detailed propositional knowledge is useful in assessing more elaborate presentations from the students. Apart from the factual information revealed in the compilation of propositional knowledge, the author searches for the students' understanding (Yarroch, 1985; Shulman, 1986) reasoning ability (Tsui and Treagust, 2003; Sins *et al*., 2005; Pata and Sarapuu,

2006), integration of domain-specific information (Novak, 1984; Wandersee *et al*., 1994; Herron, 1996) and an ability to communicate information (Zohar and Nemet, 2002; Yore and Treagust, 2006) on symbolism. Thus an evaluation of student learning difficulties becomes complex when learning patterns (Chapter1) are influenced by unique mental schemata and an imposing challenge of the nature and diversity of information that symbolism presents in Molecular Biology.

CHAPTER 4

DIFFICULTIES WITH POLYSEMIC SYMBOLS AND SHORTHAND NOTATION

4.1 Interpretation of the symbols "Δ"and ": :"

4.1.1 Introduction

The studies reported in Chapter 2 into the use of the symbols "Δ"and "**: :**" illustrate their polysemic nature and the challenge they impose regarding interpretation. It is expected that students understand their meaning in specific contexts and communicate information about them correctly. The visual cue (Gilbert, 2005) that symbols provide becomes meaningful when students draw appropriately on mental schemata to provide this communication (Taber, 2003; Yore and Treagust, 2006), demonstrating the attribute of being "visually literate" (Christopherson, 1997; Takayama, 2005; Mathewson, 2005; Schönborn and Anderson, 2006a). Learning information correctly for integration into mental schemata will be a prerequisite for such a demonstration (Thorndyke and Stasz, 1985). The expression of symbolic language requires clear identification and differentiation of symbol use, coupled with an understanding of the underlying information (Kozma, 2003; Treagust, 2003) that symbols depict. Hence the decoding of symbolism such as "Δ"and "**: :**" requires "bootstrapping" of the relevant knowledge (Cheng *et al*., 2001) when differentiating their use in specific contexts. Symbols such as "Δ"and "**: :**" are uncommon in textbooks (Chapter 2); however, there is sufficient evidence in other literature that their use is well-defined in various contexts (Chapter 3; Table 3.1). Examples of the varied use of symbols are presented by the students themselves (4.1.2). In this study, the use of written and oral probes reflects that, among some students, the understanding of symbol use can be at variance with acceptable propositional knowledge (Chapter 3; Table 3.1) on the subject. In this Chapter, the author presents two types of evidence: firstly, free-response data on students' most dominant interpretation of the symbols "Δ" and "**: :**" in any science context and genetic

engineering, respectively, and secondly, data from more focused probes on their interpretation of these symbols in specific scientific contexts, including Molecular Biology. In both cases, the data will be backed up by interviews that focus on the responses given by individual students to the free-response and focused probes. The rationale was to establish the students' understanding of polysemic symbol use in a general scientific context, coupled with specific use, differentiation and interpretation in more specific contexts of Science including Molecular Biology.

4.1.2 Variation in the interpretation of symbols

A superficial response, comprising four categories (Table 4.1), was obtained when students were asked to explain the meaning of the symbol "Δ" in a general scientific context (3.5.1; Table 3.1). Over sixty percent of the students referred exclusively to delta as a symbol depicting "difference" in the measurement of a parameter. Variations of the parameter included temperature, optical density, pH and force as indicated in Table 4.1. Some students provided more than one example, illustrating that the symbol finds use in the measurement of difference in various parameters. Examples illustrating their use of delta included;

i. $\Delta T = T_2 - T_1$ (T = temperature, ^oC)

ii. $\Delta T = T_{final} - T_{initial}$ where T refers to temperature

iii. $\Delta OD = OD_2 - OD_1$ (delta is the difference in optical density readings taken initially (OD_1) and subsequently (OD_2)

iv. ΔpH = difference in pH reading, for example, the addition of base can increase a second pH reading, therefore Δ pH = pH (after base addition) - pH (before base addition)

v. $\Delta F = F_A - F_B$, Δ is a difference in force between points A and B.

At least 30 percent of the students (n_{1-4}) (3.1) produced no written response to the free-response probe (Table 4.1). Uncommon references that delta depicts "hazard", "a triangle in mathematics", "heat" and "gene deletion" were also recorded at low incidence over 4

Table 4.1 Students' response on the meaning of "Δ" in a scientific context

	Student number (n_{1-4})			
	$n_1 = 54$	$n_2 = 35$	$n_3 = 31$	$n_4 = 44$
Categories				
Common responses				
Difference in parameter:				
temperature	15	18	14	15
optical density	6	15	4	12
pH	6	1	4	4
force	2	1	1	3
Uncommon responses				
hazard	0	0	1	0
triangle in mathematics	1	0	0	0
heat	2	0	0	0
gene deletion	0	1	0	3
Erroneous responses				
ΔOD = total optical density	0	1	1	1
Δtemperature = total temperature	0	1	0	0
No response	22	13	6	9

years (Table 4.1). From the above, it is clear that students reflect on prior knowledge derived from a range of subject areas. References to "difference in force as depicted by ΔF", "triangle in mathematics", "heat" or "gene deletion" are encountered in Physics, Geometry, Chemistry and Biotechnology or Genetics, respectively. Further deliberation on the use of "Δ" to reflect placement of double bonds, for example in steroid structures, or its use to show proton gradient (3.5.1) was not recorded. Instead, erroneous responses were recorded which indicated that ΔOD and ΔT represented "total optical density" and "total temperature", respectively (Table 4.1). Further insight into these interpretations was obtained by interviewing students (4.1.4).

As regards the symbol "**: :**", the free-response probe (3.5.2) elicited responses which could be placed in four categories (Table 4.2), namely gene fusion symbol, gene integration symbol, no response and inappropriate responses. Over seventy percent of all students (n_{1-4}), associated the symbol "**: :**" with gene fusion and relatively few gave information on its combined use to indicate gene fusion and integration. Over thirty percent of all students referred to "**: :**" as a symbol depicting gene integration and approximately seven percent showed no response. Of interest, few students provided inappropriate responses which required further investigation to achieve clarity on the meaning of their responses. As regards the use of a symbol to indicate gene fusion, there is currently a lack of standardization concerning the choice of symbol (2.5.2). Since "**: :**" finds use to describe both gene fusion and gene integration, its use must be clearly defined in both contexts (Table 3.1). Therefore, a focused probe (3.5.2) was designed to collect information from the students (4.1.5) regarding its use in specific gene constructs. Rather obscure and inappropriate responses were also recorded (Table 4.2) using the free response probe where the symbol "**: :**" was defined as " a gene separator", "chromosome component" and "electrons between genes". Students also referred to the symbol as one which showed "bonding of genes" and "gene linkage".

Table 4.2 Students' response on the use of the symbol "**: :**" in genetic engineering

	Student number (n)			
	$n_1 = 54$	$n_2 = 35$	$n_3 = 31$	$n_4 = 44$
Categories				
Gene fusion symbol	39	30	22	30
Gene integration symbol	15	13	11	16
No response	7	0	4	3
Inappropriate responses				
Gene separator	1	0	0	0
Chromosome component	2	0	0	1
Electrons between genes	0	2	1	1
Bonding of genes	0	0	3	1
Gene linkage	0	0	0	2

4.1.3 "Δ" in different contexts

The more focused probe (3.5.1) analysed the students' interpretation of the symbol delta in different contexts since the meaning of this symbol can be different or polysemic. Polysemy requires students to carefully find use of a word or symbol in different contexts, without compromising the meaning of either in a particular context. The findings are presented in Table 4.3 and investigated further using interviews (3.4.1; 4.1.4). With reference to the use of delta alongside the parameters, for example Δ T (^{o} C) or ΔOD, the large majority of students indicated that delta denotes a measure of difference in temperature or optical density, respectively (Table 4.3). Consistent with the findings of the free-response probe, a few students reported that delta meant "total temperature" and "total optical density" in the context used. This was expressed at low incidences of 5 and 3 percent in groups n_2 and n_3, respectively (Table 4.3). Among all respondents of groups n_1 - n_4, only one student referred incorrectly to delta as a symbol denoting "speed of the reaction" in the thermal denaturation reaction of DNA (Kornberg and Baker, 1992; Stryer, 1995). All other students indicated correctly that delta meant "heat" (3.5.1) in the reaction showing the conversion of double-stranded DNA to single-stranded DNA (Table 4.3). The parameter Δ*leu* elicited varied inappropriate responses (Table 4.3); however, at least 75% of all the respondents (n_{1-4}) correctly indicated that Δ*leu* denoted deletion of the leucine biosynthesis gene. In the design of this probe, any reference to gene designation or allele or locus or laboratory number (3.5.1) was excluded in order to simplify the format for students. This simplified format was routinely presented in class among other variations bearing the above. Examples of the varied inappropriate responses, which were recorded at low incidences (Table 4.3), included "changes to leucine", "hazard to leucine", "change in concentration of leucine", "different isomer of leucine", abbreviation for isoleucine" , "purest form of leucine" and " area of graph of leucine". It is apparent, from some of these responses, that the meaning of the symbol delta has been inappropriately transferred in order to explain this parameter. Further, there is failure to associate the abbreviated italics, *leu* with the nomenclature of a specific gene marker, let alone provide its function (3.5.1). Instead, since delta also represents "hazard" and "difference in the measurement of a parameter",

Table 4.3 Students' response on the meaning of "Δ" in different contexts

Parameter	Type of response and incidence
1. ΔOD	difference in optical density
	n_1 = 50/54 n_2 = 34/35 n_3 = 29/31 n_4 = 43/44
	total optical density
	n_1 = 0/54 n_2 = 1/35 n_3 = 1/31 n_4 = 1/44
	no response
	n_1 = 4/54 n_2 = 0/35 n_3 = 1/31 n_4 = 0/44
$\overset{\Delta}{\rightarrow}$	
2. dsDNA → ssDNA	delta refers to heat
	n_1 = 54/54 n_2 = 35/35 n_3 =30/31 n_4 = 44/44
	delta refers to speed of the reaction
	n_1 = 0/54 n_2 = 0/35 n_3 = 1/31 n_4 = 0/44
3. Δ *leu*	deletion of a leucine biosynthesis gene
	n_1 = 44/54 n_2 = 23/35 n_3 = 25/31 n_4 = 34/44
	hazard to leucine
	n_1 = 1/54 n_2 = 0/35 n_3 = 1/31 n_4 = 1/44
	purest form of leucine
	n_1 = 2/54 n_2 = 1/35 n_3 = 1/31 n_4 = 1/44

Table 4.3 continued

	abbreviation for isoleucine			
	n_1 = 0/54	n_2 = 1/35	n_3 = 0/31	n_4 = 0/44
	different isomer of leucine			
	n_1 = 0/54	n_2 = 0/35	n_3 = 2/31	n_4 = 1/44
	area of graph of leucine			
	n_1 = 1/54	n_2 = 0/35	n_3 = 0/31	n_4 = 0/44
	change to leucine			
	n_1 = 4/54	n_2 = 1/35	n_3 = 2/31	n_4 = 3/44
	no response			
	n_1 = 2/54	n_2 = 9/35	n_3 = 0/31	n_4 = 4/44
4. Δ T (^{o}C)	change in temperature			
	n_1 = 54/54	n_2 = 34/35	n_3 = 30/31	n_4 = 44/44
	total temperature			
	n_1 = 0/54	n_2 = 1/35	n_3 = 1/31	n_4 = 0/44

students inappropriately associated “hazard” or “change” to leucine with the gene marker notation presented (Table 4.3). The inappropriate transfer of knowledge (Grayson, 1995) might be influenced partly by the polysemic nature of this symbol; however, it is imperative that the meaning of the symbol be understood in a specific or unrelated context. In order to find plausible explanations for the other inappropriate responses (Table 4.3), further information had to be obtained from interviews (4.1.4).

4.1.4 Inappropriate responses using interviews

The following excerpts of students’ dialogue from selected interviews allow for further comment on the nature of inappropriate responses (4.1.2; 4.1.3) and clarification of the meaning (Table 3.1) of delta as a symbolic representation (Treagust *et al*., 2003; Yore and Treagust, 2006). During the course of analyses, various categories of difficulties were identified based on a dominant type of difficulty. There could be deeper underlying factors, beyond the scope of this study, for example psychological factors, which could contribute to learning difficulties. Superficially, the obvious difficulty is listed as a category, with supporting interview transcripts and integrated comment on each as follows.

4.1.4.1 Inapt mental schemata

The tendency to draw from mental schemata (Nurrenbern, 2001), possibly developed early in life or from daily life experiences (Kasanda *et al*., 2005), can manifest in a manner of communication which is inappropriate, erroneous or unacceptable when facets of inappropriately transferred information tend to impact on scientific interpretations. Student S1 indicated that the symbol delta could depict “hazard” (Table 4.1) as indicated by a road sign board and by a triangle found on an automobile dashboard switch. This response emerged in the general scientific enquiry (3.5.1) on symbol use. It would

therefore appear as a mental model conjured through life experiences which the student reflects upon. Interview 1 presents the following supporting dialogue.

I: I noticed that you indicated the symbol delta to represent hazard.
S1: Yes.
I: Where did you come across this fact?
S1: Well, Sir [pause], you can find it on a triangular road sign showing cattle crossing. Also you see on an emergency flasher switch in a car.

In another response to the probe on symbol use in a general scientific context (3.5.1), student S2 gave delta as a symbol depicting heat. Observations in daily life can also impact on the interpretation of scientific information including symbolism in an idiosyncratic manner. This is illustrated in Interview 2 as follows.

I: Where did you first come across the symbol delta to represent heat?
S2: In school, eh secondary school. Probably in standard 6 or 7.
I: Really! That's a long time ago. How do you remember that the symbol delta is used to show heat?
S2: I think that a triangle relates to the tripod stand, you know the triangular base and Bunsen burner.

Student S2 acknowledged that he had learnt at secondary school level that delta was the symbol for heat. Further, he associated a triangle with "heat" simply because it related to the triangular base of a tripod stand and a Bunsen burner. The construction and integration of knowledge concerning "heat" and "the symbol delta" are not only idiosyncratic (Thorndyke and Stasz, 1985; Pearsall, 1997) but influenced both by enactive and iconic learning principles (Bruner, 1960; 1973; cited by Cooper, 2005; Georghiades, 2004). The enactive would be knowledge that a Bunsen burner generates heat when lit while the iconic would be the visual impact of the tripod base and the flame.

4.1.4.2 Poor differentiation

Some students may show difficulty expressing the measure of difference or change in a parameter (Mammino, 2001). Such a difference or change may be illustrated using a graphical plot to show an increased or decreased second value. Students do not view an "increment" as shown using delta (Table 4.3) but erroneously interpret it as a total measure of an initial reading plus the increment or difference. Students' failure to differentiate entropy (S) from entropy changes (ΔS) or temperature (T) from temperature changes (ΔT) has already appeared in the literature (Mammimo, 2001). Poor differentiation of the difference in optical density and the erroneous formulation of the concept of "total optical density" are illustrated in the following excerpt of Interview 3. This follows the earlier analysis of responses to the focused probe (Table 4.3) on the use of delta in different scientific contexts (3.5.1).

I: Can you tell me why delta OD means total optical density as you have it?
S3: When you measure a second absorbance or OD value and it is different from the first, then you add the difference to the first value to give you total OD.
I: How do you show the difference, assuming the second value is greater than the first?
S3: [silent for a while] I think I learnt this incorrectly.
I: Now, take a look at ΔT where you indicate it means total temperature.
S3: Yes, Yes. That, [pause] that I must look at again.

4.1.4.3 Inappropriate word association

Word association is a phenomenon which is commonly associated with concept mapping (1.4.3) as single words can act as "cues" and prompt a re-call of information from mental schemata (White and Gunstone, 1992; Van Zele *et al*., 2004). The inappropriate association or linkage of information from mental schemata can be expressed in written or oral communications as shown below.

I: Why have you indicated that Δleu is the purest form of leucine?

S4: When something [pause] like a chemical is pure, you show purity with a triangle.

I: My friend, where did you see this type of information?

S4: In our laboratories [pause] you see "Analar" pure reagents showing a triangle or "A"

I: Really?

S4: It's true, Sir.

On examination of several reagent bottles marked "Analar", none showed a "Δ" or an "A". Speculatively, the information presented may have been an association of the symbol type "Δ" with an apex of an "A". Further, the student found a link between "purity" of a chemical and a trade name ("Analar") suggesting that word association (White and Gunstone, 1992) could provide a cue in the construction of this response. Alternate mental schemata (Taber, 1998) and the inappropriate transfer of information may be coupled with word association.

4.1.4.4 Constructivism

Students may find unique ways of constructing knowledge and further express this in a manner which is scientifically unacceptable. This manner of learning and expression of knowledge are explained by the theory of constructivism (Driver and Bell, 1986; Von Glasersfeld, 1992; Kogut, 1996). The following interview illustrates constructivism as the student constructs knowledge on delta which is used in the focused probe to depict deletion of the leucine biosynthesis gene (3.5.1; Table 3.1). Reports on this example of constructivism cannot be found in the literature. The student refers to Δ*leu* as the short form of isoleucine, indicating erroneously that the triangle nomenclature influences the abridged name of isoleucine.

I: Can you explain the fact that you show Δleu as an abbreviation for isoleucine?

S5: You see "leu" means leucine. [pause] The triangle is " iso" because two sides are equal. That is why it must be isoleucine in short.
I: What do you mean, my friend?
S5: I forget the full name. Iso [pause], Iso [pause]
I: You mean isosceles triangle?
S5: Yah! [Yes!]

4.1.4.5 Linguistic difficulties

It would appear that the meaning of Δ*leu* (Table 4.3) is misinterpreted entirely due to linguistic difficulties (Simonneaux, 2000;Yore and Treagust, 2006) and a failure to recognize that the italic abbreviation denotes a gene marker. The following dialogue presents supporting evidence although it is extremely difficult to find supporting literature which reports linguistic difficulties linked to misinterpretation of a symbol such as delta. There appears to be confusion in the student's understanding of the term "difference" which can be indicated by delta but in this context its placement in the form "Δ*leu*" indicates deletion of a leucine biosynthesis gene (Table 3.1). The italics "*leu*" should ideally provide a cue that this refers to gene nomenclature; however, the student confuses the usage of terminology to associate "difference" with "different" and incorrectly interprets "*leu*" as an abbreviation for leucine. The coupling of such misinterpretation yields "different leucine", which the student expresses as "different isomer of leucine".

I: Tell me more about the abbreviation Δleu which you indicate as "different isomer of leucine".
S6: Well, delta implies "difference", so I think Δleu is like a different type or isomer of leucine.

4.1.4.6 Inappropriate information transfer

Inappropriate information transfer is illustrated in the following interview where student S7 draws information from an unrelated series of lectures presented by the author. It is appropriate that students be guided to retrieve information from appropriate knowledge domains (Chapter 8) and reflect correctly on taught lecture content. The inappropriate response from student S7 is shown below.

I: I see that you refer to Δleu as the "area of the graph of leucine". What do you mean by this?
S7: You see when leucine can be separated by HPLC you get a peak. You can draw a triangle fitting under the peak with dT_2, dT_3 and dT_4.
I: Yes, yes. Go on, tell me more.
S7: The area of the triangle can give you an idea of the concentration of leucine. But you need a standard of known concentration to compare area and concentration.

Student S7 refers to oligonucleotide analyses where digestion products of the oligomer, dT_4 may be subject to HPLC analyses. Peak area integration, which can be achieved using an integrator-plotter facility, could be used to estimate concentration equivalents or a ratio of various separated components (Gupthar, 1983). The information drawn from this example is one which reflects inappropriate transfer of knowledge from aspects of work received in a different section of the course (Salomon and Perkins, 1989; Grayson, 1995).

4.1.4.7 Conjecture

The meaningful interpretation and integration of knowledge into mental schemata (Thorndyke and Stasz, 1985) are often seen in the quality and depth of student communications. Where such knowledge is deficient, this manifests poorly in the type of response students present when challenged metacognitively (Zohar and Nemet, 2002;

Georghiades, 2004). The following response illustrates an example of conjecture or guessing (Yore and Treagust, 2006).

I: Why have you shown Δ to mean speed of the reaction where double-stranded DNA is converted to single-stranded DNA?
S8: I did not know. I had to guess that it might be speed or something.

4.1.5 ": :" within two gene constructs

Using a focused probe (3.5.2) the students were challenged with differentiation of the meaning of "**: :**" within two gene constructs. The first illustrated "gene integration" by the transposable element Tn5, inactivating *lac(z)* as a consequence of transposition (2.5.2; Kadonaga *et al*., 1987). The second construct illustrated "in tandem gene fusion" which generates an open reading frame, expressing both gene functions in a single translation product (3.5.2). The presentation of a more specific probe elicited an improved response from the students. Students were able to comment in larger numbers on the specific use of "**: :**" in both transposition and gene fusion (Table 4.4) than when requested to write about the symbol's use in genetic engineering (Table 4.2). The exact trend was recorded with respect to inappropriate responses when comparing the use of the free-response and focused probes (Tables 4.2 and 4.4).

4.1.5.1 More evidence of combined difficulties

The following interviews were absolutely essential in providing supplementary information to assist with the analysis of inappropriate student responses about the symbol "**: :**". The analysis of the meaning of the double colon indicated gross misunderstanding of symbol use in specific contexts. The following interview illustrates misunderstanding of symbol function and substitution of information (Grayson, 1996; Pittman, 1999) that is inappropriate.

Table 4.4 Students' response on the use of the symbol "**: :**" in different gene constructs

	Student number (n)			
	$n_1 = 54$	$n_2 = 35$	$n_3 = 31$	$n_4 = 44$
Categories				
Gene integration symbol	25	18	18	24
Gene fusion symbol	39	30	22	30
No response	3	0	2	3
Inappropriate responses: prototype definitions				
Gene separator	1	0	0	0
Chromosome component	2	0	0	1
Electrons between genes	0	2	1	1
Bonding of genes	0	0	3	1
Gene linkage	0	0	0	2

I: Are you familiar with these notations? [Interviewer presents the notations lac (z) :: Tn5 and 5′operator-promoter-GENE1::GENE2-terminator sequence-3′as shown in the focused probe (3.5.2)]

S9: Yes. They are genes [pause] eh, in the first case showing transposon action and the second means fusion.

I: Tell me more about these genes.

S9: Lac (z) stands for beta-galactosidase biosynthesis and Tn5 is a transposon. The other genes are not described except the operator, promoter and terminator sequences.

I: What is a transposon?

S9: Aah, it's a hopping gene. It can move from one place in DNA to another place in the DNA.

I: I noticed that you labelled the symbol "::" as a gene separator in both examples.

S9: Yes, Sir. You see a transposon like Tn5 can move next to the lac (z) gene so we separate the two with this symbol.

I: Let's look at the second example. You did say earlier that this is gene fusion.

S9: Yes, Sir.

I: But what does this symbol mean in this example?

S9: You see two genes are fused but we use this symbol like a separator to show that they are separate genes now linked.

The interview above reveals misunderstanding of symbol use in two contexts. Firstly, the student correctly identifies a transposon and the *lac (z)* gene, however, fails to indicate that the transposon integrates within the domain of the *lac(z)* gene, thereby inactivating it (3.5.2; Table 3.1). While identifying both genes in the first example, the student refers to the notation "*lac(z): : Tn5*" as one which shows "transposon action", which is restricted to a "hopping" action and its movement "next" to the *lac(z)* gene. Given the incorrect explanation of transposition in this context, the student erroneously substitutes (Grayson, 1995) the symbol "**: :**" to mean " separator" of the genes. In the second example, the student correctly identifies that genes 1 and 2 are fused (3.5.2; Table 3.1) or "linked" but maintains that the "separator" symbol is used to show the genes separately. While symbolic representations and shorthand notations carry a wealth of information, these

must be learnt with an understanding in particular contexts (Kozma and Russell, 1997; Ferk *et al*., 2003). The identification and meaning of all sub-components of shorthand notation require special teaching emphasis as students or novices do not always fully understand these as assumed by the teacher or expert (Treagust *et al*., 2003; Oh, 2005).

An inappropriate transfer of information and surface level understanding of the facts (Chi *et al*., 1981; Kozma and Russell, 1997; Schönborn *et al*., 2002a) are evident in the following interview.

I: I noticed that you referred to the symbol ": :" as a chromosome component in both these gene notations. [Interviewer refers to the constructs given in the focused probe (3.5.2)]. Could you explain the meaning of the symbol in these gene notations?
S10: You see in the first example, [pause] the transposon which is a hopping gene has linked next to the lac(z) gene, so the symbol shows like a protein joiner in a chromosome. You see the same thing in the second example where the joiner is holding two genes.
I: Tell me, are there other proteins that you know of [pause] eh, found on the chromosome?
S10: Yah [Yes], you see in diagrams [pause] like a dot for the centromere protein [pause] that is like a joiner of the chromosome to the mitosis spindle.

There are indications, from the excerpt above, that student S10 has not understood the mechanism of transposition but acknowledges that genes 1 and 2 are held together. The term "gene fusion" was absent in this response. The symbol "**: :**" is perceived to be a "protein joiner" linking genes in a chromosome. Further, the student compares the double colon with a "dot" representing the centromere within the chromosome structure. The "dot" is perceived to be a protein component of the chromosome, namely the centromere, which joins the chromosome to the mitotic spindle. The student infers, from a diagrammatic feature, that the double colon represents a protein joiner of genes in the constructs shown. While diagrams can be simply illustrated to show structural features of a chromosome or the process of chromosome recombination or segregation, they must be skilfully interpreted to promote learning (Cho *et al*., 1985; Kindfield, 1994).

Surface-level understanding and poor interpretation of symbol use are evident in the following response from student S11.

I: Tell me more about these gene constructs [Interviewer refers student S11 to the focused probe (3.5.2)].
S11: The first example shows a transposon or hopping gene joined to lac(z) by a joiner [pause] I mean chromosome joiner, here again in the second example there it is joining two genes.
I: What is a joiner, my friend?
S11: It is protein found in the chromosome and joins genes like this [points to examples given in probe].
I: Where did you get this information?
S11: I read that DNA is associated with protein that is why genes are nucleoproteins. Proteins join all genes together.

Student S11 defines the symbol "**: :**" as a proteinaceous chromosome joiner, linking genes. The concepts of gene integration and gene fusion were not mentioned during the course of the interview. The student identifies protein as a component of DNA and indicates that "proteins join all genes together". While it is generally true that DNA is found as a nucleoprotein complex, this example clearly illustrates the tendency of a student to "go beyond the information" at hand (Simonneaux, 2001; Bruner, 1973, cited by Cooper, 2005) and present facts that are incorrect (1.5).

Student S12 demonstrates surface-level understanding and poor differentiation ability during the course the following interview. The latter could well manifest as a consequence of surface-level understanding (Chi *et al*., 1981; Kozma and Russell, 1997; Schönborn *et al*., 2002a).

I: Does this symbol mean gene linkage in both examples? [Interviewer refers to the focused probe (3.5.2) and S12's written response]

S12: Yes.

I: What is Tn5? And what does it do?

S12: It is a gene that can move from one area of DNA to another. Here you see it linked to the lac(z) gene.

I: Okay. Tell me more about genes 1 and 2 in the second example.

S12: Here you see two genes linked to each other.

I: Take a look at this double colon. What does it mean?

S12: It shows linkage.

Student S12 identifies the double colon as a symbol showing linkage between genes. The student has some idea of the property of transposon Tn5 but fails to comment that it integrates within the domain or sequence of the *lac(z)* gene thereby inactivating it (Table 3.1). Student S12 is unable to differentiate a notation which shows transposition from one which illustrates in tandem gene fusion, presumably because the same symbol is used for more than one purpose. Surface-level understanding (Chi *et al*., 1981; Kozma and Russell, 1997) is evident as the student makes the observation that the respective genes are simply next to each other in both constructs and therefore must be linked.

The following interviews illustrate inappropriate information transfer coupled with poor understanding of symbol use. Both interviews involving students S13 and S14 show the inappropriate transfer of information from bonding theory which is introduced in level 1 Chemistry modules (3.1).

I: Are you familiar with these notations? [Interviewer refers student S13 to the focused probe (3.5.2.)]

S13: Yes. This lac(z): : Tn5 is a transposon bonded to the lac(z)gene and the second one [pause] shows a bond between two genes.

I: Why do you say the genes are bonded?

S13: This symbol [points to "::"] shows the electrons involved in the bond.

I: Tell me, where did you come across this symbol to show such a bond?

S13: Atoms can share electrons like this in a bond, [pause] so genes also follow this pattern.

The notation which illustrates transposition and the resultant inactivation of the *lac(z)* gene (3.5.2) were poorly understood by student S13. Further, gene fusion was interpreted as "a bond between two genes" because "electrons" between the genes were shared in the same way as done by atoms in a covalent bond. Symbolically, the double colon does show resemblance to conventional Lewis dot diagrams which illustrate covalent bonding involving the sharing of electrons between atoms (Peterson and Treagust, 1989; Peterson *et al*., 1989; Birk and Kurtz, 1999). This is another example of inappropriate transfer of information, given that such symbol use is encountered in Chemistry. Often the sub-microscopic such as atoms and electrons are shown by symbolic representations to give meaning of a process or arrangement of a molecule (Treagust *et al*., 2003; Ferk *et al*., 2003). The visual role of such representations must be understood in a particular context (Schönborn and Anderson, 2006a; 2006b); however, the information they depict to the learner might be used inappropriately to explain a different observation (Treagust *et al*., 2003; Mathewson, 2005).

Student S14 gave the following response on the use of the double colon.

I: Are you able to tell me why you refer to this symbol [pause] as a bond between genes in both examples? [Interviewer refers S14 to the gene structure mentioned in his response to the focused probe (3.5.2)]

S14: This symbol is used to show bonds [pause]. Sometimes you see only a colon in non covalent bonds.

I: Really? Give me an example. Can you write it down?

S14: In protein synthesis, you can show a non covalent link [pause] like the way you showed us [pause] I mean a bond like this.

S14 writes: "aa-tRNA: enz."

I: What is this, my friend?

S14: Aminoacyl tRNA to which the specific aminoacyl tRNA synthetase is non covalently attached.

I: Okay. Can you show me a covalent bond?
S14 : Say A and B are atoms sharing electrons, then we show it like this [pause].
S14 writes: "A: :B".
I: Thank you.

This interview reveals inappropriate information transfer which resembles the response obtained from student S13. The notations illustrating transposition and gene fusion (3.5.2) are interpreted as "bonds between genes" in the same way that Lewis dot diagrams show covalent bonds (Peterson and Treagust, 1989; Birk and Kurtz, 1999). The use of a single colon to show non covalent association of aminoacyl tRNA to a specific aminoacyl tRNA synthetase is uncommon in various Biochemistry textbooks except one (Garrett and Grisham, 1995, page 1028). The author acknowledges the use of a single colon or dash in previous lectures to illustrate this particular non covalent association. This might be the source of the information inappropriately transferred.

4.1.6 Discussion

Virtually all the reported difficulties concerning the use of delta (4.1.2; 4.1.3; 4.1.4) are classified at level 2, or suspected, in accordance with the four-level classification scheme described by Grayson *et al*. (2001). These were anticipated from several years of teaching Molecular Biology at the University of Durban-Westville, SA. On examination of the literature, there appeared to be limited information on the misuse of the symbol delta and the nature of similar difficulties experienced by students based elsewhere. Mammino (2001) reports that students at the University of Venda, former apartheid homeland of Venda, SA, were unable to clearly differentiate temperature (T) from temperature changes (ΔT) or entropy (S) from entropy changes (ΔS).

The conjecture or assumption (Yore and Treagust, 2006) made by student S8 that delta means "speed of the reaction" as shown in the denaturation of double-stranded DNA was unexpected and is therefore classified at level 1. The inappropriate transfer and

application of knowledge (Grayson, 1995; Schönborn *et al*., 2002a) that Δ*leu* denotes "area of graph of leucine" with respect to the determination of concentration was also recorded for the first time. This difficulty is also classified at level 1. Both these difficulties, therefore, require extensive further research in multiple teaching contexts (de Jong *et al*., 2000; Rhemtula and Rollnick, 2002; Darby, 2005) in order to be classified higher on the framework.

The nature of difficulties (Tables 4.2 and 4.4) reported with the regard to use of the symbol "**: :**" (4.1.2; 4.1.5) can be regarded as "anticipated" (Grayson *et al*., 2001) on the basis of teaching experience. It remains classified at level 2 on the classification framework described by Grayson *et al*. (2001). Currently, there is no similar report on the difficulties associated with the use of this symbol to indicate gene integration and gene fusion, respectively. Also experts have not standardized the use of a particular symbol to indicate in tandem gene fusion although limited supporting literature has been presented which shows use of the double colon in this regard (2.5.2). When used, it is imperative that the double colon be clearly defined to indicate in tandem gene fusion as the same symbol can also be used to indicate transposition (3.5.2). The examples of inappropriate information transfer tend to manifest in at least 5 sub-categories (Tables 4.2 and 4.4) which are also unique to this study. Although the reference to "gene linkage" (Tables 4.2 and 4.4) was recorded in two responses from student group n_4 only, it is to be classified at level 2 (Grayson *et al*., 2001) as this writer and teacher of Molecular Biology can attest to its prevalence over the last 15 years.

4.2 Interpretation of oligonucleotide shorthand notations

4.2.1 Introduction

Organic structures can be complex in composition and may present a challenge to draw in full detail. Often such structures can be represented by empirical formulae (Garnett *et al*., 1985; Garnett and Hackling, 1995) or drawn in shorthand or abbreviated form. The

interpretation of such symbolic representations can be done meaningfully especially when prior learning of the expanded structural form is well-integrated into mental schemata (Thorndyke and Stasz, 1985). This enables the recognition of abridged features of an alternate representation, correlation with the expanded structural form and the offer of verbal or written communications on it (Lewalter, 2003; Crisp and Sweiry, 2006). An understanding (Shulman, 1986; Ametler and Pinto´, 2002) of both forms of structural depictions and the ability to transfer knowledge correctly (Salomon and Perkins, 1989; Taber, 2003; Cook *et al*., 2006) are essential in this exercise. In this study, students were tested in their ability to differentiate between the phosphodiester and phosphite groups found illustrated in shorthand notations of synthetic oligonucleotide intermediates. Apart from the visual cue (Gilbert, 2005; Takayama, 2005) that such unique oligonucleotide structures offer, a correlation with the expanded organic structures of such groups would be useful (Garnett and Hackling, 1995). Coupled with such an exercise is the understanding of the molecular basis that contributes to differences in bonding structure (Peterson and Treagust, 1989; Garnett and Hackling, 1995; Birk and Kurtz, 1999) and the reasons for differentiation (Harrison *et al*., 1999; Grayson, 2004). The understanding of the concept of valency (3.5.3; Table 3.1) and oxidation of the phosphite group to the phosphophodiester link is vital in this differentiation (Table 3.1). The identification of structural features in symbolic nucleotide representation becomes important in the analysis of synthetic oligonucleotides as shorthand notation is commonly used in different sources of literature (Chapters 2 and 3; Table 2.1)

4.2.2 Students' understanding of the concept of valency

The free-response probe (3.5.3) which delved into the students' understanding of the concept of "valency of an atom" produced a wide range of responses which could fit certain categories (Table 4.5). Superficial, incomplete responses constituted the large majority, where students referred to valency as "bonding ability of an atom" or "number of bonds an atom can form". Approximately 25 to 30 percent of students produced no

response. Examples of inappropriate responses, documented over a period of four years, included:

i. "It is the number given to an element or atom in a periodic table".
ii. "Valency refers to the outermost energy levels in which electrons are found".
iii. "Valency is the number of electrons an atom can bond".
iv. "It is the number of electrons an element is running short of, so that it can bind with other elements".
v. "It is the number of electrons present in the last orbital of an atom or molecule".
vi. "Valency is the number of free electrons available for bonding".
vii. "Valency refers to the charge on an atom".
viii. "Valency is the extra electrons in an atom".
ix. "Valency is the number of electrons in an atom".
x. "Valency is the number of electrons an atom can accept".
xi. "Valency is the number of electrons an atom can carry".

It is evident, from the above, that a large number of answers were erroneously constructed involving electrons of an atom. Although this was a basic enquiry, none of the students (n_{1-4}) produced any information on VSEPR theory or the Octet Rule (3.5.3) affecting the combination of atoms to yield a bond (Peterson and Treagust, 1989; Birk and Kurtz, 1999). From the above statements, it would be reasonable to assume that the concept of valency and the theory behind bond formation were poorly understood from Chemistry modules offered at levels 1 and 2 (3.1). A deficiency in their understanding of vital foundation knowledge and development of an alternate conceptual framework (Taber, 1998) point to the need for epistemic scaffolding (Martinez-Gracia *et al*., 2003; Sins *et al*., 2005; Oh, 2005), perhaps through remedial exercises and explanation-driven inquiry (Sandoval and Reiser, 2004). Further information to support this assumption is presented from student interviews (4.2.4), following an attempt to correlate the students' definition of valency and their ability to predict the valency of the phosphorus atom in oligonucleotide intermediates (4.2.3).

4.2.3 Students' difficulties with phosphodiester and phosphite groups in symbolic shorthand structures

The author attempts to correlate the students' definition of valency and their ability to predict the valency of the phosphorus atom in oligonucleotide intermediates (Figure 3.1). Four categories of students' response on the meaning of valency are presented in Table 4.5. Students, who referred to valency as the "bonding ability of an atom" or "number of bonds that an atom can form" (Table 4.5), correctly differentiated phosphite and phosphodiester groups in oligonucleotide intermediates A, B and C (Figure 3.1; 3.5.3). In addition, they also indicated that the valency of the phosphorus atom in the phosphite (A) and phosphodiester (B) intermediates as three and five, respectively. With respect to the shorthand notation of the phosphodiester group in intermediate C, all but two, one and three students of groups n_2, n_3 and n_4, respectively indicated that the valency of the phosphorus atom is two. While the identity of the phosphodiester group was not construed differently in intermediate C, certainly the visual impact of its shorthand format indicated to some students that there are two bonds to the phosphorus atom and therefore the valency is deceptively two. When factual information is learnt, it is imperative that learners be able to interpret and communicate this information meaningfully. Understanding shorthand notation such as an alternative structural representation is a form of visual literacy (Ferk *et al*., 2003; Schönborn and Anderson, 2006a). Structural shorthand notations can carry a wealth of information which is generally or implicitly no less than that shown by an expanded representation in most cases. Being visually literate about it, involves the meaningful interpretation of both its form and the information it bears (Ferk *et al*., 2003; Treagust *et al*., 2003). This is actually the decoding of symbolism. A mental model or comparative reflection on its expanded or alternative form is also useful in this type of analysis (Schönborn and Anderson, 2006a; b). Of the students who showed no response to the probe on the meaning of valency (Table 4.5), six, four, three and nine students respectively of groups $n_{1\text{-}4}$ could differentiate the phosphite group from phosphodiester group in intermediates A, B and C (Figure 3.1) without providing any meaning of valency nor any indication of the valency of the phosphorus atom in each functionality. Some of these students were interviewed in order

Table 4.5 Students' response on the meaning of valency

	Student number (n)			
	$n_1 = 54$	$n_2 = 35$	$n_3 = 31$	$n_4 = 44$
Categories				
Bonding ability of an atom	10	5	11	6
Number of bonds an atom can form	18	15	7	19
No response	14	11	9	14
Inappropriate Responses containing common factual information *				
Number of atom in periodic table	1	0	0	0
Charge on atom	0	0	1	1
Extra electrons in atom	0	0	1	0
Free electrons available for bonding	1	1	0	1
Electrons in last orbital of atom	1	0	0	0
Number of electrons an atom can bond	1	2	1	0
Outer most energy levels in which electrons are found	1	0	1	1
Number of deficient electrons in an atom	2	1	0	2
Number of electrons in an atom	1	0	0	0
Number of electrons an atom can accept	2	0	0	0
Number of electrons an atom can carry	2	0	0	0

*The exact statement from candidates differed each year. The description of inappropriate responses above presents the common information derived from various expressions.

to establish the basis for their conclusions (4.2.4). Twenty three of the twenty five students (n_{1-4}), who produced inappropriate responses on the meaning of valency (Table 4.5), failed to provide the valency of the phosphorus atom in the phosphite and phosphodiester functionality in intermediates A, B and C, respectively (Figure 3.1). Neither could they identify nor name the phosphite group in intermediate A; yet all students in this category could identify the phosphodiester group in intermediate B and shorthand notation of it in intermediate C. Two students of group n_1, who defined valency as "the number of electrons an atom can accept" indicated that the valency of the phosphorus atom is three in the phosphite group which was also identified correctly. As regards intermediates B and C, both students identified the phosphodiester group in each but indicated that the valency of the phosphorus atom is five and two, respectively in these intermediates. From the above, it is apparent that some students memorize particular structures, assign the correct name to them but cannot find the basis for differentiation using valency of an atom. The meaning of a shorthand notation depicting a 3,5 phosphodiester link between nucleotides is poorly understood. The lack of critical thinking (Kogut, 1996), poor integration of knowledge (Thorndyke and Stasz, 1985) and development of an alternate conceptual framework (Taber, 1998) by several students are evident in this work. Similarly, while Chemistry students may be adept at balancing equations or commenting on reaction products, many do not understand the basis of bond formation nor are they able to draw diagrammatic representations illustrating the molecular basis of chemical reactions (Yarroch, 1985; Garnett *et al*., 1985; Peterson and Treagust, 1989; Garnett and Hackling, 1995). Since the sub-microscopic or molecular levels of organisation cannot be seen (Bowen, 1998; Treagust *et al*., 2003) an understanding of factual information about these should ideally contribute to mental models and an expression of the visual in the form of drawings or diagrams (Stylianidou and Boohan, 1998; Perini, 2005; Schönborn and Anderson, 2006a; b).

4.2.4 Analyses of selected interviews

The following interviews provide further information on the nature of students' understanding of valency and differentiation of the phosphite and phosphodiester groups in the oligonucleotide shorthand notation (Figure 3.1). The most apparent difficulty associated with the interpretation of the shorthand notation of the oligonucleotide structure was surface-level understanding of the molecular attributes of such a representation.

4.2.4.1 Surface-level understanding

Student S15 gave the following response to questions concerning valency and the differentiation of structural components of the oligonucleotide shorthand notation.

I: I noticed that you defined valency as "the bonding ability of an atom".
S15: Yes.
I: What is it that contributes to this ability?
S15: I don't understand what you mean.
I: Can you relate the issue of valency with VSEPR theory and the Octet Rule?
S15: The information for different atoms is complex. I can't remember how electrons fit energy shells and subshells but basically unpaired electrons can pair another from a different atom to form a bond [pause]. I can't tell you more about it.
I: Okay. I see that you correctly identified the phosphite and phosphodiester groups in intermediates A, B and C [Interviewer refers S15 to illustrations shown in Figure 3.1]. But tell me why you indicate the valency of the phosphorus atom to be 3, 5 and 2 in intermediates A, B and C [pause]. Take a look at this. [Interviewer shows S15 his written response]
S15: Eissh! [Local expression meaning Gosh!]. I know C should be 5 like B but you see when I looked at the two bonds, I wrote 2. [S15 points to shorthand notation of 3,5 phosphodiester link].

Student S15 provides remnants of information towards VSEPR theory (3.5.3), making no comment on the Octet Rule. Although the student uses a count of the bonds around an atom to indicate "bonding ability" or valency, he nevertheless makes a mistake in assigning the phosphorus atom a valency of 2 because the shorthand notation of the phosphodiester group shows "two linkage arms" to carbon positions 3 and 5 between consecutive nucleotides. When challenged metacognitively (Georghiades, 2004), there is an admission that the shorthand notation presented a form that resembled two bonds at the phosphorus atom.

Student S16 gave the following response on the shorthand notation.

I: I see that you did not define valency. [Interviewer refers to S16's lack of response to the both free-response and focused probes (3.5.3) on valency]
S16: Yes. I cannot remember details of valency shells and energy levels affecting bond formation.
I: But you surely remember some information on electron sharing between atoms in a bond [pause] lets say covalent bond.
S16. Sure. But I still can't tell you the exact meaning of valency. Some atoms can share space in a shell where there are unpaired electrons and when they do this a bond forms between them, so valency could be an indication of how many electrons can be shared between atoms but I am not sure.
I: I see you could differentiate the phosphite group [Interviewer points to intermediate A as shown in Figure 3.1] from phosphodiester groups in intermediates B and C].
S16: Aah that's easy. I learnt the structures and names.
I: What did you learn about the structures?
S16: This form is phosphite [points to intermediate A], that form is the, the [stammer] phosphodiester form [points to intermediate B] and this [pause] a short form [points to intermediate C] showing the same structure in B [points to intermediates B and C].

Student S16 was unable to provide relevant information on the molecular basis of bond formation or valency of the phosphorus atom yet the student could identify structural

intermediates of nucleotides by form and name. The visual form of a chemical structure can be learnt superficially as shown in this example. A symbolic representation can be memorized yet not fully understood as theory relating to the sub-microscopic components of the symbolic form may be difficult to comprehend (Treagust *et al*., 2003). This is also evident in an interview with student S17 as indicated below.

I: I see that you defined valency as the "number of electrons an atom can accept".
S17: Ehh yes, yes.
I: Can you relate this to VSEPR theory or the Octet Rule affecting bond formation?
S17: No. I can't remember details of it. I know that if an atom's energy shell has space to accept electrons from another atom then it can form a bond with that atom.
I: I see that you indicate that the valency of the phosphorus atom in intermediates A, B and C are 3, 5, and 2 [pause], respectively. Also you named the intermediate A as phosphite, [pause] B and C as phosphodiester groups.
S17: Yes.
I: How did you determine the valency of the phosphorus atom in each?
S17: The number of bonds tells us it is 3, 5 and 2 [points to respective intermediate].
I: But B and C are the same groups, not so?
S17: Oh yah! It's my mistake. I see now.
I: What do you see, my friend?
S17: C is the short form but I made a mistake by counting the bonds.

The interview with student S17 indicates a poor understanding of the molecular or sub-microscopic basis for bond formation. The concept of valency is not explained. There is also vague reference to the acceptance of electrons into "an energy shell" which subsequently facilitates bonding. Student S17 is able to identify functional groups of the oligonucleotide intermediates and assign a name to each. The visual impact of the structural form of the shorthand notation which shows the C3 and C5 linkage of the phosphodiester bond deceives student S17 on the valency of the phosphorus atom.

Interviews conducted with a few more candidates, who gave inappropriate responses for valency (Table 4.5), showed similar trends (transcripts not shown). The molecular basis of bonding was generally poorly understood. Students appear to memorize representations of oligonucleotides without understanding the finer, molecular detail associated with these structures.

4.2.5 Discussion

This study reports on difficulties associated with the differentiation of the phosphite and phosphodiester groups in oligonucleotides on the basis of valency and the role of an established shorthand notation (White *et al*., 1959; Conn *et al*., 1987; Stryer, 1995; 3.5.3) used to show the 3,5 phosphodiester link between nucleotides. The nature of the student difficulties concerning valency (Table 4.5) remains classified as level 2 or "anticipated" (Grayson *et al*., 2001) on the basis of teaching experience. The difficulties concerning valency, as described using the 3,5 phosphodiester shorthand notation (4.2.2; 4.2.3), are also classified at level 2 meaning that further research is required to fully clarify the nature of each difficulty. The exact difficulty, within the framework of oligonucleotide chemistry, has not been reported in the literature. However, published literature indicates problems with the phosphate abbreviation, characterized by a lack of standardization in the use of abbreviated forms (Akers and Smith, 1987; Carusi, 1992). The phosphate group can be found represented by the letter "P", the abbreviation "ph" or "Ph" or "Pi" (Akers and Smith, 1987), encircled P (Garrett and Grisham, 1995; Weaver, 1999), encased in a square (Turner *et al*., 1997) or diamond-shaped frame (Mathews and van Holde, 1990; cited by Carusi, 1992) and highlighted by squiggly lines (Carusi, 1992). There is also poor differentiation of the phosphate group ($HOPO_3^{2-}$) from phosphoryl group ($HOPO_2^{1-}$) which are both indicated by an encircled "P" (Watson *et al*., 1987; Akers and Smith, 1987; Carusi, 1992). In the ionized form, pyrophosphate is written as $P_2O_7^{-4}$ or $HP_2O_7^{-3}$, yet the use of "PPi" will not specify the ionized form (Akers and Smith, 1987). In this study, there is clear evidence that the molecular basis of shorthand structural representations is poorly understood. Differentiation of the phosphodiester and

phosphite groups is done on the basis of rote learning of the structural or symbolic form or visualization without an understanding (Collis *et al*., 1998; Ametler and Pinto´, 2002) of the molecular information that accompanies the shorthand representation (Treagust *et al*., 2003; Ferk *et al*., 2003). A similar trend has been shown in Chemistry where students are able to balance chemical equations and give products of a reaction; however, they falter when asked to explain the molecular basis of the reactions concerning bond formation. Further, the students were unable to draw representations which show the molecular basis of reactions (Yarroch, 1985; Garnett *et al*., 1985; Peterson and Treagust, 1989; Garnett and Hackling, 1995). The development of mental models concerning the sub-microscopic element is important especially when learners are challenged to demonstrate their understanding of the unseen (Treagust *et al*., 2003). Symbolic representations are simply an abridged form of illustration; however, the meaning and depth of molecular information they depict about the sub-microscopic must be revealed clearly by learners (Ferk *et al*., 2003; Barak and Dori, 2005). Invariably, such information reveals why a symbolic form is actually structured in a particular manner.

CHAPTER 5

DIFFICULTIES WITH SYMBOLIC REPRESENTATIONS OF PLASMIDS, RESTRICTION MAPS AND GENE MARKERS

5.1 Introduction

Symbolic representations of plasmid structures are varied (2.5.4; 3.5.4). They indicate physical form *in vivo*, different structural forms when isolated or cleaved by enzymes and functional domains which play a role in the expression of genes they carry. Thus plasmids are complex entities which present learning challenges with regard to their structure and function. There is evidence that structure-function related expression of genes is also associated with plasmids (3.5.4). Plasmid domains tend to be demarcated discretely in plasmid diagrams or maps and are associated with specific functions (Table 3.2). A learner must understand the functions of such domains and their role in recombinant DNA technology. As plasmids are nucleic acids, their molecular structure must also be understood to enable an explanation of their manipulation as vectors or vehicles (2.5.7.2) in recombinant DNA technology. Symbolic representations of plasmid can take the form of shorthand notation for nomenclature (2.5.4); however, they are represented by linear or circular maps which must be interpreted correctly in terms of placement of functional domains and cleavage sites which tend to be manipulated for gene insertion strategies, DNA sequencing (Brown, 1986; Turner *et al*., 1997) or mapping (Weaver, 1999; Szeberényi, 2002; Walsh *et al*., 2007; Table 3.3). In this chapter, students' interpretation of symbolic plasmid structures, gene expression and mapping are described. Recent literature on the design of tutorials and laboratory-based exercises on these issues suggest that there is a need to test students on the interpretation of plasmid structures, gene expression and mapping (Szeberényi, 2002; Walsh *et al*., 2007).

5.2 Symbolic plasmid forms and gene expression

The free-response probe (3.5.4), concerning plasmid form and the ability to express genes of interest, generated a range of views from the students. These could be placed in four categories and various sub-categories (Table 5.1). Many students provided speculative comment on different plasmid forms, some comments being discretely different while others were somewhat convoluted. Speculative views were analysed in terms of the supporting information provided by the students. Special attention is also given to alternate and inappropriate viewpoints which illustrate constructivism (Driver and Bell, 1986; Von Glasersfeld, 1992). Characteristically, these responses showed a synthesis of information based on idiosyncratic mental models and the expression of data which was not only incorrect but scientifically unacceptable.

5.2.1 Alternate conceptual framework and constructivism

Approximately 55 percent of respondents incorrectly indicated that gene expression was not possible using a linear form of the plasmid pBCH 301 (Figure 3.2; 3.5.4). Some of the reasons given for this incorrect prediction were as follows.

i. "Linear plasmids cannot enter cells".
ii. "Linear plasmids cannot be transformed".
iii. "Linear plasmids present an abnormal conformation and cannot be read by enzymes for gene expression".
iv. "Linear plasmids have a disrupted flow of genetic information - for gene expression to take place, plasmids must be circular and not broken at any point".
v. "Linear plasmids cannot enter the nucleus of the cell therefore expression of genes from it is not possible".
vi. "The moment linear plasmids enter a cell, they can be destroyed by nucleases".

vii. "Linear plasmids take up too much space in a cell therefore they cannot be expressed or multiply inside".

The above statements are analysed critically as follows to determine which are true, untrue or speculative. The assertion that linear plasmids "cannot enter cells" or "cannot be transformed" is not true; however, they can present a challenge to transform (Brown, 1986; 3.5.4). Linear plasmids have been expressed in cells (2.5.4; 3.5.4) although it is likely that the covalently closed form which assumes a supercoiled conformation *in vivo* is most efficiently expressed or replicated (Weaver, 1999; 3.5.4). Not all plasmids are expressed via integration into the nucleus. Linear plasmids, encoding a killer phenotype, have been expressed cytosolically in certain yeasts (Ligon *et al*., 1989; McCracken *et al*., 1994; 3.5.4). The integration of linear plasmids or vectors such as YACs (yeast artificial chromosomes) into the nucleus and expression of genes cloned in such constructs are also possible (Gardiner *et al*., 1984; Pretorius, 2000). It is not true that linear plasmids do not express a phenotype nor replicate because they occupy "too much space in a cell". Certainly, smaller supercoiled plasmids might have a higher copy number and be suited, in terms of conformation, for enzyme interaction during transcription (3.5.4). The destruction of newly transformed linear plasmids in cells by nucleases is a speculative comment. The statement which reads "Linear plasmids have a disrupted flow of genetic information - for gene expression to take place, plasmids must be circular and not broken at any point" illustrates constructivism (Driver and Bell, 1986; Von Glasersfeld, 1992). The student presents an idiosyncratic mental model that genetic information is confined within the structural domains of circular plasmid DNA and not a linear plasmid. This student was interviewed (5.6) to gather further information on these statements and to establish her overall understanding of gene expression. In the interview (5.6.1), the student says "*Once you disrupt the plasmid or make it linear, the flow of genetic information does not carry well in the plasmid. The plasmid must be circular for the information to flow through it"*. Eleven students of groups $n_{1\text{-}4}$ (Table 5.1) indicated that the linear form of the plasmid would give good expression of the genes within this construct (Figure 3.2). Noting that the plasmid gene sequences are uninterrupted in this form, expression is a possibility (3.5.4) although a prediction of the level of expression

Table 5.1 Students' response on the link between plasmid form and gene expression

	Student number (n)			
	$n_1 = 54$	$n_2 = 35$	$n_3 = 31$	$n_4 = 44$
Categories*				
Linear form				
Good expression	2	3	2	4
Average expression	2	3	1	2
Expression not possible	30	19	18	24
Some expression	0	0	1	3
Circular and covalently closed				
Good expression	25	20	14	18
Expression not possible	0	0	3	1
Other forms				
Alternate views on expression	5	6	6	8
No response	0	1	0	2

*Several individual answers contained information fitting different categories.

would be speculative. The *Pvu* II-cohesive ends of the plasmid indicate that recircularization and supercoiling are likely to take place *in vivo* (Gardiner *et al*., 1984; Brown, 1986). This is the conformation associated with gene expression by most plasmids (3.5.4; Brown, 1986). Students speculated on various levels of expression (Table 5.1) which is acceptable; however, the reasons which accompany such viewpoints are untrue. Seven of these students indicated that expression is possible because "linear DNA can be transformed easily". The others attributed this level of efficiency in gene expression because:

i. "The sequences are intact and unhindered in a linear form of DNA".

ii. "The enzyme DNA dependent-RNA polymerase moves easily on a linear DNA template to facilitate transcription".

iii. "Linear DNA fits nicely in a cell".

iv. "Linear DNA transcribes better because its straight form is less complex than supercoiled DNA".

The above statements indicate a misunderstanding of the role of conformational forms of DNA and gene expression (3.5.4; Table 3.3). Linearity is perceived to be a form that facilitates transcription of a "straight", "unhindered" template as the RNA polymerase moves from one end of the molecule to the other. There were fifteen responses from groups $n_{1\text{-}4}$ indicating that there would be "some or average expression" of genes in the linear form of the plasmid. The only reason extracted from these responses was that the transformation of the linear structure would be difficult to accomplish hence a poor uptake of the plasmid would simply result in the adverse or poor expression of the genes. The students were unable to provide any discussion on what really contributes to adverse gene expression which is really a separate issue from poor transformation. Approximately 45 percent of students (Table 5.1) indicated that the closed covalent form of the plasmid would facilitate good expression of the cloned genes X and Z(α), encoding amylase and β-galactosidase, respectively. The most common reason, extracted from these responses,

was that this particular plasmid conformation favoured the transcription of genes. As plasmids are generally found supercoiled *in vivo*, this belief is untrue. Other interesting comments were also extracted from student responses. These included:

i. "The closed covalent plasmid is the most relaxed structure. Being relaxed, it can comfortably express the genes it has because enzymes can bind to it easily".

ii. "A covalently closed form of the plasmid expresses its genes the best because the genetic information is contained within the structure".

These statements indicate an alternative conceptual framework (Taber, 1998) and constructivism (Driver and Bell, 1986; Von Glasersfeld, 1992) concerning plasmid conformation and gene expression. The first statement illustrates an idiosyncratic mental model that a "relaxed" plasmid form is associated with "comfortable gene expression" and "ease at which enzymes bind". Being "relaxed", "comfortable" or showing an "ease" towards a function might be attributes of a personal feeling towards "relaxation" as an experience (Kasanda *et al*., 2005). The second statement associates the containment of genetic information within the confines of a circular structure. These students were interviewed (5.6) to gather further information on their understanding of plasmid conformation and gene expression. When asked about circular covalently closed plasmids and gene expression, student S19 remarks (5.6.2); "*The plasmids are relaxed and in this form they offer no problems with gene expression"*. In addition, the student says;*"You see the enzyme which transcribe a gene is not blocked by coils or a difficult twist in the DNA [pause]. The enzyme moves easily along the DNA because it is smooth"*.

Four students indicated that the covalently closed form of the plasmid would not be expressed in the cell unless this structural form would revert to a "supertwisted" or "supercoiled" state in the cell as topoisomerases or gyrases act on them. This view is speculative, nevertheless, acceptable in view of the influence of such enzymes on plasmids *in vivo* (3.5.4). Twenty five students of groups $n_{1\text{-}4}$ gave alternative but intelligent views on gene expression (Table 5.1). Their answers focused on the difficulty transforming a linear plasmid but indicated the possibility that a linearized plasmid such

as pBCH 301, cleaved at one site by *Pvu* II, would re-ligate or "repair" in the host to generate a circular form which will supercoil, given the action of topoisomerases or gyrases. A supercoiled form would then facilitate optimum gene expression as other native, supercoiled plasmids would do. This is an imaginative view based on the understanding of both plasmid form and processes affecting the expression of plasmid-borne genes.

5.3 Gene fusion in plasmid constructs

This aspect of research supplements the findings (4.1.5; Table 4.4) recorded earlier on the meaning of the symbol, "**: :**". In essence, students were tested further on the use of this symbol in a plasmid gene construct (3.5.4; Table 3.4). The plasmid construct generates an open reading frame (ORF), allowing a single transcript and translation product with dual enzyme activities, viz., amylase and β-galactosidase, encoded by genes X and Z (α), respectively (2.5.2; 3.5.4).

5.3.1 Erroneous information transfer and superficial understanding

Over seventy percent of students (n_1 = 40/54; n_2 = 30/35; n_3 = 24/31 and n_4 = 33/ 44) correctly identified the symbol "**: :**" as one which indicates fusion of genes X and Z(α). In addition, they predicted that the in tandem fusion of such genes would generate an ORF and that transcription would yield mRNA, followed by translation where the transcript is decoded to a single protein with dual function as described (2.5.2; 3.5.4). Of the total numbers, 6, 0, 2 and 0 students of groups n_{1-4} respectively gave no response to the focused probe (3.5.4). Five, three, one and six students of groups n_{1-4} respectively incorrectly indicated that gene Z(α) integrates into gene X in the construct, thereby inactivating it. They explained further that transcription of the construct gives mRNA which is translated to yield a single protein, expressing the β-galactosidase function only. This type of theory would be applicable if gene Z(α) was a transposon (2.5.2; 3.5.2);

however, there appeared to be disregard for the supplementary information, which accompanied the probe (3.5.4), defining this gene among other plasmid-borne sequences (3.5.4; Figure 3.2). As gene fusion and integration can be shown using the same symbol, this could be confusing to students although such symbols are defined when used in specific contexts. Erroneous information transfer (Grayson, 1995) can be expected when students reflect superficially on taught concepts and engage poorly (Chi *et al*., 1981; Darby, 2005) with questions that probe a deeper understanding of coursework (Salomon and Perkins, 1989, Sandoval and Reiser, 2004).

Three, two, four and five students respectively of groups $n_{1\text{-}4}$ gave inappropriate definitions of the symbol "**: :**". The responses were disturbingly identical to those reported earlier for the respective groups (Table 4.2). The same candidates interpreted the symbol "**: :**" as "a gene separator", "a chromosome component holding genes", "electrons between genes", "gene linkage" and "bonding of genes". Given their earlier responses reported on these definitions (4.1.5), shorter but incisive interviews were conducted to establish their understanding (5.6) of gene expression involving the plasmid construct pBCH 301 (3.5.4; Figure 3.2).

5.4 Linearization and re-circularization of plasmids

5.4.1 Erroneous substitution

All students understood that when circular plasmids are cleaved at one point by a restriction endonuclease, such plasmids linearize. Re-circularization is possible unless plasmids are treated with alkaline phosphatase which removes the phosphate group either at the 3′ or 5′ end of the cleavage point (3.5.4). When the students were asked how they would prevent re-circularization of the *Pvu* II-cleaved plasmid pBCH 301 (3.5.4), all but three students of group n_4, correctly indicated that alkaline phosphatase would be used to treat the plasmid to enable cleavage of a phosphate group as described. In contrast, the three students from group n_4 wrongly indicated that the plasmid would be

treated with alkaline phosphorylase. This information was clearly learnt incorrectly as students erroneously substituted phosphatase with phosphorylase. The latter enzyme (phosphorylase) catalyses the cleavage of its substrate(s) via nucleophilic attack by inorganic phosphate (Scism, 1996).

5.5 Plasmid functional domains

5.5.1 Rote learning and erroneous substitution

Students were tested on the role of ARS, ori, operator and promoter sequences found in plasmid, pBCH 301 (3.5.4; Figure 3.2). In general, most students understood the role of the selected sequences in plasmid pBCH 301 (Table 5.2). Two types of difficulties emerged among the different student groups. Four and three students of groups n_1 and n_2 had difficulty providing the role of ARS or even writing the full name of the sequence ("Autonomous Replicating Sequence") which suggests the function (3.5.4). The second difficulty, affecting four, three, one and two students respectively of groups $n_{1\text{-}4}$, concerned the substitution of definitions for operator and promoter. It does seem that some students engage in rote learning (Grayson, 1995) as the definitions provided for each are distinctly different (3.5.4). A definition such as "sequence which controls the rate of mRNA formation" was erroneously assigned to the promoter. Likewise, these students referred to the operator as "a sequence which facilitates the binding of DNA dependent-RNA polymerase for transcription to occur", demonstrating a failure to "disentangle" (Grayson, 2004) such definitions which are discretely different. Some researchers reflect more deeply on similar trends as a tendency to superimpose one concept upon another (Grayson, 2004; Schönborn and Anderson, 2006b). Perhaps this example might illustrate that the students had simply not rote learnt the definitions properly. The author does not encourage rote learning in his teaching practice.

Table 5.2 Students' response on the role of selected plasmid-borne sequences

	Student number (n)			
	$n_1 = 54$	$n_2 = 35$	$n_3 = 31$	$n_4 = 44$
Categories				
Role understood				
ARS	50	32	31	44
ori	54	35	31	44
operator	50	32	30	42
promoter	50	32	30	42
Problems				
Acronym: ARS	4	3	0	0
Substitution Errors				
operator-promoter	4	3	1	2

5.6 Difficulties associated with the integrated nature of topics

Interviews were conducted to consolidate the findings of the current investigations. Specifically, the students' understanding of plasmid conformation, gene expression and a plasmid construct were probed in an integrated manner.

5.6.1 Misunderstanding, substitution and poor knowledge integration

The following interview excerpt illustrates a student's misunderstanding of plasmid structure, conformation and gene expression. The analysis of the response reveals erroneous substitution of facts and poor knowledge of gene expression by a plasmid.

I: Do you want to tell me why you feel that the linearized form of the plasmid pBCH301 cannot express these genes? [Interviewer presents maps of pBCH301as shown in Figure 3.2].
S18: Well, the plasmid must be intact [pause] I mean circular because it contains the genetic information.
I: What information is this?[Interviewer refers to supplementary information given in probe (3.5.4)]
S18: Gene X codes for amylase and gene Z codes for beta-galactosidase. Once you disrupt the plasmid or make it linear, the flow of genetic information does not carry well in the plasmid. The plasmid must be circular for the information to flow through it.
I: Tell me more about genes X and Z. When you say they code for amylase and beta-galactosidase [pause] what do you mean?
S18: Gene X contains amylase and the other gene contains beta-galactosidase.
I: So, what are the roles of the operator and promoter sequences?
S18: The operator facilitates transcription and the promoter controls the rate of mRNA formation [hastily].
I: Are these sequences [points to operator and promoter] linked to the expression of genes X and Z.

S18: No, they control transcription of all the DNA in the plasmid.
I: Again, what do you mean by expression of genes X and Z?
S18: These gene , you see[pause] are bonded by electrons [points to "::"] but each one contains an enzyme.
I: Does mRNA have anything to do with gene expression?
S18: Let me think now [pause] I came across it in protein synthesis but not plasmids. No. I don't think so.
I: Tell me [pause] the two enzymes amylase and β-galactosidase, [pause], how are they formed? Is mRNA implicated in their synthesis?
S18: [Silence]. They are protein but found in those two genes.

DNA is often described as a "molecule of heredity" or molecules associated with "the storage, transmission and expression of genetic information" (Stryer, 1995). It is also associated with the "flow of genetic information". It is imperative that these terms be understood in a scientific context. Student S18 presents a misunderstanding that the flow of genetic information is contained in a circular plasmid and that this is lost when such a plasmid is linearized. Certainly, the cleavage of a gene within the plasmid could result in the inactivation of the gene or lack of expression (3.5.4); however, the student is unable to associate "genes" with the mechanism of phenotypic expression. Instead, the genes, Z(α) and X, are believed to contain the enzymes β-galactosidase and amylase, respectively. The student is unable to reflect meaningfully on the mechanism of transcription of plasmid-borne genes and erroneously substitutes (Grayson, 2004) the specific transcription-related functions of the operator and promoter sequences. Further, the student incorrectly indicates that the operator and promoter sequences control the transcription of "all the DNA in the plasmid". Clearly the concept of gene expression is not understood by student S18. The revelation that mRNA is encountered in protein synthesis and not in the expression of plasmid-borne genes is untrue and illustrates inappropriate knowledge transfer (Salomon and Perkins, 1989). The student is unable to integrate knowledge (Thorndyke and Stasz, 1985) or reflect on appropriate prior knowledge or underlying schema (Kirschner, 2002; Ploetzner and Lowe, 2004) which adequately explains gene expression. While external representations of various kinds can

assist students in making links with prior knowledge (Kirschner, 2002; Schönborn and Anderson, 2006b), the symbol "**::**" was construed as "electrons" between bonded genes perhaps in the manner Lewis dot diagrams show covalent bonding between atoms (Peterson and Treagust, 1989; Peterson *et al*., 1989). The concept of an "open reading frame" (ORF) and single transcript expressing a protein with dual enzyme function was not revealed (3.5.2).

5.6.2 Alternate mental schemata

Molecular processes can evoke a form of mental imagery which may assist learners in explaining the underlying conceptual knowledge (Kozma, 2003) about the process itself. Conversely, when conceptual knowledge is not clearly understood, the form of mental imagery that manifests can be rather contentious to explain. Explanations, offered by student S19, were at variance with propositional knowledge on plasmid form and gene expression (3.5.4; Tables 3.3 and 3.4) as can be seen in the following interview excerpt.

I: Are circular covalently closed plasmids the best for gene expression?
S19: I think so.
I: Why?
S19: The plasmids are relaxed and in this form they offer no problems with gene expression.
I: What do you mean by problems, my friend?
S19: You see the enzyme which transcribe a gene is not blocked by coils or a difficult twist in the DNA [pause]. The enzyme moves easily along the DNA because it is smooth.
I: Okay. Take a look at this plasmid construct. In linear form, could it be transcribed? [pause] What do you think?
S19: I don't think that you can get it into a cell like the relaxed form, so transcription is unlikely.
I: Tell me more about these sequences marked "o" and "p".

S19: O is the operator which controls the rate of mRNA formation and p is the promoter where RNA polymerase [pause]eh, DNA- dependent RNA polymerase binds to synthesize mRNA.
I: Tell me more about this symbol [Interviewer points to "::" in construct].
S19: That is a joiner. It holds the genes together in the plasmid like a protein in a chromosome. It must be because DNA is like a nucleoprotein complex, so here in this plasmid, the same kind of protein is found.
I: Tell me more about transcription and these two genes [Interviewer points to Z(α) and X] .
S19: I don't understand.
I: Okay. You did tell me about sequence "o" and "p", now tell me more about the transcription of Z and X.
S19: Ohh! I see. Yes. Yes. Gene Z is joined to X after integration. It is held by the joiner but gene X is inactivated by Z , so mRNA formed from these two genes will only produce the protein from Z?
I: What protein is that?
S19: Beta-galactosidase.
I: Tell me, is gene Z a transposon [pause]? You know the type that integrates from one position to another, resulting in inactivation.
S19: Yes. Yes. It must be because transposons are joined like this after integration.
I: Thank you, my friend, for your input.

Student S19 imagines that the transcription enzyme is unhindered in its movement along a "circular covalently closed" DNA template because it is "not blocked by coils or a difficult twist". External representations, such as diagrams and symbols, may assist the learner in understanding conceptual information about a process or phenomenon (Cho *et al*., 1985; Kindfield, 1994) and hence conjure an improved visual perspective about it (Mathewson, 2005; Schönborn and Anderson, 2006a). However, some external representations tend to present information which is too complex to understand, resulting in "cognitive overload", suppression of the learning process (diSessa, 1993; Kirschner, 2002) or compromise in expression on the subject (Crisp and Sweiry, 2006; Yore and

Treagust, 2006). Other external representations may be poorly designed, leading to the construction of alternative (Taber, 1998) or faulty schemata (Schnotz and Bannert, 2003) thus obscuring a link to the underlying concept (s) (Kozma, 2003; Treagust *et al*., 2003; Perini, 2005).

The symbol, "**: :**" is incorrectly perceived to be a proteinaceous joiner of two genes as DNA generally exists as a nucleoprotein complex. Further, the symbol is interpreted incorrectly to denote integration of the Z(α) gene into gene X, thereby inactivating it as a transposon would do (3.5.2). The student affirms that the transcript will produce the protein encoded by gene Z(α). The use of a double colon to denote in tandem gene fusion or integration as shown by transposition is clearly emphasized during teaching. It remains a difficult symbol to interpret when the role and characteristics of gene sequences are not understood.

5.7 Restriction mapping

5.7.1 Combined learning difficulties: Substitution errors, erroneous information transfer and poor reasoning

The free response probe (3.5.5) indicated that students had understood the meaning of the term "restriction mapping" (Table 3.3). However, some definitions for restriction mapping were awkwardly expressed as illustrated by the following examples:

i. "Restriction mapping is a technique to make a map or diagram of a plasmid which shows all the cutting or restriction sites".

ii. "Restriction mapping is a method which is used to transfer restricted band patterns from a gel to paper. Depending on the migration pattern, the size of plasmid fragments can be put in correct order to show where cutting sites lie on the plasmid diagram".

iii. "Restriction mapping is a way to form a plasmid map. The map can be linear or circular to show where a plasmid can be cut by restriction enzymes and what size of fragments can be generated".

While the concept of restriction mapping was understood, this study reports further on difficulties associated with practical restriction mapping (Szeberényi, 2002; Walsh *et al*., 2007). In this regard, the three-part focused probe (3.5.5) presented a greater challenge to the students. Part one of the probe required a prediction of the electrophoretic separation of the fragments of a plasmid subject to six separate digests. A hypothetical arrangement of fragments as shown in Figure 3.3 would suffice, accompanied by labels of the negatively-charged cathodic and positively-charged anodic ends, respectively, a relative scale alongside the length of the electrophoretogram (drawing of electrophoresed gel) to indicate fragment size and relative migration. The following trends were recorded from students' written responses.

i. Correctly shown arbitrary kilobase (kb) scale alongside an electrophoretogram, coupled with the migration of fragments from a negatively-charged cathodic end to a positively-charged anodic end. Larger fragments were shown to be retarded in their migration relative to smaller ones as shown in Figure 3.3.
Incidence: n_1 = 50 /54; n_2 = 32 /35; n_3 = 31 /31; n_4 = 40 /44

ii. Electrophoretogram and fragment electrophoretic separation as described above (i) except that migration was incorrectly shown from (a) a negatively-charged anode to a positively-charged cathode and (b) a positively-charged cathode to a negatively-charged anode.
Incidence involving the error shown in (a): n_1 = 3 /54; n_2 = 3/35; n_3 = 0 /31; n_4 = 3 /44
Incidence involving the error shown in (b): n_1 = 0 /54; n_2 = 0/35; n_3 = 0 /31; n_4 = 1 /44

iii. Presentation of an electrophoretogram without a kilobase scale but correct indication of the retardation of larger DNA fragments relative to smaller ones. The kb value was inserted above each band to show size. In addition, the migration of bands was correctly shown to be between a negatively-charged cathode to a positively-charged anode. However, a set of annotated notes, relating to thin layer chromatography (Plummer, 1987), accompanied the diagram. This included a label to the gel as "stationary agarose gel" and reference to "mobile buffer solvent carrying the DNA fragments towards the anode under the influence of electric current". Further, the student makes reference to the calculation of the R_f value of each DNA fragment using the "distance of fragment migration from the gel slot" and "the length of the gel as solvent migrates throughout its matrix".

Incidence involving the erroneous information transfer: $n_1 = 1/54$; $n_2 = 0/35$; $n_3 = 0/31$; $n_4 = 0/44$.

From the above, it is evident that most students understood the basis for DNA fragment separation using agarose gel-electrophoresis. This is an important pre-requisite to restriction mapping as banding trends provide information for the determination of restriction sites. Substitution errors and erroneous information transfer (Salomon and Perkins, 1989; Grayson, 2004) were also prevalent. There appears to be confusion regarding the polarity and nomenclature of electrodes as some students ($n_1 = 3/54$; $n_2 = 3/35$; $n_3 = 0/31$; $n_4 = 3/44$) indicated that the migration of DNA is from a negatively-charged anode to positively-charged cathode. One student of group n_4 indicated that DNA migrates from a positively-charged cathode to a negatively-charged anode. There appears to be disregard of the fact that electrophoresis is performed using specific buffers and pH values, usually between 7.6 and 8.0, which facilitate the dissociation of the sugar-phosphate backbone in DNA , giving it an overall negative charge. This promotes its electrophoresis towards a positively-charged anode (Plummer, 1987; Kornberg and Baker, 1992; Garnett and Treagust, 1992b). The amalgamation of theory which applies to thin layer chromatography with that applying to agarose gel-electrophoresis (Plummer,

1987) is another example illustrating a student's failure to disentangle concepts (Grayson, 2004; Schönborn and Anderson, 2006b). Inappropriate information transfer (Schönborn *et al*., 2002a) can manifest in constructivism and an alternate conceptual framework (Driver and Bell, 1986; Taber, 1998) as shown by the example above.

The second part of the focused probe (3.5.5) required interpretation on two bands of the same undigested plasmid, corresponding to size estimates of *ca*. 4.0 and 4.1 kb, respectively. Generally, the nicked form of a circular plasmid which linearizes, tends to migrate closely behind the supercoiled species while the relaxed, covalently closed circular (cccDNA) conformer of the plasmid is retarded in its migration (3.5.4). The students' views on this matter were placed in the different categories. These included:

i. Most students predicted that the cccDNA would be retarded in its migration and those bands as close as 4.1and 4.0 kb would correspond to the rapidly-migrating, nicked linear form of the plasmid and supercoiled conformer, respectively. Conformational size and physical constraints were the correct and main reasons given, contributing to the migration trends observed through the gel matrix.
   Incidence: n_1 = 54 /54; n_2 = 35 /35; n_3 = 30 /31; n_4 = 44 /44

ii. One student speculated that the 4.0kb supercoiled DNA would develop a nick. Subsequent closure of a nick by a topoisomerase would unfold the plasmid to give rise to the circular, relaxed form. This relaxed form would migrate slower than the supercoiled species at a size equivalent to 4.1 kb.
   Incidence: n_1 = 0 /54; n_2 = 0 /35; n_3 = 1/31; n_4 = 0 /44

The migration patterns of different plasmid conformers were generally well-understood by students expressing answers similar to those given in category "i" above. Given that the word, "undigested" was used in the probe to describe the plasmid (3.5.5), there exits a possibility that any reference to linear DNA can be omitted as shown in category "ii". The structure of concept maps or answers to a probe can be influenced when "word

association" acts a presumptive cue or prompt (White and Gunstone, 1992; Szeberenyi, 2003). Category "ii" describes one student's view which goes "beyond the information presented" (1.2; Bruner, 1973, cited by Cooper, 2005), reflecting changes to DNA topology, possibly induced by topoisomerases (3.5.4) to yield cccDNA from the supercoiled form (Table 3.3). It would be reasonable to speculate that a nick could be induced physically when loading a plasmid sample into the gel using a hypodermic syringe or micro-pipette. A nicked circular plasmid (ncDNA) (3.5.4) migrates closely behind the supercoiled form of the plasmid.

Part 3 of the focused probe (3.5.5) revealed difficulties with the practical aspects of restriction mapping (Figure 3.3). Although most students could interpret restriction digests and the electrophoretic profiles of different fragments, an average of 33% of students from groups n_{1-4} were unable to construct a plasmid map. The following incidence of the difficulty is reported, namely $n_1 = 20/54$; $n_2 = 9/35$; $n_3 = 11/31$ and $n_4 = 15/44$. The difficulty was characterised by the following trends.

i. A tendency to produce a circular drawing of the plasmid, traversed by a short line at one point, indicating "*Hin*d III" and "0/4.1" (Figure 5.1). This is a way to show the single restriction point (3.5.5) of *Hin*d III and the 4.1 kb fragment it generates on linearizing the circular plasmid. Students correctly provide this illustration but fail to explain that the *Hin*d III digestion fragment gives an indication of the plasmid size of 4.1kb. Circularity is interpreted taking into account the overlap of restriction patterns produced by the digest. This information is given by the probe (3.5.5).

ii. Following the placement of "*Hin*d III and 0/4.1", students attempted to fit the other fragment sizes as concentric rings around the initial circle drawing. This can give a rather convoluted pattern (Figure 5.1), characterised by incorrect placement of restriction sites. There was also a tendency to label information describing the digest and not engage in

reasoning (Figure 3.3 - aspect labelled "rough work") as to where restriction sites should fit in order to generate the different fragment sizes (Szeberényi, 2002). Some students use coloured pencils to demarcate fragment areas around a circular form. This may be beneficial in visualizing fragment placement, including overlap; however, there will always be a need to engage in reasoning (Kindfield, 1994; Weaver, 1999; Zohar and Nemet, 2002) and problem-solving (Taconis *et al*., 2001; Tsui and Treagust, 2003) when mapping DNA fragments (Szeberényi, 2002; Walsh *et al*., 2007).

iii. Two students of group n_3 mapped the restriction sites in an anti-clockwise manner. This is rarely encountered; however, students are made aware of this alternate map or mirror image of the clockwise indication of restriction sites through verbal communications.

It is well-known that colour, labels or shape may constitute the most salient visual features of a diagram, thereby enhancing visual cognition and the students' ability to interpret the representation (Gilbert, 2005; Cook *et al*., 2006; Crisp and Sweiry, 2006).Researchers (Lord, 1985; Wu and Shah, 2004; Jones *et al*., 2006; Cook *et al*., 2006) have reported that such features tend to enhance both lower and higher order cognitive skills as outlined by Anderson and Krathwohl (2001). Textual information, the underlying molecular theory on mapping and verbal information on this concept contributes to the verbal mental representation of the learner. Collectively, both representations contribute to dual processing of a mental model, a characteristic coding feature of external information in accordance with Paivio's dual coding theory (Paivio, 1986, cited by Schönborn and Anderson, 2006a).

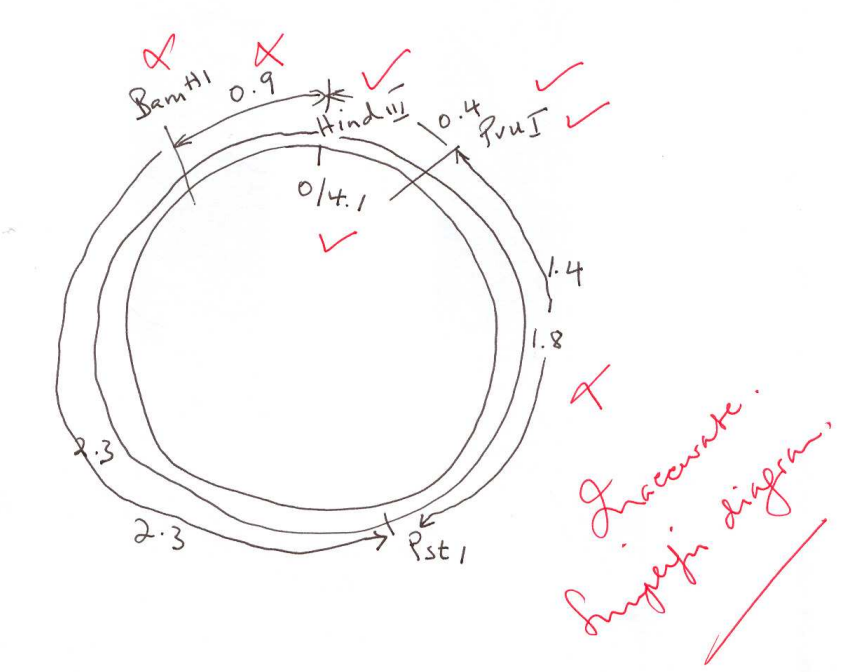


Figure 5.1 Illustration of a typical convoluted pattern generated from a restriction mapping exercise.

5.7.2 Procedural and reasoning difficulties

The following interview transcript shows evidence of procedural and reasoning difficulties which impeded the student's task on restriction mapping.

I: What is restriction mapping?
S20: It's a technique where DNA [pause] say a plasmid is cut by restriction endonucleases[pause] to give fragments. You electrophorese the fragments and this gives you the separation based on size. From the gel, you tell where each fragment fits in the plasmid and there [then] you can mark the site of cleavage.
I: So how do you determine fragment size?
S20: DNA markers can be used to estimate size depending on where fragments migrate.
I: Tell me more about this migration.
S20: You see large fragments are held back and small ones migrate faster through the gel pores.
I: What influences the migration?
S20. Well, electric current moves between the cathode to the anode and carries the DNA.
I: How is the DNA carried?
S20. It moves, you see, to the anode which positively charged because the DNA is negatively charged.
I: Are you able to see the DNA?
S20: Yes. You can use a dye like ethidium bromide. It binds to the DNA and when you shine UV light on the gel, you see the DNA bands shining bright.
I: Okay, but I see you are having trouble with mapping, not so?
S20: Ahhg! This thing is like a jigsaw puzzle.
I: Why do you say that?
S20: It takes time to match fragments around a circle. Like finding matching pieces of a puzzle. It can be tricky but not really difficult.
I: Okay. Thank you, my friend.

Student S20 shows an understanding of restriction mapping but expresses theory on electrophoresis in an awkward manner. There is a clear indication that practical aspects of restriction mapping can be "tricky but not really difficult". Student S20 draws similarity between restriction mapping and matching pieces of a jigsaw puzzle. Further, the cognitive challenge imposed by the technique could well necessitate the allocation of additional "time to match fragments around a circle". This interview demonstrates that while "conceptual knowledge" on restriction mapping could be fairly sound, the closely linked "procedural knowledge" associated with the technique could be cognitively more demanding (Grayson, 1995; Kirschner, 2002).

The following interview also shows that a demonstration of procedural knowledge and reasoning ability can be cognitively demanding. Student S21 understands the theoretical basis of restriction mapping but falters in the practical aspect of mapping.

I: I noticed that you explained restriction mapping fairly well [pause], also your understanding of restriction endonucleases and electrophoresis is good. Tell me, why the difficulty with mapping? [Interviewer browses through the written effort of student S21].
S21: I needed more time to complete the mapping.
I: Is it difficult to achieve?
S21: No. There is nothing factual about it. I mean I know the theory but this mapping issue, phew! [pause]It can take you in circles!
I: What do you mean, my friend? Your reference to circles is of interest to me.
S21: Well, plasmids are generally mapped in circles but you have to work out where a fragment fits. Then when you fit one fragment, you find there is an overlap with another fragment. It is very confusing and tricky.
I: Okay. Thank you for your input.

The revelation that mapping "can take you in circles", indicates confusion as attempts are made to fit fragments in a circular plasmid configuration. Students appear to draw several concentric rings in an attempt to find the correct placement of restriction sites (Figure 5.1). As shown in Figure 3.3 (Chapter 3), restriction cuts may be placed in separate

drawings for the sake of clarity, followed by reasoning with regard to their placement in a composite form. Weaver (1999) shows separate diagrams of linear fragment cuts, an assessment of overlap and reasoning ((Szeberényi, 2002; Pata and Sarapuu, 2006; Walsh *et al*., 2007) as to how these should fit in circular form.

5.8 Gene markers and phenotypic expression

5.8.1 Inappropriate information transfer and superficial understanding

The free response probe (3.5.6.1), investigating the definition of a "gene marker", elicited a range of responses which could fit the following categories;

i. Association of DNA and an encoded property
ii. Association of DNA with the containment of a property, entity or characteristic
iii. Association of DNA fragments as an indicator of size when performing electrophoresis
iv. DNA elements which can be grown when cells are cultured.

These categories of responses are assessed for their correct or incorrect content as follows.

Almost ninety percent of students (n_1 = 48/54; n_2 = 30/35; n_3 = 30/31; n_4 = 40/44), defined "gene marker" correctly, providing a link between "a gene" and "an encoded property" or phenotype. Examples of responses fitting the first category included;

i. "A gene marker is a piece of DNA, usually about 1000bp, which encodes a property that defines a cell type. For example, a cell may produce an enzyme such as thymidine kinase, therefore the marker is Tk^+. Other cells may not produce this enzyme, therefore such cells are indicated by the marker, Tk^-".

ii. "A gene marker" tells us that a gene is responsible for the coding of a property, say resistance to the antibiotic ampicillin. This marker is indicated as *Amp* ^{r}".

iii. "A gene marker like *LEU* shows a dominant leucine biosynthesis gene. A recessive form is *leu* in small letters. When cells like auxotrophs cannot synthesize leucine, we show the marker as *leu* $^{-}$".

iv. "If gene X encodes an enzyme x, we can say this gene marker can be shown as X^{+} because it expresses x (the enzyme). X^{-} means that perhaps a mutation of the gene prevents expression of the enzyme, so the marker is X^{-}".

In the second category of difficulty, students (n_1 = 2 /54; n_2 = 0/35; n_3 = 1/31; n_4 = 2/44), incorrectly referred to "gene marker" as DNA which contains an enzyme, an amino acid or simply a property such as antibiotic resistance or sensitivity. Examples of such incorrect expressions are given below.

i. "A gene marker such as *lac(z)*$^{+}$ indicates DNA which contains the enzyme, β-galactosidase inside the cell".

ii. "*LEU* would be a marker gene since it contains the amino acid leucine".

iii. "A marker gene is a gene containing a rare property such as resistance to the antibiotic, ampicillin. A gene marker such as *Amp*^{r} is found in bacteria".

iv. "If a cell is sensitive to an antibiotic like ampicillin, this sensitivity is found in the marker genes of the cell, like the *ampicillin*-sensitive (*Amp*^{S}) gene".

v. "A gene marker contains a property such as antibiotic resistance which characterizes a cell. Therefore, the cell has a marker like *Tet* ^{r} which is tetracycline resistance".

Category (iii) describes the association of "gene marker" with DNA marker fragments which are used to estimate the size of DNA during electrophoresis (Figure 3.3). This is an example of inappropriate transfer of information (Salomon and Perkins, 1989; Schönborn

et al., 2002a; Grayson, 2004) perhaps influenced by "word association", in this case "marker" being the presumptive cue or prompt (White and Gunstone, 1992). The incidence of the difficulty recorded among groups n_{1-4} was $n_1 = 4/54$; $n_2 = 3/35$; $n_3 = 0/31$ and $n_4 = 1/44$, respectively.

The incorrect perception that gene markers are in fact DNA elements which could be grown in cells when cultured, is presented in category (iv). The incidence of this difficulty among groups n_{1-4} was $n_1 = 0/54$; $n_2 = 2/35$; $n_3 = 0/31$ and $n_4 = 1/44$, respectively. Students were interviewed in order to gather further information on this difficulty and to assist with its interpretation (5.8.2). Student S22 says in her interview; "*But the marker is in the cell, so it also grows in the medium"* (5.8.2).

The use of focused probes 1 and 2 (3.5.6.1) provided similar data on the students' understanding of gene markers and phenotypic expression. Forty five of a total of fifty four students of group n_1 gave various examples of gene markers and their respective functions or phenotypes as shown in Tables 3.2 and 3.4. These included *LEU* for leucine biosynthesis, Tet^r and Amp^r for tetracycline and ampicillin resistance, respectively, Tet^S and Amp^S for tetracycline and ampicillin sensitivity, respectively, $lac(z)^+$ and $lac(z)^-$ for production or non production of β-galactosidase, respectively, leu^+ and leu^- to indicate leucine prototrophy and auxotrophy, respectively, Tk^+ and $Hgprt^+$ for production of thymidine kinase and hypoxanthine guanine phosphoribosyl transferase, respectively and Tk^- and $Hgprt^-$ for non production of the enzymes thymidine kinase and hypoxanthine guanine phosphoribosyl transferase, respectively. Three students of group n_1 gave $Hgprt^+$ but provided no further comment or indication of phenotype. Students of groups n_{2-4} also provided similar examples of gene markers and their respective phenotypes as shown for group n_1 above. The ratio of students, showing this level of proficiency, correlated with the data shown earlier on those who had understood the concept of a "gene marker", viz., 30/35, 30/31 and 40/44 respectively for groups n_2 , n_3 and n_4. Identical response patterns ($n_1 = 2/54$; $n_2 = 0/35$; $n_3 = 1/31$; $n_4 = 2/44$), were also recorded regarding the wrong perception that "gene markers are in fact genes which contain a property or entity". The "property or entity" would be the "phenotype". These included gene markers (Table 3.2)

such as *LEU, lac(z)*$^{+}$, *Amp*^{r}, *Amp*^{S} and *Tet*^{r} which "contain" leucine, β-galactosidase, ampicillin resistance, ampicillin sensitivity and tetracycline resistance, respectively.

The inappropriate reference to "gene markers" as determinants of molecular weight or size of DNA was also recorded. No names of gene markers were recorded. The incidence of this difficulty recurred among groups $n_{1\text{-}4}$ as shown with the previous free-response probe, viz., $n_1 = 4/54$; $n_2 = 3/35$; $n_3 = 0/31$ and $n_4 = 1/44$, respectively. In response to the third focused probe (3.5.6.1), only this exclusive group of students indicated that gene markers could be visualized on a gel using ultraviolet light. Two and one student of groups n_2 and n_4 respectively listed both *Hgprt*$^{+}$ and *Tk*$^{+}$ as markers which could be grown in cell cultures. Neither definition of phenotype nor nomenclature of these markers was provided.

5.8.2 Superficial learning and poor differentiation

The following interview with student S22 revealed that an abbreviation or acronym could be memorised yet the more important and informative nomenclature about it is not learnt.

I: Are you really able to grow gene markers?
S22: Yes.
I: But [pause] what is a gene marker?
S22: It is DNA in a cell. It may contain an enzyme such as Hgprt, so the marker is Hgprt^{+}.
I: What is Hgprt?
S22: I can't remember the full name of the enzyme.
I: Where did first come across Hgprt?
S22: In the monoclonal antibody lectures. You know the th' [pause] hybridoma cells.
I: Surely, you grow cells and not the marker.
S22: But the marker is in the cell, so it also grows in the medium.

I: Thank you for your input. I want you to see me some time, we need to talk about gene markers and the hybridoma. Okay?
S22: Yes, yes. I will come. Ha ha, I know you always full of trick questions.

Hypoxanthine guanine phosphoribosyl transferase (Hgprt) is the name of a purine salvage enzyme which converts hypoxanthine to IMP or guanine to GMP, utilizing phosphoribosylpyrophosphate in both conversions (3.5.6.2; Köhler and Milstein, 1975; Staines, 1983). The unabbreviated name of the enzyme actually gives the substrates involved in the pathway. The superficial learning (Chi *et al*., 1981; Schönborn *et al*., 2002a) of acronyms is not beneficial to the student. Further, student S22 showed no understanding of the concept of a gene marker or phenotypic expression, indicating that a gene marker contains an enzyme and that it could be grown as an endogenous component of the cell during cell culture.

In the following interview transcript, student S23 demonstrates the inability to differentiate or "disentangle" the concepts (Grayson, 2004), "gene marker" and "molecular weight marker" until challenged to do so (Posner and Gertzog, 1982; Hewson and Hewson, 1984; Duit and Treagust, 2003).

I: You refer to gene markers as DNA fragments that can be seen on a gel [pause] using UV light and that eh[pause] they could be used to determine the molecular weight of DNA.
S23: Yes.
I: Are you not confusing molecular weight markers with gene markers?
S23:[Silent at least 10 seconds] Well, a gene marker of known molecular weight can be used during agarose gel electrophoresis to estimate the size of other DNA.
I: Sure, but I want to find the definition of a gene marker within the framework of phenotypic expression. Can you give me such an example? And differentiate it from DNA molecular weight marker.
S23: My mind is blank now. I was thinking about markers in a gel.
I: Think about marker genes, my friend. Remember the lectures on plasmids?

S23: You mean something like Tet^r or the tetracycline resistance gene.
I: Exactly. They are different from molecular weight markers.
S23: I see what you mean. I was confused.

The student's assertion that "a gene marker of known molecular weight can be used during agarose gel electrophoresis to estimate the size of other DNA" is true (3.5.6.1). However, DNA molecular weight markers need not be genes whose molecular weights are known. Often these are DNA fragments of known molecular weight generated from the digestion of phage λ DNA which is commercially available (Boehringer-Mannheim, FRG). Towards the near end of the above interview, a prompt linking gene markers and lectures on plasmids (White and Gunstone, 1992), led the student to engage with information received at lectures (Shulman, 1986; Rhemtula and Rollnick, 2002; Guterman, 2003), possibly influencing metacognition (Zohar and Nemet, 2002; Georghiades, 2000; 2004) or "cognitive dissonance" (Cho *et al*., 1985; 1.2) and the ensuing retrieval of correct information from mental schemata (Carey, 1986).

5.9 Discussion

5.9.1 Symbolic plasmid form and gene expression

A wide range of difficulties are presented on symbolic plasmid form and gene expression. Based on their prevalence (Table 5.1), the following are currently classified at level 2 as "anticipated" in accordance with the four-level classification scheme described by Grayson *et al*. (2001). The views that linear plasmids cannot be expressed because they "are difficult to transform", "take up too much space in a cell", "present an abnormal conformation that cannot be read by enzymes" or present a "disrupted flow of genetic information" are typical level 2 difficulties which the author suspected from his teaching experience. Unexpected, level 1 difficulties associated with the non expression of genes in linear plasmids included inability of linear plasmids to enter the nucleus of cells,

destruction of linear plasmids by endogenous nucleases and inability of linear plasmids to multiply (replicate) due to physical space constraints within a cell.

Viewpoints associated with plasmid linearity and positive expression of genes because "such DNA can be transformed easily" are classified as a level 2 difficulty. Unexpected level 1 difficulties are those that associated linear plasmid strands as unhindered templates for transcription. These included views that efficient gene expression is possible because:

i. "the sequences are intact and unhindered in a linear form of DNA".
ii. "the enzyme DNA dependent-RNA polymerase moves easily on a linear DNA template to facilitate transcription".
` iii. "linear DNA fits nicely in a cell".
iv. "linear DNA transcribes better because its straight form is less complex than supercoiled DNA".

With regard to the covalently closed circular (ccc) plasmid conformation (3.5.4), the belief that this form contains the genetic information within its circular structure is classified as a level 2 "anticipated" difficulty. Further, the assertion that its relaxed form allows for easy enzyme binding and subsequent transcription is also classified as a level 2 difficulty. The classification of difficulties associated with the use of the symbol, "**::**" has received attention in section 4.1.5. It remains confusing to students who are unable to differentiate its use to show in tandem gene fusion within a plasmid construct and the mechanism of transposition. This difficulty is classified at level 2 or "anticipated". Student S18 presents a difficulty with phenotypic expression by revealing that genes X and Z (α) contain the enzymes amylase and β-galactosidase, respectively. Problems associated with phenotypic expression are not uncommon (1.5) as some students believe naively (Caramazza *et al*., 1981) that properties or entities encoded by genes are found within the genes themselves (Lewis and Wood-Robinson, 2000; Lewis and Kattmann, 2004). This difficulty is placed at level 3 as "partially established" (Grayson *et al*., 2001). The substitution of functions between operator and promoter sequences, or that which

relates to alkaline phosphatase and not alkaline phosphorylase (5.4.1), is classified as a level 2 difficulty. Similarly, the use of the acronym, "ARS" and inability to provide its full name, meaning or function is also classified as a level 2 difficulty.

5.9.2 Map construction

The difficulty associated with practical aspects of map construction is classified at level 2 or "anticipated" in accordance with the four-level classification scheme described by Grayson *et al*. (2001). Difficulties (5.7.1), associated with electrode nomenclature and polarity, as well as the migration patterns of negatively charged components such as the dissociated ionic form of DNA, are now classified at level 3 or as "partially established". The need to devise practical exercises, encompassing the theory of this topic, has become evident (Szeberenyi, 2002; Walsh *et al*., 2007). Similar difficulties have also been reported in electrochemistry (Garnett and Treagust, 1992b; Ozkaya, 2002) where students were unable to explain the functions of anode and cathode in electrochemical cells or the effect of current on ion migration. The inappropriate transfer of theory relating to thin layer chromatography, and its amalgamation with that pertaining to electrophoresis, is classified as a level 1 or "unexpected" difficulty (Grayson *et al*., 2001). The student's inability to predict the conformation of an undigested plasmid which bands closely behind the supercoiled form is also classified as level 1 or an "unexpected" difficulty and, therefore, requires further research in order to be substantiated.

5.9.3 Gene markers and phenotypic expression

Three types of difficulties emerged from the current investigations on gene markers and phenotypic expression (5.8). The difficulty arising from the students' inability to explain phenotypic expression, coupled with the assertion that properties or entities are found in genes or gene markers, is classified at level 3 or as "partially established" in accordance with the four-level classification scheme by Grayson *et al*. (2001). Lewis and Kattmann

(2004) reported on a similar difficulty, highlighting a student's belief that chromosomes contain pigments for eye colour. The concept of phenotypic expression is generally poorly understood as students find difficulty linking "expression of a trait" with mechanisms such as transcription and translation (Rotbain *et al*., 2006). Students fail to understand that a codified protein may be responsible for a trait or phenotype (Fox, 1996; Lewis and Wood-Robinson, 2000; Lewis *et al*., 2000; Marbach-Ad, 2001; Lewis and Kattmann, 2004; Rotbain *et al*., 2006). The erroneous association of gene marker as an indicator of DNA size instead of the use of DNA molecular weight markers is classified as a level 2 or "anticipated" difficulty. The difficulty that gene markers are in fact DNA elements which grow when cells are cultured, is also classified as a level 2 difficulty.

CHAPTER 6

DIFFICULTIES WITH SYMBOLISM AND THE SELECTION OF HYBRIDOMAS AND TRANSFORMANTS

6.1 Introduction

Gene marker symbolism commonly illustrates an abridged format describing the function or property of a gene. A property or phenotypic trait can manifest in numerous ways (Chapter 2) and must be understood when used in molecular cell manipulation techniques. The selection of hybridomas and bacterial transformants receive attention in this study. Phenotypic traits can be manipulated in cell selection processes and this involves a clear understanding of the interplay between the various phenotypes represented by gene markers and factors influencing cell survival or death (Köhler and Milstein, 1975; Staines, 1983). Such factors can include, *inter alia*, epigenetic or environmental factors including chemicals or nutrients of growth media or metabolism which is influenced by enzymes making up complex pathways. The selection of both hybridomas and bacterial transformants requires the ability to explain the fate of cells under defined growth conditions. The fate of hybridomas, for example, is influenced in an integrated manner by gene-encoded properties, the ability or inability to assimilate media constituents and the expression of enzymes implicated in salvage biochemical pathways (3.5.6.2; Table 3.5). Similarly, a combination of host and plasmid-encoded properties, coupled with media constituents play a role in the selection of transformants (3.5.6.3). Students are tested in their ability to reflect on an interplay of multiple factors, engage in reasoning (de Bono, 1967, cited by TIP theories, 2005; Sins *et al*., 2005; Kang *et al*., 2005; Pata and Sarapuu, 2006) and predict the survival of such manipulated cells.

6.2 Difficulties with hybridomas and their selection

The free-response probe (3.5.6.2), on the cell fusion hybridoma technology, indicated that students (n_{1-4}) did understand both the concept of a "hybridoma" (Table 3.5) and the fusion methodology associated with its production. However, there appeared to be notable problems explaining the selection of the hybridoma, taking in effect the integrated role of gene marker information, biochemical pathways and media components. In total, the free-response probe yielded five categories of difficulties and varying incidences associated with each.

6.2.1 A failure to define abbreviations, acronyms and marker symbolism

Abbreviations, acronyms and marker symbolism, concerning the cell fusion hybridoma technology, are commonly used in literature (2.5.2; 3.5.6.2). This is generally acceptable following an indication of full names alongside the abbreviation or acronym at first mention. Their use obviates the need to spell out lengthy names which makes writing and reading a lot easier. However, the author has given repeated instructions to students that abbreviations or acronyms must always be defined when first mentioned. In this study, the category (i) difficulty reports on those students who are oblivious of the full meaning of such symbolism (Table 3.5), yet these find common use in their superficial written explanations concerning the subject matter. In response to the free-response probe, several examples of abbreviations and acronyms have been recorded unnamed. Some students are adept at memorising the following abbreviations but show difficulty in defining them. These included Hgprt and Tk in differently printed format with superscripts "+ "or " –" for the gene markers and for the salvage enzymes, hypoxanthine guanine phosphoribosyltransferase and thymidine kinase respectively, HX for hypoxanthine, PRPP for phosphoribosylpyrophosphate, IMP for inosine monophosphate, GMP for guanosine monophosphate, AMP for adenosine monophosphate, T for the nucleoside thymidine, HAT for components hypoxanthine, aminopterin and thymidine in HAT medium (3.5.6.2), dUMP for deoxyuridine monophosphate and dTMP for

deoxythymidine. This category of difficulty showed an incidence among groups n_{1-4} as follows; n_1 = 19/54; n_2 = 11/35; n_3 = 15/31 and n_4 = 13/44. Some students attempted to provide definitions for the abbreviations but these were characterised poor spelling mistakes and syntheses of new words (6.2.2).

6.2.2 Syntheses of new unfound words

The second category of difficulty presents the syntheses of new words (highlighted in bold) or spelling mistakes, vaguely resembling those encountered in the science of cell fusion technology. These included:

i. "**Heterones** of *Hgprt*$^+$ *Tk*$^+$ are viable in HAT medium".
Incidence: n_1 = 1/54

ii. "**Heterokaryones** are viable in HAT medium".
Incidence: n_1 = 1/54

iii. "HX is **hydroxyxanthine**".
Incidence: n_2 = 1/35

iv. HX is **hapoxanthine**".
Incidence: n_2 = 1/35

v. "HX is **hypoxonthine**".
Incidence: n_3 = 1/31

vi. "The enzyme HGPRT, which is **hydroxyglutarylphosphoribosylotransferase**, converts HX to IMP, using PRPP".
Incidence: n_2 = 1/35

vii. "**Aminoptic** acid is found in HAT medium with hypoxanthine and thymine".
Incidence: n_3 = 1/31

viii. "TK, the enzyme **thymidylic kinase** phoshorylates deoxythymidine".
Incidence: n_4 = 1/44

ix. "PRPP is **phosphorylase pyrophosphatidase**".
Incidence: n_1 = 1/54

The newly synthesized words, "heterones" and "heterokaryones", resemble "heterokaryons", transient cell fusion products which are highly unstable as nuclei of different parental cells are found in a common cytoplasm or cybrid state (Köhler and Milstein, 1975). These nuclei may be induced to fuse, generating a hybrid through karyogamy or they may promote the dissociation of parental progenitors with or without gene exchange (Gupthar and Garnett, 1987; Gupthar, 1987; 1989; 1992). The abbreviation "HX" is generally used in place of hypoxanthine and not "hydroxyxanthine", "hypoxonthine" or "hapoxanthine". HX is the purine precursor to IMP synthesis via the enzyme, hypoxanthine guanine phosphoribosyltransferase (HGPRT) and certainly not "hydroxyglutarylphosphoribosylotransferase" as indicated above. "Aminoptic acid" is unknown as a compound but finds its way substituting "aminopterin" as a component of HAT medium (3.5.6.2). Lastly, thymidine kinase phosphorylates the nucleoside thymidine to the nucleotide, thymidine monophosphate (3.5.6.2). "Thymidylic kinase" and "phosphorylase pyrophosphatidase" are unknown enzymes in the scientific literature. These forms of nomenclature are erroneous and may indicate a constructivist learning approach where scientific information is incorrectly learnt and expressed in ways which are undesired or unacceptable (Carey, 1986; Driver and Bell, 1986; Von Glasersfeld, 1992). The following incidence of this difficulty was recorded; n_1 = 3/54; n_2 = 3/35; n_3 = 2/31 and n_4 = 1/44.

6.2.3 Erroneous substitution of nomenclature and related facts

The third category of difficulty presents with numerous examples of erroneous substitution (Grayson, 1996; 2004) of nomenclature and related facts. This category is closely related to category (ii) except that individual words presented do exist in the scientific literature but find meaningless use in the current context. These include:

i. "TK is tyrosine kinase".
Incidence: $n_1 = 1/54$

ii. "IMP is inositol monophosphate".
Incidence: $n_4 = 1/44$

iii. "HGPRT is hypoxanthine guanine pyrophosphate thymidine".
Incidence: $n_4 = 1/44$

iv. "The enzyme thymidylate synthase converts thymine to dTMP".
Incidence: $n_2 = 1/54$

v. "TK^- is involved in the conversion of dUMP to dTMP, ie. TK phosphorylates the dUMP".
Incidence: $n_2 = 1/54$

vi. "Thymidine kinase converts T to THF".
Incidence: $n_3 = 1/54$

vii. "Thymidine kinase converts T (thymine) to dTMP using deoxyribose sugar".
Incidence: $n_1 = 1/54$

viii. "T + ATP + TK → HGPRT .....this is a salvage reaction".

Incidence: $n_4 = 1/54$

ix. "Thymidylate synthase converts DHF to THF" ("DHF – dihydrofumarate, THF- tetrahydrofumarate".
Incidence: $n_2 = 1/54$

x. "Thymidine kinase is involved as follows:

$$G + ATP \xrightarrow{TK} IMP$$ ".

Incidence: $n_2 = 1/54$

Examples (i)-(iii) above show the erroneous and meaningless substitution of scientific vocabulary to abbreviations that find common use in the field of the hybridoma cell fusion technology (3.5.6.2). As aminopterin inhibits the conversion of dUMP to dTMP via thymidylate synthase, the salvage enzyme thymidine kinase converts the nucleoside deoxythymidine to the nucleotide dTMP, utilizing ATP to effect this phosphorylation. Instead, some students substituted deoxythymidine with either the free base thymine, dUMP, guanine or guanosine (undifferentiated G) as substrates for thymidine kinase. Reactant and product combinations linked to this enzymatic reaction included guanine / guanosine (undifferentiated G) plus ATP to generate IMP and, thymine / thymidine (undifferentiated T) plus ATP to generate HGPRT. Other nonsensical equations linked to thymidine kinase included:

i. The conversion of thymine or thymidine (undifferentiated T) to THF (assumed abbreviation of tetrahydrofolate, a derivative of which (3.5.6.2) is implicated in the conversion of dUMP to dTMP via thymidylate synthase).
ii. The conversion of thymine plus deoxyribose sugar to dTMP.

In addition, there appeared to be confusion between the activities of thymidine kinase and thymidylate synthase. Example (iv) displays two substitution errors, firstly the enzyme thymidylate synthase is used in place of thymidine kinase and secondly, the substrate

thymine is used instead of thymidine. Example (v) shows the erroneous substitution of thymidylate synthase by thymidine kinase facilitating the conversion of dUMP to dTMP. This is in fact a methylation-type reaction (3.5.6.2) and not a phosphorylation reaction as indicated by the student. Example (ix) carries two substitution errors. Firstly, the enzyme dihydrofolate reductase is implicated in the reduction of dihydrofolate (DHF) to tetrahydrofolate (THF), using NADPH as the reductant (3.5.6.2), and certainly not thymidylate synthase as indicated by the student. However, it is well-known that THF which converts to N^5, N^{10}- methylene tetrahydrofolate , using serine or glycine, acts as a methyl group donor in the conversion of dUMP to dTMP via thymidylate synthase (3.5.6.2). The assignment of nomenclature such as "dihydrofumarate" and "tetrahydrofumarate" to the abbreviations DHF and THF, respectively also constitutes unwanted syntheses and substitution in example (ix).

Incidence: $n_1 = 2/54$; $n_2 = 4/35$; $n_3 = 1/31$ and $n_4 = 3/44$

6.2.4 Poor understanding of gene marker symbolism

Category (iv) difficulties are characterized by poor conceptual understanding of a "gene marker" (Table 3.4). The following statements were isolated from students' responses to the free-response probe (3.5.6.2):

i. "*Hgprt*$^+$ is able to take up toxic levels of 8-azoguanidine and 6-thioguanidine and convert it to GMP which results in cell death".
   Incidence: $n_3 = 1/31$

ii. "*Hgprt*$^+$ / *Tk*$^-$ cannot grow in HAT medium".
   Incidence: $n_1 = 2/54$; $n_2 = 1/35$; $n_3 = 1/31$ and $n_4 = 1/44$

iii. "To ensure that *Hgprt*$^-$ does not revert to *Hgprt*$^+$ one must use media with ^{3}H- hypoxanthine".
   Incidence: $n_4 = 1/44$

iv. "*Hgprt*$^{+}$ is the enzyme involved in the synthesis of purine derivatives, guanine, hypoxanthine and adenine".
Incidence: n_3 = 1/31

v. "*Hgprt*$^{-}$ / *Tk* $^{+}$ will not grow in HAT medium".
Incidence: n_1 = 2/54; n_2 = 1/35; n_3 = 1/31 and n_4 = 1/44

vi. "*Tk*$^{+}$ is the enzyme involved in the synthesis of pyrimidine bases thymidine and uracil".
Incidence: n_1 = 1/54

vii. "*Hgprt*$^{+}$ / *Tk* $^{+}$ will grow in HAT medium".
Incidence: n_2 = 2/35 and n_4 = 1/44

The concept of a "gene marker" appears to be elusive to some students. There are perceptions that markers can be grown in a medium, a property indicated by the superscript "+" as shown in the example, "*Hgprt*$^{+}$ / *Tk* $^{+}$ ". In addition, there appears to be a false correlation that the superscript "-", as shown in "*Hgprt*$^{-}$ / *Tk* $^{+}$ ", denotes a lack of growth of the marker in a growth medium. Interviews were conducted to verify, whether or not, the lack of growth was associated with the combination of both complementary markers. There are also differentiation problems illustrated by the students' reference that *Hgprt*$^{+}$ and *Tk* $^{+}$ are enzymes and not the marker. Examples (iv) and (vi) above show this difficulty, coupled with an indication of erroneous enzyme activity. The enzyme hypoxanthine guanine phosphoribosyltransferase (HGPRT) converts hypoxanthine (HX) to inosine monophosphate (IMP) using phosphoribosylpyrophosphate (PRPP) or the free base guanine to guanosine monophosphate using PRPP (3.5.6.2). HGPRT is not associated with the synthesis of hypoxanthine, guanine and adenine as indicated above. Thymidine kinase phosphorylates thymidine and has nothing to do with the synthesis of the same nucleoside or the free base uracil as indicated in example (vi) above. Further, example (i) presents *Hgprt*$^{+}$ as an entity or presumably a cell capable of incorporating toxic analogues of guanine, which

should read 8-azaguanine (8-AZG) and 6-thioguanine (6-TG). These base analogues may be incorporated into the respective toxic nucleotide forms, 8-AZGMP and 6-TGMP, in the presence of phosphoribosylpyrophosphate and HGPRT. Generally, the stability of the *Hgprt*$^{+}$ marker may be tested on a small fraction of cells which die as a result of the above incorporation (Köhler and Milstein, 1975; Staines, 1983; Kornberg and Baker, 1992; Stryer, 1995). The claims made in statement (iii) above, which presumably applies to a cell, that the reversion or back-mutation of *Hgprt*$^{-}$ to *Hgprt*$^{+}$ may be averted using ^{3}H- hypoxanthine in a medium, cannot be supported by relevant scientific literature.
Incidence: n_1 = 5/54; n_2 = 4/35; n_3 = 4/31 and n_4 = 4/44

6.2.5 Poor integration of knowledge and problems with reasoning

Category (v) difficulties concern the poor integration of knowledge (Thorndyke and Stasz, 1985; Sandoval and Reiser, 2004) and problems with reasoning (Taconis *et al*., 2001; Kang *et al*., 2005). A higher incidence of such difficulties was recorded (n_1 = 17/54; n_2 = 6/35; n_3 = 7/31 and n_4 = 10/44) in comparison with those placed in the other four categories. The nature of category (v) difficulties is illustrated by examples of the following statements which unfortunately provide no further explanation on the hybridoma selection process:

i. "Cells which are *Hgprt*$^{-}$ / *Tk* $^{+}$ die in HAT medium because the enzyme *Hgprt* is not produced, so the purine nucleotides cannot be synthesized. The hybridoma *Hgprt*$^{+}$ / *Tk* $^{+}$ produces both salvage enzymes, so all purine and pyrimidine nucleotides can be produced".

ii. "We want to select for a B cell fused to a myeloma cell. So we grow the cells in HAT medium containing hypoxanthine, aminopterin and thymidine which supports the growth only of the fused cell".

iii. "Hybridomas will survive in HAT medium because they produce the enzymes Hgprt and Tk".

iv. "*Hgprt*$^{+}$ and *Tk* $^{+}$ grow in the hybridoma therefore it survives in HAT medium".

There is a clear indication from the above that some students find it difficult to integrate the role of media components, gene markers and biochemical pathways in determining the fate of cells after the fusion process (3.5.6.2). Statement (i) correctly associates hypoxanthine guanine phosphoribosyl transferase with purine nucleotide biosyntheses but gives no detail on the salvage purine pathway nor on the role of aminopterin or hypoxanthine in the selection medium. The submission that the hybridoma produces both salvage enzymes to effect purine and pyrimidine nucleotide biosynthesis is correct, however, this information is rather superficial. Specific detail is required on the inhibitory mechanism of aminopterin and the role of medium supplements hypoxanthine and thymidine in biosynthetic pathways (3.5.6.2). Statement (ii) provides no mechanism which supports the survival of the hybridoma except a mention of hypoxanthine, aminopterin and thymidine which selectively allow for hybridoma proliferation. Statement (iii) suggests that hybridoma survival is due to the salvage enzymes but no further information is given as to how this happens. Statement (iv) erroneously indicates that the growth of markers in the hybridoma invariably contributes to its survival.

6.2.6 Recurring difficulties

The more focused probe (3.5.6.2) yielded the same five categories of difficulties; however, these presented with some variation in the incidence as shown in Table 6.1. In comparison with the free-response probe, a general decrease in the incidence of the difficulties was recorded. This could be attributed to an ongoing learning process among the students. Difficulties, characteristic of categories (i) to (v) (Table 6.1), re-emerged among all four student groups and will not be discussed in great detail to avoid

Table 6.1 Students' difficulties with the hybridoma cell fusion technology

Category	Type of difficulty and incidence
i.	Use of abbreviations, acronyms and marker symbolism without definition Incidence: n_1 = 15/54 n_2 = 8/35 n_3 = 10/31 n_4 = 10/44
ii.	Syntheses of new words, some vaguely resembling those encountered in the science of cell fusion technology Incidence: n_1 = 3/54 n_2 = 3/35 n_3 = 2/31 n_4 = 1/44
iii.	Erroneous substitution of nomenclature and related facts Incidence: n_1 = 2/54 n_2 = 2/35 n_3 = 1/31 n_4 = 1/44
iv.	Poor understanding of the concept of a marker and symbolism as reflected by $Hgprt^+$ / Tk^- or $Hgprt^-$/ Tk^+ Incidence: n_1 = 1/54 n_2 = 2/35 n_3 = 3/31 n_4 = 4/44
v.	Poor integration of knowledge and problems with reasoning Incidence: n_1 = 13/54 n_2 = 6/35 n_3 = 4/31 n_4 = 8/44

Total student number in groups n_{1-4}

n_1 = 54 n_2 = 35 n_3 = 31 n_4 = 44

duplication. For example, abbreviations and acronyms such as Hgprt, Tk, IMP, HAT, PRPP, dUMP, HX, DHF, THF and dTMP were used without definition as described previously although there have been repeated instructions to the students that abbreviations must always be defined when first used. The syntheses of new words included "hydroxyanthen", "hypoxonthine", "hapoxanthine", "hyperzanthine", "aminoptic acid", "heterones" and "heterokaryones". Category (iii) and (iv) difficulties relating to erroneous substitution and misunderstanding of the concept of a gene marker were also prevalent in the exact form as described for the free-response probe. The focused probe yielded more detailed answers on the mechanism of hybridoma selection but some of these were characterized by poor explanations on the integrated role of media components, gene marker symbolism and biochemical pathways, illustrating imprecise and alternate conceptual frameworks (Duchovic, 1998; Taber, 1998). The following excerpts from student responses illustrate the above.

i. "Parent cells do not survive in HAT medium. They have a complimentary gene make-up, such as $Hgprt^{+}\ Tk^{-}$ or $Hgprt^{-}\ Tk^{+}$, meaning they can grow one enzyme only but not the other. The hybridoma which is a fusion of a B cell and a myeloma (cancerous, long-living B cell) grows both enzymes therefore it is $Hgprt^{+}\ Tk^{+}$. These enzymes allow for use of hypoxanthen and thymine from HAT medium to grow the cells. Cells which do not produce these enzymes die because aminopterin in HAT medium is a poison".

ii. "HAT medium allows for growth of the hybridoma and not the parent cells which fuse to form the hybridoma. The hybridoma is not affected by aminopterin which can block the enzyme Tk so that it becomes Tk^{-} and also Hgprt which becomes $Hgprt^{-}$. The hybridoma ($Hgprt^{+}$ Tk^{+}) produces both enzymes. Hgprt produces IMP from HX and PRPP and Tk produces TMP from T using ATP. Parent cells which have either $Hgprt^{-}$ or Tk^{-} cannot form IMP and TMP therefore they die in HAT medium".

Excerpt (i) indicates that the gene markers, *Hgprt*$^{+}$ *Tk* $^{-}$ or *Hgprt*$^{-}$ *Tk*$^{+}$, simply translate to a cell's ability (+) or inability (-) to "grow" an enzyme. The symbolism which applies to gene markers and enzymes are neither defined nor differentiated. HAT medium components are given erroneously as "hypoxanthen" and "thymine" instead of hypoxanthine and thymidine (3.5.6.2). Cell survival is linked to the production of salvage pathway enzymes which counteract the effect of a "poison" (aminopterin). The inhibitory role of aminopterin is not discussed. There is also poor reasoning and a failure to link cell survival or death to gene markers, biochemical pathways and media components in an integrated manner (3.5.6.2).

Excerpt (ii) demonstrates the use of gene marker symbolism which is undefined. The role of aminopterin and concept of a gene marker are misunderstood. Aminopterin appears to be associated with the blockage of salvage enzyme activity and not *de novo* purine and pyrimidine biosynthesis (3.5.6.2). Gene marker notations are erroneously assigned to describe this blockage.

The following interviews also show recurring difficulties with the concept of gene marker symbolism and the role of such markers in the hybridoma selection process.

I: I want to focus on your written effort [pause] where you indicate that markers are grown in HAT medium. Here you indicate that "Hgprt $^{-}$ */ Tk* $^{+}$*" denotes a lack of growth of the marker in a growth medium.*
S24: Yes.
I: Which marker?
S24: Both Hgprt and Tk because one of them in the combination has a minus sign.
I: What is a marker?
S24: It can be a gene which shows a property like enzyme production.
I: Why do you say that a marker is grown or not grown in a medium?
S24: It is found in a cell which grows or not grows [hastily]. I mean it grows if the cell grows.

I: Have you come across the term complementary markers and perhaps the reason as why they are used?
S24: No. I have no idea.
I: Let us take the selection of cells in HAT medium [pause], are you troubled by any aspect concerning the hybridoma selection?
S24: It's too complex. You have markers which ensure cell survival if they produce enzymes like hypoxanthine guaninepyro [Eish!] its too complex. Then you have thymidine kinase.
I: What do these enzymes do?
S24: Ehh really Sir, I can't give any answer. I mean there is a link to pathways and the HAT medium but I find this thing confusing.

Student S24 indicates that a marker grows concomitantly with the cell, producing an enzyme. A marker gene is defined superficially, precluding any reference to the mechanism of gene expression in the statement, "*It can be a gene which shows a property like enzyme production"*. Student S24 reveals difficulty with enzyme nomenclature and the selection of the hybridoma. Hybridoma selection requires knowledge of pathways, the role of the medium supplements and the use of gene markers to achieve selection. Being able to explain these "elements" in an integrated manner (3.5.6.2) involves "lateral thinking" (de Bono, 1967, cited by TIP Theories, 2005) (Website in reference list) and a clear consolidation of the knowledge domain which applies to each is essential. Predicting cell survival is an elaboration of a composite mental scheme which could be cognitively demanding especially when the individual knowledge domains are poorly structured (Kirschner, 2002; Kozma, 2003; Kang *et al*., 2005).

The following interview illustrates erroneous substitution (Grayson, 1996). However, when challenged metacognitively (Georghiades, 2004), student S25 responds positively, giving precise information on the enzymology influencing hybridoma selection.

I: Let us talk about your written effort on hybridoma selection. I notice that you describe the cell types correctly [pause] but there is some confusion when you relate the

mechanism of selection. Here you show that aminopterin blocks the conversion of dUMP to dTMP via thymidine kinase.

S25: Yes.

I: What does a kinase do?

S25: Which kinase, Sir?

I: Any kinase you know [pause] take thymidine kinase.

S25: It is linked to phosphorylation.

I: Excellent! Now take a look at the conversion of dUMP to dTMP. Does it involve phosphorylation?

S25: Oh! It should be thymidylate synthase and the methyl transfer reaction.

I: Tell me more about the reaction.

S25: I was confused between thymidine kinase and thymidylate synthase and the blockage at..at [stammer] the thymidylate synthase reaction by aminopterin.

I: But what about this methyl transfer reaction you mention earlier.

S25:Dihydrofolate is reduced to a tetrahydrofolate form by dihydrofolate reductase [pause] you see this is inhibited by aminopterin therefore you can't get methyl group transfer from the tetrahydrofolate donor.

I: What does the methyl group transfer do?

S25: It is needed to convert dUMP to dTMP [pause] I mean T in TMP is a methylated uracil.

I: Excellent. See me again should you have problems with this section.

Student S25 substitutes the enzyme thymidylate synthase with thymidine kinase in a reaction involving methyl group donation (3.5.6.2). When challenged to give the function of a kinase, S25 correctly associates this with phosphorylation but recognizes that the reaction being considered does not involve phosphorylation. Word association can be regarded as a cue (White and Gunstone, 1992) in this instance. S25 makes further input on the methylation reaction but gives no information on the inhibitory role of aminopterin on thymidylate synthase. However, the inhibitory role of aminopterin is mentioned in a second reaction involving dihydrofolate reductase. In addition, N^5, N^{10} - methylene tetrahydrofolate is not mentioned by full name as the methyl group donor although the

student understands the concept of methylation. "Cognitive load" can impact on the learning process, exposing the student's inability to cope with the assimilation of the apparently more intricate information (Zohar and Nemet, 2002; Kirschner, 2002).

6.3 Interpretation of symbolic gene markers and the selection of transformants

6.3.1 Varying levels of understanding

The free-response probe (3.5.6.3) demonstrated that all students (n_{1-4}) had a good understanding of the term "transformant" (Table 3.5). Students provided a range of descriptions, essentially highlighting a cell's receipt or accommodation of a plasmid. No students referred to the cells' accommodation or receipt of a fragment of DNA. Examples of scientifically sound conceptions included the following:

i. "A transformant is a cell which receives a plasmid and changes as a result because it expresses a gene found on the plasmid".
ii. "Transformants are cells (e.g. bacterial cells) which receive a plasmid by heat-shock treatment".
i. Transformants are those cells that can keep a plasmid once they receive it".
iv. "Transformants are cells receiving a plasmid".
v. "When a host cell receives a plasmid, it becomes a transformant".

The second part of the free-response probe (3.5.6.3) yielded a range of answers which could fit three different categories. Category (i) answers focused on transformant selection with specific reference to different plasmid-encoded properties and phenotypic traits of the transformant. The incidence of students who gave category (i) answers, characterised by a satisfactory explanation of the selection mechanism, was as follows; n_1 = 50/54, n_2 = 33/35, n_3 = 30/31 and n_4 = 41/44. Examples of such acceptable responses included:

i. "Transformants are host cells which receive a plasmid. They may be selected depending on the gene expressed by a plasmid. If a host is auxotrophic for leucine (host *leu*$^-$) and we put a plasmid expressing a leucine biosynthesis gene in the transformant then the transformant can make its own leucine. It will survive in a medium without leucine but the host cannot. The transformant is said to be prototrophic for leucine (transformant *leu*$^+$)".

ii. "Transformants are host cells which retain a plasmid once they are transformed with a plasmid. If we have an amylase gene in the plasmid, then the transformants can break starch to glucose by producing amylase but the original host cells which are perhaps not starch-utilizing cannot break starch because they don't have a plasmid with the amylase gene".

iii. "A transformant (tetracycline-sensitive) receiving a plasmid with tet^r gene will develop resistance to the antibiotic tetracycline, therefore such a transformant can be selected on agar nutrient medium with tetracycline in it. If the transformant loses the plasmid it will die in this tetracycline medium".

The following categories of answers demonstrated misunderstanding of the concept of a gene and its expression. Category (ii) answers, characterised by an incidence of $n_1 = 2/54$, $n_2 = 0/35$, $n_3 = 1/31$ and $n_4 = 2/44$, made reference to transformant selection on the basis that transformant-acquired genes actually contain an entity such as an enzyme, antibiotic or amino acid and that these aid selection as follows.

i. "When a transformant receives a plasmid with the amylase gene, it can be selected on a starch medium because the gene contains amylase which can break the starch to glucose".

ii. "Transformants which keep a plasmid containing ampicillin in its ampicillin gene will survive when we grow it in a medium containing ampicillin".

iii. "A transformant (*leu*$^-$) with a plasmid can be selected on a leucine-free medium if the plasmid releases leucine found in the leucine gene".

Category (iii) answers, characterised by an incidence of n_1 = 0/54, n_2 = 2/35, n_3 = 0/31 and n_4 = 1/ 44, wrongly indicated that transformant selection could be achieved on different media as transformants grew gene markers or gene products. Examples of such alternative conceptions included the following:

i. "Transformants are modified cells. They contain plasmids which grow specific genes and gene products as the cells grow. A tetracycline gene grows with the cell therefore tetracycline accumulates in a medium where this transformant will survive".

ii. "An adenine gene of a plasmid can grow adenine as the transformant grows. Therefore it is possible to select such a transformant on a medium lacking adenine".

iii. "Transformants keeping a leucine plasmid will grow or multiply the leucine gene because of a high copy number. More leucine genes can release more leucine in the cell transformant. Transformant selection is possible because more leucine genes grow and more leucine is formed".

The difficulties, shown in categories (ii) and (iii), arise from a poor understanding of the concept of a gene and its expression as shown previously (5.8.1). When protoconcepts (Fisher, 1985) are not understood, the more complex issues relating to transformant selection are also poorly expressed. In essence, uncompromised "epistemic scaffolding"

depends on the understanding of basic knowledge which underpins the more complex (Shulman, 1986; Sandoval and Reiser, 2004; Sins *et al*., 2005).

6.3.2 Varying reasoning ability

Reasoning about a selection mechanism requires an integrated understanding of host phenotypes, their possible change due to gene acquisition or expression of plasmid-borne genes and finally, the role of media supplements in the selection or non selection of transformants. This was facilitated using the more focused probe (3.5.6.3) which provided useful data, supplementing the probe work carried out on plasmid conformation and function (3.5.4). The results were placed in different categories as follows.

Category (i) presents correct answers characterised by an integrated understanding of host or transformant phenotypes, plasmid-encoded traits and the most likely effect of media supplements on cell survival. The answers which fit category (i) demonstrate the students' ability to engage in "lateral thinking" (de Bono, 1967, cited by TIP Theories, 2005) (Website in reference list) where factual information pertaining to gene markers, plasmid-encoded traits and media supplements are initially considered as components or "elements" and subsequently as a whole or "recombination" (Perkins and Salomon, 1989, cited by Zohar and Nemet, 2002; Oh, 2005), thereby enabling an analytic comparison of their influences on cell survival. The incidence of category (i) responses was n_1 = 30/54, n_2 = 28/35, n_3 = 19/31 and n_4 = 24/ 44. Some examples of these scientifically sound conceptions included:

i. "The host will not survive in a medium with X-gal + leucine + tetracycline because it is *Lac(z)* $^{-}$ and *tet* ^{S}, meaning that it cannot produce β-galactosidase to break X-gal to a bromo-chloro indole dye plus usable galactoside and it is sensitive to the antibiotic tetracycline. The host is auxotrophic for leucine so the addition of leucine will help its requirement

for the amino acid but X-gal and tetracycline will most likely inhibit its growth".

ii. "The transformant can grow in the X-gal/leucine/ampicillin medium because it expresses the plasmid gene Z for β-galactosidase which breaks X-gal to an indole-type dye plus galactoside which it can use and Amp^r encodes resistance to ampicillin. The presence of leucine does not inhibit growth because it is a nutrient but the transformant can make its own leucine because it has the plasmid *LEU* gene".

iii. "The host *E.coli* Tet^S *Lac (z)* $^-$ *leu*$^-$ Amp^S *X* $^-$ may not grow in a medium with starch minus leucine because it does not express the amylase gene to break starch to glucose and the host cannot make its own leucine, it is auxotrophic for leucine (*leu*$^-$)".

Category (ii) presents incorrect answers characterized by a lack of understanding of host or transformant phenotypes, plasmid-encoded traits and the effect of media supplements on cell survival. The incidence of category (ii) responses was n_1 = 12/54, n_2 = 5/35, n_3 = 10/31 and n_4 = 17/ 44. Some selected examples, that illustrate student difficulties, are shown below:

i. "The host does not have leucine in its leucine gene and is sensitive to the antibiotic tetracycline, therefore it should die in a medium with Xgal, leucine and tetracycline. It may take up the leucine in the medium and break Xgal to X + gal because the Z gene contains enough β-galactosidase to break Xgal. X is a blue dye which can make the host blue".

ii. "The transformant will not grow in the medium with starch minus leucine because it needs leucine and glucose which it can produce from starch since gene X codes for amylase".

iii. "The transformant will grow on the starch-leucine-ampicillin medium because the transformant has plasmid genes which recognize starch, leucine and ampicillin. These genes are *X, LEU* and AMP^r".

Statement (i) demonstrates a poor understanding of the concept of a gene marker as reference is made to the containment of an entity a gene is meant to express. There is limited comment on all the gene markers characterizing the host and their influence on the host's survival in a specific medium. X-gal is poorly characterized as a chromogenic substrate. Statement (ii) illustrates that plasmid-encoded traits are partly recognized in the transformant. The transformant's ability to convert starch to glucose is linked to gene *X*; however, the reversal of leucine auxotrophy to prototrophy is omitted, erroneously affecting the interpretation of the transformant's survival in a starch medium devoid of leucine. Statement (iii) illustrates constructivism (Driver and Bell, 1986; Von Glasersfeld, 1992) in the sense that specific genes *X, LEU* and Amp^r, associated with amylase production, leucine biosynthesis and resistance against ampicillin (3.5.6.3), respectively are erroneously linked to "recognition" of medium supplements.

Category (iii) type of difficulties represent those students who could not give any answer to the focused probe (3.5.6.3) perhaps influenced by their limited understanding of host or transformant phenotypes, plasmid-encoded traits or the effect of media supplements on cell survival. The incidence of responses reflecting no written information to the focused probe was $n_1 = 12/54$, $n_2 = 2/35$, $n_3 = 2/31$ and $n_4 = 3/44$. Interviews were conducted to establish the nature of difficulties experienced by the students. In comparison with the free-response probe, the focused probe posed a greater cognitive challenge to the students as the phenotypic properties were made more complex. Certainly, it did appear from the results of the free-response probe that many more students could reason the mechanics of transformant selection using various simpler, single phenotypes of their own choice.

The following interviews yielded additional information on students' misunderstanding of gene marker symbolism.

I: I want to help you later but let us see how much you know regarding gene markers and transformant selection.

S26: Sure. Ha, ha , ha [short laughter]. I find it confusing.

I: Lets take the host…[Interviewer writes on a sheet of paper,"host E.coli Tet^S *Lac (z)*$^-$ leu^- Amp^S X^- *"]. Firstly, what is a gene marker?*

S26: [Points to Tet^S *Lac (z)*$^-$ leu^- Amp^S X^- *on sheet of paper] These are gene markers. They tell you more about genetic properties, in this case the host.*

I: Excellent! But what are Tet^S *and* Amp^S *?*

S26: Those are gene markers that tell you that the host is sensitive to tetracycline and ampicillin.

I: Correct! And Lac (z)$^-$ *and* leu^- *?*

S26: These are markers which do not contain β-galactosidase and leucine. Here's the negative sign that shows it [S26 points to superscript on each].

I: Tell me about X^- *.*

S26: X^- *eh it [pause] it won't have amylase.*

I: Do you understand the supplements found in the various media?

S26: Yes.

I: Do you understand the gene markers of the transformant ?

S26: Yes.

I: Can you give me an idea of the fate of the transformant in the X-gal-leucine-tetracycline medium?

S26: The transformant has the Z gene in the plasmid which contains β-galactosidase therefore it can use X-gal. It cannot make its own leucine so leucine in the medium will make it grow. Tetracycline will not affect the transformant because the tetracycline sensitivity gene is removed and replaced with gene X.

I: Thank you, my friend. Let us schedule a suitable time to re-visit this issue.

Student S26 reveals difficulties with the concept of a gene marker, erroneously associating the gene-encoded product as an entity found in the gene. The student correctly identifies gene Z as plasmid-borne but associates *β-galactosidase* containment in the gene. The action of this enzyme on the chromogenic substrate X-gal is not explained. The superscript "-" is erroneously interpreted in markers *Lac (z)*$^{-}$, *leu*$^{-}$ and *X*$^{-}$ as genes lacking the entity they encode. The reference to the transformant's inability to synthesize its own leucine is untrue as this trait is also influenced by the plasmid. Clearly, multiple markers pose recognition difficulties. S26 makes reference to the excision of a "tetracycline sensitivity" gene (3.5.4) instead of *Tet*^{r} gene and fails to recognize that tetracycline sensitivity is a genomic trait of the host (3.5.6.3).

The gene marker superscript "-" can denote the non production of enzymes β-galactosidase and amylase as shown by the markers *Lac (z)*$^{-}$ *and X*$^{-}$, respectively or it can show the auxotrophic status of a host (*leu*$^{-}$). Its association with the absence of a gene as revealed by student S27 in the following interview constitutes the formation of an alternative mental framework (Taber, 1998).

I: Let's take a look at transformant selection. I know that you did not provide written answers in your script.
S27: Okay. I am having trouble with it.
I: What kind of trouble?
S27: I know about gene markers which show a property but this media issue and how the host changes to a transformant and becomes selected is really confusing. I ah! I [stammer] mean predicting this survival thing is tricky.
I: Let's take the host [Interviewer writes on a piece of paper, "host Tet^{S} *Lac (z)*$^{-}$ *leu*$^{-}$ *Amp*^{S} *X*$^{-}$*"]. Can you give me the meaning of these gene markers?*
S27: Firstly these are genomic markers.
I: Yes, yes. Go on.
S27: Tet^{S} *is the tetracycline sensitive gene marker. Lac (z)*$^{-}$*, leu*$^{-}$ *and X*$^{-}$ *means that the Z, leu and X genes are not found in the host. Amp*^{S} *means that the host is sensitive to ampicillin.*

I: Tell me how the host properties change in the transformant and why?
S27: The transformant has the plasmid which codes for β-galactosidase and amylase. Therefore, the transformant produces these enzymes.
I: Okay. What other genes on the plasmid will change the properties of the transformant?
S27: There's the ampicillin resistance gene and leucine biosynthesis gene. Oh! I see it now, I didn't think about these making the transformant ampicillin resistant and able to synthesize leucine. I m'mean [stammer] the host does not have a leucine gene so it can't make leucine.
I: Let us take the medium with X-gal, leucine and tetracycline [pause]. Do you think that the transformant will survive in this medium?
S27: [Silence for at least 10 seconds] I don't know.
I: Can you explain survival or death of the transformant on the basis of gene markers and the medium supplements?
S27: This I find too complex [S27 appears to be exhausted].
I: Okay. Let us stop here. I want to see you again sometime. I am concerned about your difficulties with this section.

Student S27 fails to explain the action of enzymes on the medium substrates and furthermore is unable to relate (Yore and Treagust, 2006) the mechanism of survival taking into account multiple gene notations and medium supplements. The elemental knowledge domain, concerning gene markers, is not fully consolidated in this case and its application in analytical reasoning manifests in a cognitively-demanding task (Kirschner, 2002; Kozma, 2003). The candidate is unable to relate fully the effect of plasmid-borne genes on the transformant. There are also difficulties explaining phenotypic expression or gene-encoded traits (Lewis and Kattmann, 2004). The student associates the "-" superscript with the absence of the respective genes in the host.

6.4 Discussion

With the exception of category (iv), all other categories of difficulties (Table 6.1) on the hybridoma cell fusion technology (3.5.6.2) are currently classified at level 2 or "anticipated" from teaching experience in accordance with the four-level classification scheme described by Grayson *et al*. (2001). The current study provides the first report on student difficulties with the hybridoma cell fusion technology. Category (iv) difficulty concerns the interpretation of symbolic representations of gene markers. As shown previously, the concept of a gene marker and expression of a trait are widely misunderstood in different areas of Science (1.5). For this reason, it is appropriate to classify this difficulty at level 3 or as "partially established" (Grayson *et al*., 2001).

The nature of difficulties, associated with transformant selection, is currently classified as "anticipated" or level 2 in accordance with the four-level classification framework presented by Grayson *et al*. (2001). As can be seen from the analyses, difficulties arise from a poor understanding of the concept of a gene marker and phenotypic expression (6.3), a phenomenon reported by many other researchers (Fox, 1996; Lewis and Wood-Robinson, 2000; Lewis *et al*., 2000; Marbach-Ad, 2001; Lewis and Kattmann, 2004; Chapter 1).

CHAPTER 7

STUDENTS' UNDERSTANDING OF THE SYMBOLIC LANGUAGE ASSOCIATED WITH DNA REPLICATION

7.1 Introduction

Symbolic forms of DNA, such as models and diagrams (Beltramini *et al*., 2006; Cook *et al*., 2006; Rotbain *et al*., 2006) illustrate a wealth of information on various structural components making up the molecule (Chapter 2). Such representations should ideally provide a visual cue (Takayama, 2005; Crisp and Sweiry, 2006) of the underlying (Kozma, 2003) molecular detail that they depict. In this context, students are expected to interpret the function of structural components and the role they play in processes such as transcription and replication. The use of diagrams becomes meaningful when accompanying theoretical knowledge (Cheng *et al*., 2001; Tsui and Treagust, 2003) is correctly expressed about the form of representation. A visual cue also requires prior learning of structural forms of DNA and the knowledge that such a bio-molecule could be represented in multiple forms (Seufert, 2003; Schnotz and Bannert, 2003; Schönborn and Anderson, 2006a).

In this study, students were tested on their differentiation of template and coding strands of DNA in terms of function in replication and transcription (Rao, 1996; Scism, 1996; Table 3.6). Using schematic diagrams of DNA replication intermediates, students were expected to interpret the structural features shown in various symbolic depictions of DNA. Specifically, their knowledge of the molecular mechanism of replication was sought, based on an understanding of elemental components of each diagram (Perini, 2005; Cook *et al*., 2006). The molecular mechanism contributing to change in structural form or component is also analysed in various symbolic representations.

7.2 Difficulties with nomenclature and function of nucleic acid templates

The free-response probe (3.5.7) investigating students' interpretation of the nomenclature and function of nucleic acid templates (Table 3.6) generated five categories of difficulties, each accompanied by an incidence as shown in Table 7.1.

7.2.1 Lack of knowledge

In response to parts 1 and 2 of the free-response probe (3.5.7), thirty three percent of the total number of students among groups $n_{1\text{-}4}$ showed the category (i) difficulty, characterised by an inability to provide answers for both strand nomenclature and function. Presumably, these students had no knowledge of the subject matter, or chose not to answer because of doubt or confusion arising from the inability to select the required information from detailed notes. Further investigations were carried out via the interview route (7.3) to establish the nature of their difficulties.

7.2.2 Lack of understanding

With reference to the difficulty placed in category (ii), the following student responses also demonstrated a lack of understanding of specific strand function in replication and transcription (Figure 3.4):

i. " A - template strand. This is the original parental strand and the template is used to synthesize a new strand. B - coding strand. This strand codes from the template and the correct bases are added according to the template".

ii. "A - template strand. It moves in a 5→3' direction.
B - coding strand. It moves in a 3→5' direction".

Table 7.1 Description of student difficulties with DNA-strand nomenclature and function

Category	Type of difficulty and incidence
i.	Failure to name and differentiate the coding and template strands on the basis of function Incidence: n_1 = 20/54; n_2 = 5/35; n_3 = 15 /31; n_4 = 15/44
ii.	Interchange of the labelling of the DNA strands with incorrect function Incidence: n_1 = 4/54; n_2 = 2/35; n_3 = 2/31 ; n_4 = 2/44
iii.	Substitution of coding and template strand labels with labels characterizing DNA replication intermediates Incidence: n_1 = 13 /54 ; n_2 = 4/35; n_3 = 7 /31; n_4 = 10/44
iv.	Atypical coupling of nomenclature involving an incorrect synonym for one strand and labels of DNA replication intermediates for the other strand Incidence: n_1 = 3/54; n_2 = 3/35; n_3 = 2 /31; n_4 = 2/44
v.	Difficulties understanding the mechanism of strand function in replication and transcription. Incidence: n_1 = 5/54; n_2 = 2/35; n_3 = 5/31; n_4 = 4/44

Clearly, the student responses above suggest that rote learning of terminology did take place.

7.2.3 Failure to differentiate

The interchange of DNA-strand labels (category iii difficulty) is accompanied by a failure to differentiate (Grayson, 2004) the strands on the basis of specific function. Several combinations of erroneous strand nomenclature were also recorded, some with or without reasons for the differentiation. For the sake of brevity, only selected examples which illustrate this category (iii) difficulty, and typical of the overall trend, are presented here. Twenty percent (34/164) of the total students ($n_{1\text{-}4}$) referred to labels A and B as "leading" and "lagging" strands, respectively. In support of their answers, some students provided the following reasons.

i. "A is the leading strand. Replication occurs in a $5\rightarrow3'$ direction within a replication bubble or fork. There is a problem with the polarity of B, resulting in the formation of Okazaki Fragments, thus B is the lagging strand".

ii. "A is the leading strength [strand] because nucleotides move from a $5\rightarrow3'$ direction. B is the lagging strength because nucleotides move from a $5\rightarrow3'$ direction".

iii. "A - leading strand. It begins from $5'\rightarrow3'$ left to right. B - lagging strand. It forms in the opposite direction to the leading strand and therefore it is from right to left in the $5\rightarrow3'$ direction".

The reference to "leading strand", "lagging strand" or "Okazaki Fragments" clearly demonstrates a substitution (Salomon and Perkins, 1989; Grayson, 1996) of DNA-strand labels with nomenclature associated with DNA replication intermediates (3.5.8; Kornberg and Baker, 1992). Examples of other terminology, associated with DNA replication, were erroneously coupled as follows:

i. "A - parent strand. It is used in replication.
B - newly synthesized daughter strand".

ii. "A is the template strand. Strand that is replicated.
B is the complementary strand because it is newly synthesized".

7.2.4 Atypical coupling of nomenclature

An atypical coupling of nomenclature involving an incorrect synonym for one strand and labels of DNA replication intermediates for the other strand (category iv difficulty) was also recorded; however, the incidence of the difficulty was low (Table 7.1). No reasons were provided for the differentiation given in the example below.

i. "A is the minus (-) strand or the template.
B is the primer strand".

The synonym "minus (-) strand" is generally associated with the nomenclature of viral DNA templates (Kornberg and Baker, 1992). It is apparent that students memorized terminology and used them out of context, clearly demonstrating erroneous information transfer (Salomon and Perkins, 1989; Grayson, 1996) and a lack of understanding of specific strand function. The precise reasons contributing to the above is speculative. The word "replication" in part (2) of the probe (3.5.7) could be viewed as the presumptive cue, influencing the misuse of terminology associated with DNA replication intermediates. "Word association" is generally regarded as a cue in concept mapping

(White and Gunstone, 1992; Van Zele *et al*., 2004). However, the current difficulty may be exacerbated by the presentation of a wide range of potentially confusing alternate nomenclature for DNA strands. Synonymous with "coding strand" are alternate nomenclature such as the "Watson strand" (Garrett and Grisham, 1995), "anti-sense strand" (Garrett and Grisham, 1995), "non template strand" (McKee and McKee, 1996; Mathew *et al*., 1997; Weaver, 1999) and "+ strand" (Kornberg and Baker, 1992; McKee and McKee, 1996). Similarly, the template strand may also be labeled as the "sense strand" (Garrett and Grisham, 1995), "copy strand" (Armstrong, 1989) and the "- strand" (Kornberg and Baker, 1992; Garrett and Grisham, 1995). The nomenclature, provided by Kornberg and Baker (1992), namely "+" and "–" strands, refers to the double-stranded RF (replicative form) ϕX174 viral DNA templates. Conflicting DNA-strand nomenclature can also be found in the literature. For example, Lewin (1994) and Turner *et al*. (1997) referred to the template strand as the anti-sense strand, contrary to the common and accepted opinion that the sense strand is transcribed to mRNA (Rao, 1996; Scism, 1996). Such conflicting nomenclature which accompanies symbolic DNA representations may also contribute to learning difficulties.

7.2.5 Poor understanding of mechanisms

The response to part (2) of the probe (3.5.7) demonstrated a poor understanding of the basic mechanisms which relate to replication and transcription (category v difficulty). The following examples of student quotes warrant a closer introspection into student difficulties.

i. "A -leading strand
B - lagging strand
Replication - it is synthesized from strand A, therefore, this means A replicates to form B. Transcription - strand B has been transcribed from information from A, therefore B is the correspondence of A nucleotides."

ii. "A is the template strand and B is the complementary strand.
Replication - complementary and template strands because cDNA is being copied on the template. Leading and lagging strands are involved in transcription".

iii. "A is the coding strand and B the template strand.
The template strand is copied during replication. The template strand binds a new strand to form a replication strand. The coding strand transcripts a new strand".

iv. "A - template strand
B - complementary strand."
Both strands are implicated in replication because A is being copied in $5\rightarrow3'$ by B to an exact copy of A. The information on A is transcribed to B strand".

v. "A - template strand
B - coding strand
So, in coding strand replication occurs $3\rightarrow5'$ and in the template strand replication occurs in the $5\rightarrow3'$ direction. The coding strand is used in replication because it allows primers to bind to it and this allows synthesis of new DNA strand".

When teaching DNA replication, it is customary to refer to both DNA strands as templates for semi-conservative DNA replication (Meselson and Stahl, 1958; Table 3.6). Indeed, this can be very confusing to students who are initially introduced to the template and coding strands as the two constituent strands of DNA. As can be seen in the students' responses above, definitions are learnt without a clear understanding of strand function which invariably influences the nomenclature. Teaching a complex topic such as replicative DNA synthesis can be challenging itself, let alone contributing to further confusion in the naming of strands. Replication intermediates can be characterized by a

range of labels such as "primer strand", "leading strand", "lagging strand", "nascent strand", "Okazaki Fragment", "parent strand" or "daughter strand" (Kornberg and Baker, 1992). It is evident that students memorize these names and use them out of context, especially when the mechanism of replication or transcription is poorly understood (Gupthar and Anderson, 2003). In a subsequent study, reported by colleagues at the University of KwaZulu-Natal, Pietermaritzburg Campus, SA, undergraduate students demonstrated similar difficulties with replication and transcription (Fossey and Hancock, 2005).

In response to part 3 of the free-response probe (3.5.7), a small number of students (n_1 = 3/54; n_2 = 3/35; n_3 = 2 /31; n_4 = 2/44) perceived DNA templates to be immobile or inert during replication simply because “they are large structures” which are scanned by enzyme complexes that are “nimble” or “considerably smaller and therefore mobile”. This perception might reflect a dynamic process of visualization in the students' mind (Hegarty, 2004; Bodemer *et al*., 2004; Lowe, 2004; Ploetzner and Lowe, 2004) that the smaller component is indeed mobile. Unfortunately, many external representations or diagrams of replisome-enzyme complexes (Campbell and Smith, 1993; Stryer, 1995; Campbell, 1995; Horton *et al*., 1996; Turner *et al*., 1997) present a stage-specific feature of DNA replication which can be interpreted incorrectly. The students' attention is usually drawn to an arrow at the forefront of a replisome complex on a nascent strand of DNA (Campbell and Smith, 1993, p.95; Stryer, 1995, p.809; Campbell, 1995, p.611). Presumably, this gives the deceptive impression that the enzyme complex is mobile as nascent DNA is synthesized 5 $\rightarrow$ 3' (Campbell and Smith, 1993; Stryer, 1995; Campbell, 1995; Horton *et al*., 1996; Turner *et al*., 1997). In contrast, however, there is sufficient evidence in bacterial cells that DNA polymerase is actually anchored in place as the DNA template moves through the enzyme (Lemon and Grossman, 1998; Losick and Shapiro, 1998). Lemon and Grossman visualized DNA polymerase in living cells of *Bacillus subtilis* by the creation of a fusion protein comprising the catalytic subunit Pol C and green fluorescent protein (GFP).The Pol C-GFP complex was localized at discrete midcell positions rather than being distributed randomly, supporting an earlier model that the DNA template moves through the polymerase (3.5.7). Often, students may not consult

lecture notes or the relevant textual information which accompanies a diagram and hence develop a mental image or internal representation that is untrue (Salomon and Perkins, 1989) or superficial (Chi *et al*., 1981; Lowe, 1993; Chapman, 2001; Schönborn *et al*., 2002a).

7.3 Conceptual understanding of DNA-strand nomenclature and function using student-generated diagrams

The focused probe (3.5.7) was used to gain further information on the students' conceptual understanding and consolidation of the facts concerning DNA-strand nomenclature and function. The probe demanded a clear differentiation of DNA-strand function in the processes of transcription and replication. Diagrams or external representations, generated by students, afforded an opportunity to assess individual mental models (Stewart, 1985; Carey, 1986) or idiosyncratic depictions of these biological processes. Invariably, these may be influenced by prior knowledge or learning (Kindfield, 1994) but can be expressed incorrectly as explained by the theory of constructivism (Driver and Bell, 1986; Von Glasersfeld, 1992). Diagrams or external representations therefore provide useful information which can be used to categorize the nature of various difficulties with these concepts.

The focused probe (3.5.7) generated over 300 student diagrams dealing with replication and transcription. The categories of difficulties (i-v) were identical to those which emerged when using the free-response probe (Table 7.1); however, the incidence of difficulties fitting categories (i-iv) differed slightly. The focused probe produced a higher incidence of the category (v) difficulty. Overall, the following incidences were recorded, evidently showing that a number of students experienced more than one type of difficulty:

Category (i): Incidence: n_1 = 22/54; n_2 = 4/35; n_3 = 16 /31; n_4 = 15/44

Category (ii): Incidence: n_1 = 4/54 ; n_2 = 2/35; n_3 = 3/31 ; n_4 = 2/44

Category (iii): Incidence: n_1 = 13 /54 ; n_2 = 4/35; n_3 = 6 /31 ; n_4 = 10/44

Category (iv): Incidence: $n_1 = 3/54$; $n_2 = 3/35$; $n_3 = 2/31$; $n_4 = 2/44$

Category (iv): Incidence: $n_1 = 16/54$; $n_2 = 11/35$; $n_3 = 14/31$; $n_4 = 24/44$

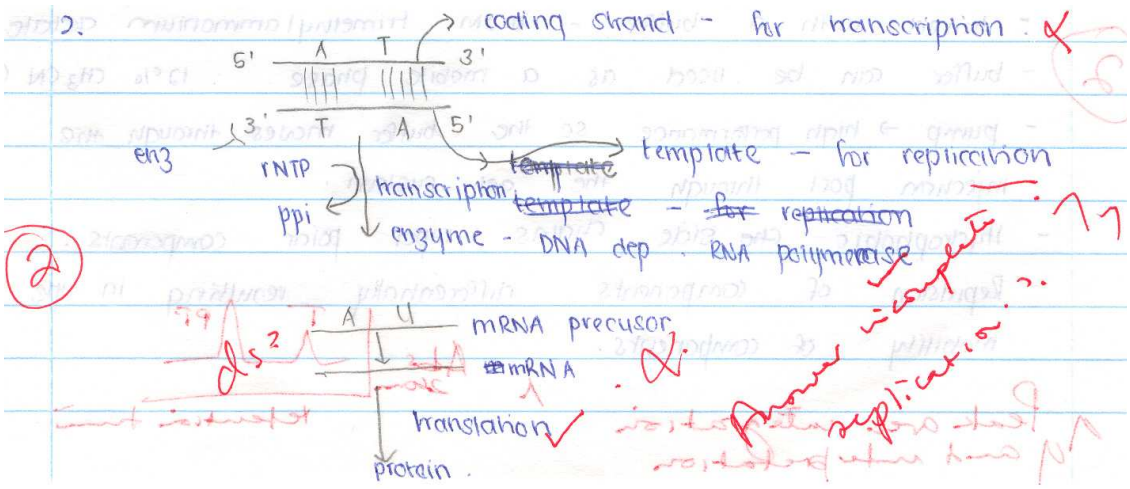
To avoid repetition of similar student responses (7.2), only selected examples highlighting the different categories of difficulties are presented here. These were extracted from annotated diagrams (Figures 7.1 and 7.2) that students had presented in response to the focused probe, commenting on either transcription or replication or both processes.

Student S28 provides the correct labeling of the DNA strands in diagram (A) but links the coding strand with transcription and the template strand with replication, respectively (Figure 7.1). Contrary to the label, the template strand is shown to be read by DNA - dependent RNA polymerase from the 3′ end to synthesize an mRNA precursor which gives rise to mRNA. At first glance, the precursor mRNA appears to be linked to the mRNA but a poorly illustrated short arrow barely shows the separation. It is not clear from diagram A that both parental DNA strands could act as templates for DNA replication (Figure 7.1). Student S28 presents an incomplete answer demonstrating the category (v) difficulty.

Student S29 presents with difficulties of both categories (i) and (v). Diagram B shows a failure to name the DNA strands except that "1 and 2" are provided without any explanation. Further, a false double-stranded mRNA transcript is shown instead of a single-stranded one (Figure 7.1 B). The erroneous insertion of thymine on one strand of the mRNA transcript is not known to exist in nature except in transfer RNA (tRNA) (Stryer, 1995). Student S29 provided no further comment on strand function and DNA replication.

Diagram C (Figure 7.1) shows student S30's attempt at providing multiple labels for each DNA strand. These included "template", "sense" and "+" for one of the strands and "coding strand", "-" and "antisense" for the other. Student S30 gives no reasons for the strand labels but indicates erroneously that transcription produces an mRNA transcript

A


coding strand - for transcription
5' A T 3'
3' T A 5'
template - for replication
enz
rNTP
ppi
transcription
template - for replication
enzyme - DNA dep. RNA polymerase
mRNA precusor
mRNA
translation
protein

B

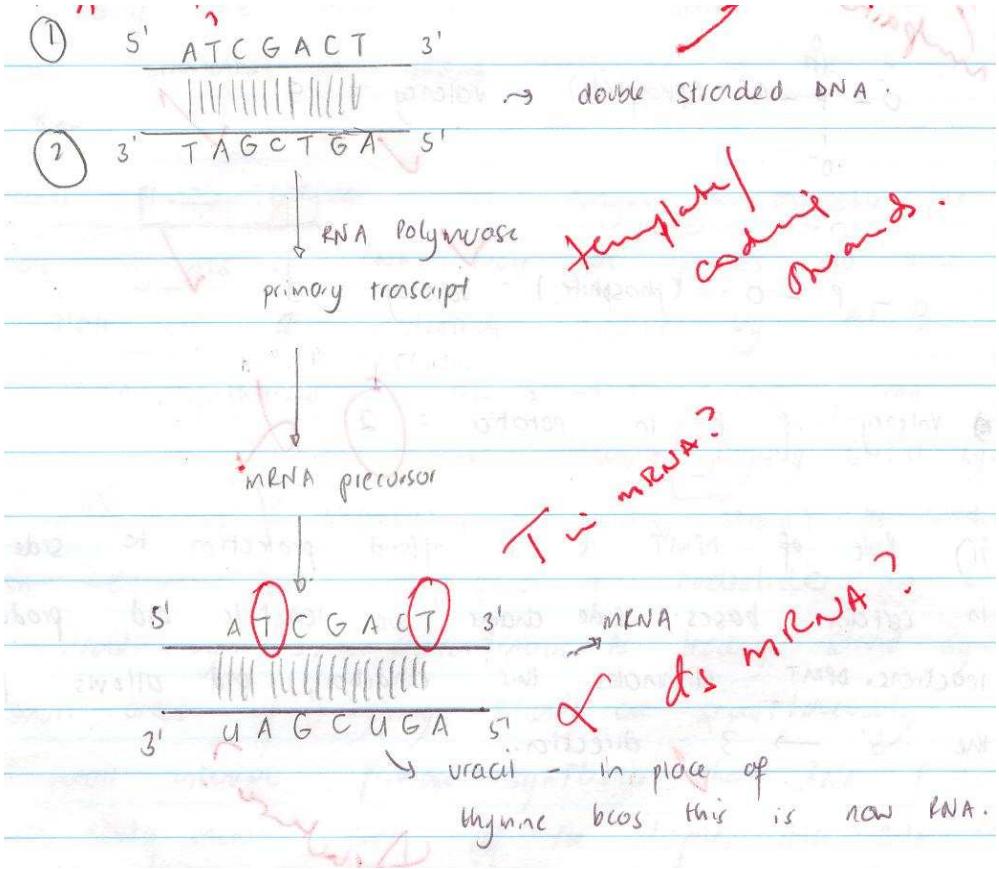
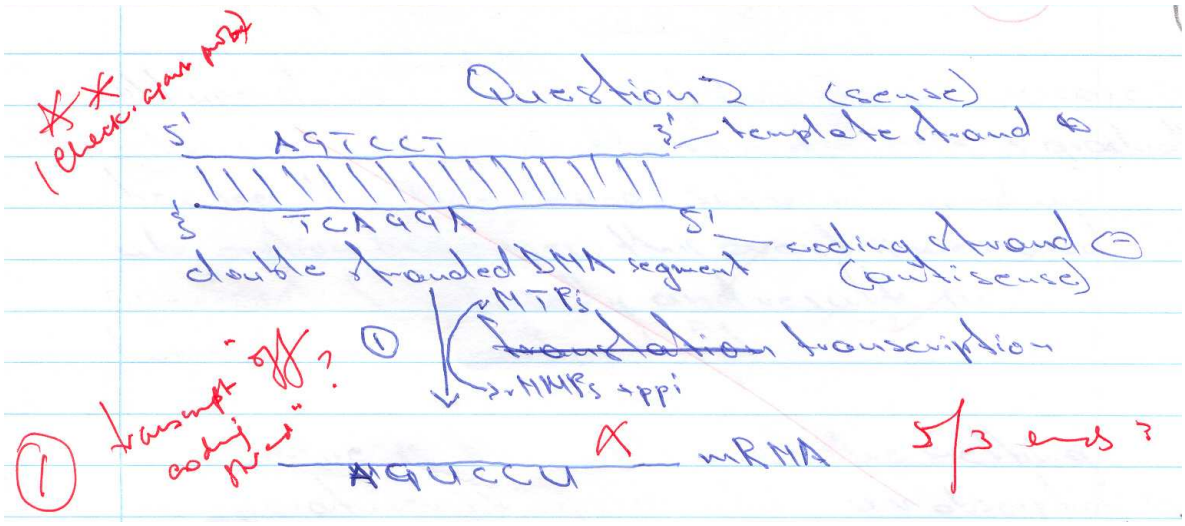

1 5' ATCGACT 3'
double stranded DNA.
2 3' TAGCTGA 5'
RNA Polymerase
primary transcript
mRNA precursor
5' ATCGACT 3'
mRNA
3' UAGCUGA 5'
uracil - in place of thymine bcos this is now RNA.

Figure 7.1 continued

C



D

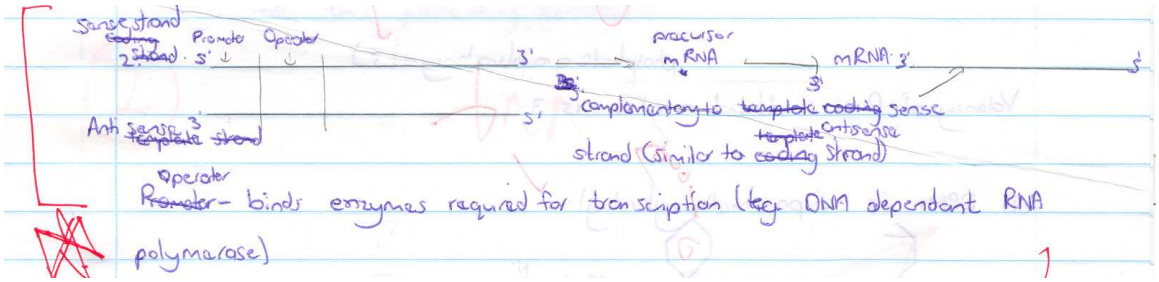


Figure 7.1 Students' illustration of DNA-strand function with exclusive reference to transcription.

complementary in sequence to the coding strand. No comment was given on DNA-strand function and replication. Student S30 shows difficulties which fit categories (ii) and (v), respectively (Table 7.1).

Diagram D (Figure 7.1) illustrates confusion in the mind of student S31. The left hand side of the diagram shows deletions of the "coding strand" and "template strand" labels and erroneous substitutions (Salomon and Perkins, 1989; Grayson, 1996) with "sense" and "antisense" labels, respectively. The student indicates that mRNA is complimentary to the "template" strand which is deleted and replaced by "coding" strand. It appears subsequently that "coding" strand is also deleted and substituted with "sense" strand. S31 indicates in parenthesis that the sense strand is similar to the antisense strand and this conclusion arises from deletions of "coding" and "template" in the same line. Student S31 shows difficulties which fit categories (ii) and (v), respectively. No comment was given regarding DNA replication. S31 also indicates erroneously, and with yet another deletion which suggests uncertainty, that the operator binds DNA-dependent RNA polymerase to facilitate transcription. This is not true as the operator simply influences the rate of mRNA formation once DNA-dependent RNA polymerase binds to the promoter in order to facilitate transcription of a prokaryotic gene (3.5.4; Gardner *et al*., 1984; Weaver, 1999).

Among the incomplete answers were rather peculiar depictions of the semi-conservative mode of DNA replication (Meselson and Stahl, 1958), illustrated by a few students (Incidence: n_1 = 1/54; n_2 = 2/35; n_3 = 2/31; n_4 = 2/44) who labelled both DNA strands as "parental DNA" or "parent strands A and B". In terms of the mechanism of semi-conservative replication, it is reasonable to show the parental strand (A and B) "template" distribution as follows; AB → An + Bn, where n = the nascent cDNA strand. Instead, depictions showing the following patterns of strand distribution were recorded (Figure 7.2);

i. AB → AB (Diagram A does not differentiate template B from nascent DNA).

ii. AB → AB + AB (Diagram B presents a duplication of parental DNA templates which are distributed as "daughter molecules", following replication).

iii. AB → AB + A[B] → n (Diagram C indicates that a daughter cell simply receives copies of the parental DNA; however, strand B in one of two progeny cells gives rise to nascent DNA (n). The distribution of n was not explained.

iv. AB → An + An or AB → Bn + Bn (Diagram D indicates the distribution of one template in combination with the nascent cDNA).

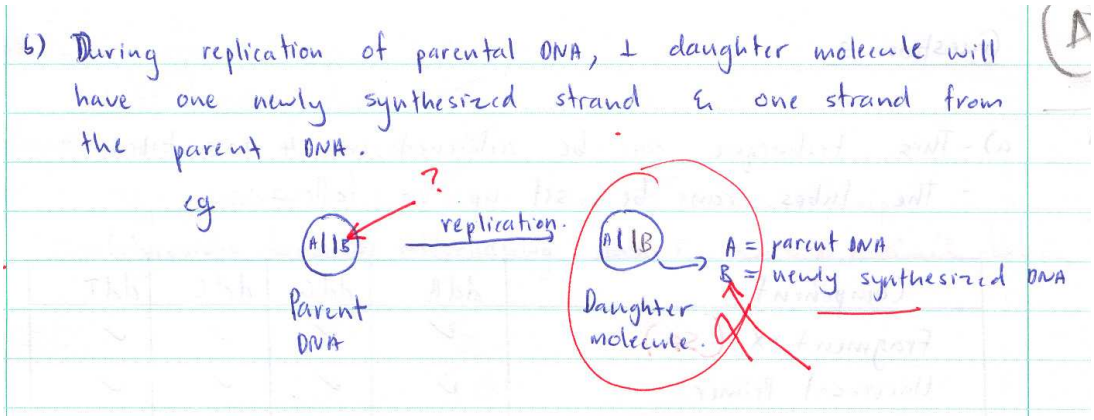
v. AB → Ax + Bx (Diagram E does not define "x" and fails to differentiate DNA template distribution from parent and progeny cells where A and B are regarded as cells and not template strands of DNA).

The students, who generated diagrams A-E (Figure 7.2), provided no comment on transcription. Interviews reveal further that DNA-strand labelling and function are poorly understood by certain students and there appears to be difficulty explaining specific DNA-strand function in different processes (Table 3.6).

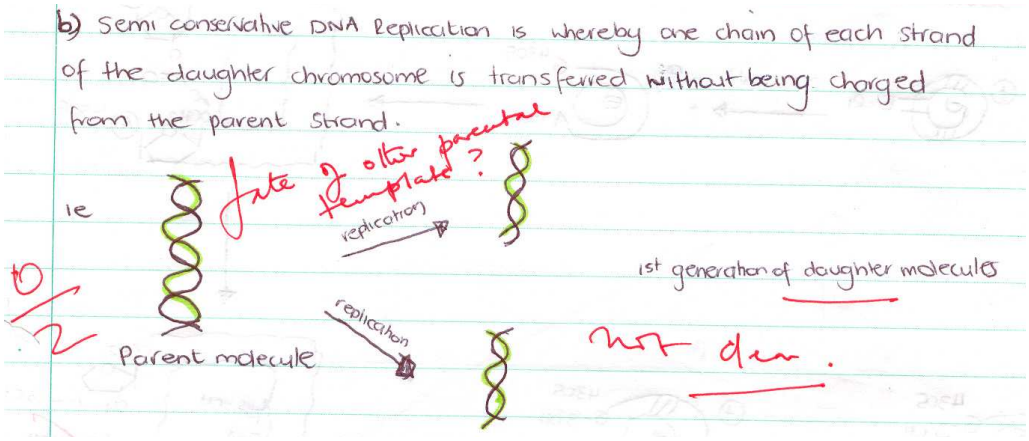
The following interview, about a student-generated diagram, reveals a category (iii) (Table 7.1.) difficulty, characterized by the erroneous substitution of DNA strand labels with one usually associated with DNA replication intermediates.

I: Let us take a look at your labels for double-stranded DNA [pause]. I see that you have labelled the DNA strands [pause] template and primer, not so?
S32: Yes, Sir.
I: Could you tell me why you used these labels?

A



B



C

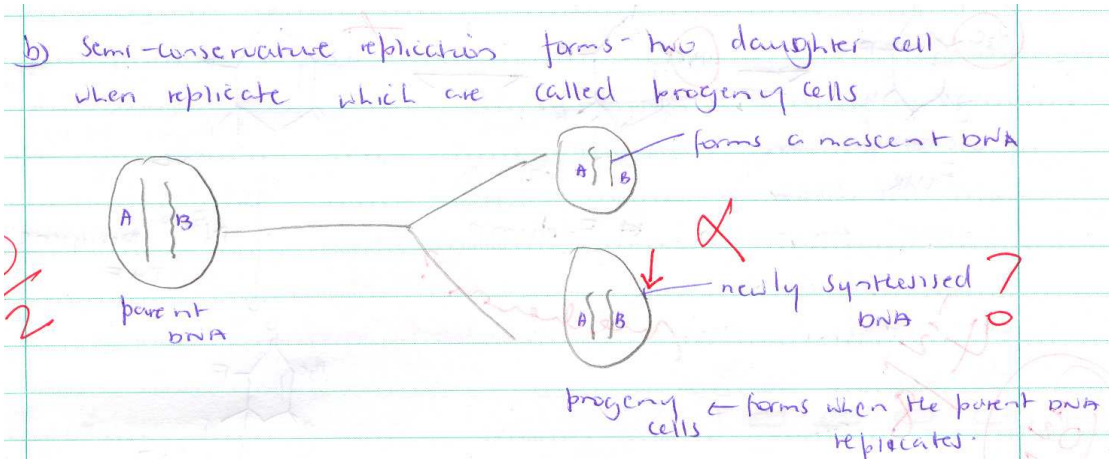
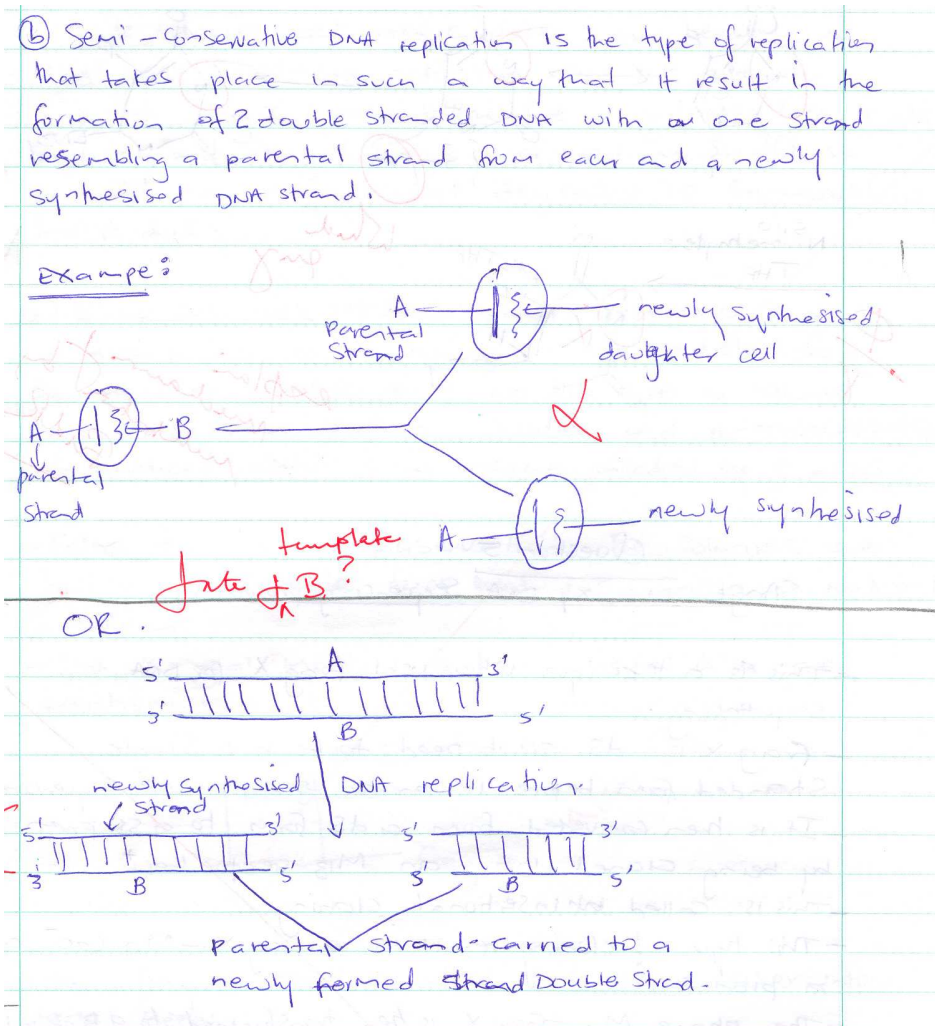


Figure 7.2 continued overleaf

D



E

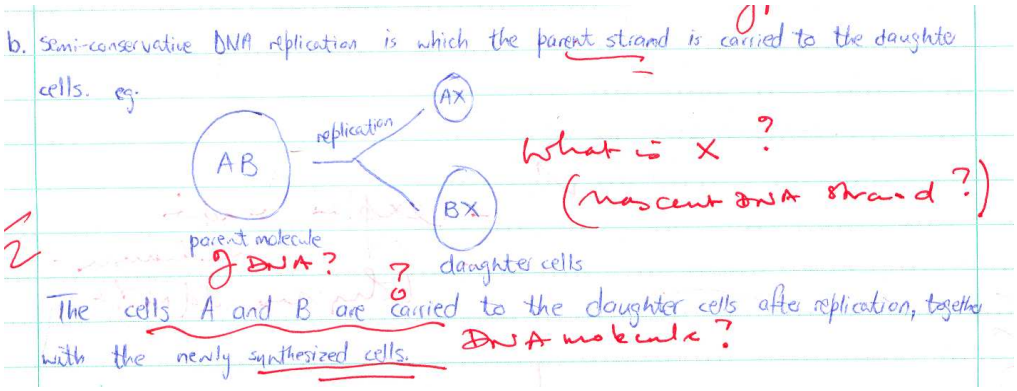


Figure 7.2 Students' illustration of DNA-strand function and replication.

S32: Well, you see the primer is needed for replication. I know the template strand gives mRNA during transcription.
I: Do you not think that a template strand is needed to give you complementary DNA or newly synthesized cDNA?
S32: No. I think the primer starts the DNA replication process.
I: Have you come across the coding strand?
S32: Eh [pause] maybe, no I don't think so.
I: Okay. Let's stop here.

"Primers", typically short RNA fragments, are required for the initiation of continuous or discontinuous DNA synthesis (3.5.8; Table 3.7). Student S32 correctly associates the term "primer" with the start of DNA replication but fails to indicate that nascent cDNA bears the complementary base sequence to a "template" strand of DNA. The association of a DNA "template" strand that gives rise to mRNA during transcription is acceptable.

An interview with student S33 indicates that meaningful learning from diagrams (Lowe, 1988; Kindfield, 1994; Henderson, 1999) did not take place. This is evident from the following interview excerpt.

I: Your written attempt at explaining DNA strand function stops at replication. Here I notice that you actually show [pause]..eh.. that each parental strand is distributed separately to progeny cells together with the nascent DNA [I refers to Figure 7.2, diagram D]. What about strand functions and transcription?
S33: I thought about it but I wasn't sure which parent strand to show.
I: What do you mean by this?
S33: I was doubtful about the strand that will give mRNA during transcription.
I: You mean the parent strands A and B?
S33: Yes.
I: Tell me, besides the parent strand label, what other strand names can you give to double-stranded DNA?
S33: You have template and primer strands, also Okazaki Fragments.

I: What is the function of a template strand?

S33: I know it joins the primer which becomes extended with new DNA.

I: So could a parent strand of DNA be called a template?

S33: I found the label in a replication structure and not in the intact double-stranded DNA.

I: Okay. Let us stop here. Thank you.

Student S33 initially labels DNA strands as "parental strands" but fails to interpret that these are in fact template strands implicated in replicative DNA synthesis. There is acknowledgement; however, that "template" strand labels are found in DNA replication intermediates. Clearly, the student is unable to assign and interpret common factual information which relates to different diagrams of DNA (Barlex and Carre, 1985; Lowe, 1988; Henderson. 1999; Bodemer *et al*., 2004). A poorly established "knowledge domain" is known to impact negatively on a student's ability to interpret or produce diagrams which relate to a specific topic or process (Kindfield, 1994; Henderson. 1999; Bodemer *et al*., 2004; Cook *et al*., 2006).

7.4 Interpretation of schematic DNA replication diagrams

7.4.1 Students' multiple difficulties with the mechanism of DNA replication

Free-response probes 1 and 2 (3.5.8) yielded a wide range of student difficulties associated with DNA replication (Table 7.2). Students ($n_{1\text{-}4}$) chose to explain structural features of either the *E.coli* replication bubble model (3.5.8; Figure 3.6) or the ΦX174 RF I → RF II replication model (3.5.8; Figure 3.7). This exercise proved to be cognitively challenging to some students. Annotated diagrams, reflecting transitional features or sequential change to the structure of DNA, were presented but some of these reflected a poor understanding (Fossey and Hancock, 2005; Beltramini *et al*., 2006) of the role of enzymes and auxiliary proteins in the molecular process. The response was further exacerbated by inaccurate labelling of structural replication intermediates such as RNA

primer, nascent cDNA, Okazaki Fragment, template strand, leading and lagging strands, giving false interpretations of the mechanism of replicative DNA synthesis (Gupthar and Anderson, 2003). In response to free-response probe 1 (3.5.8), a student presented diagram A (Figure 7.3) illustrating a replication bubble where poor differentiation is made between the leading and lagging strands within the structure. In addition, diagram A (Figure 7.3) does not indicate that RNA primers are precursor fragments required for all nascent cDNA synthesis. Okazaki Fragments are erroneously labelled as a combination of an RNA primer and nascent cDNA when it clearly bears reference to the latter only (3.5.8; Table 3.7). Furthermore, the direction of DNA synthesis is not indicated within the replication eye or bubble. This diagram lacked information on the auxiliary proteins implicated in the unravelling the replication eye or enzymes facilitating RNA primer and cDNA synthesis. Reasons for the generation of "leading" and "lagging" strands were not given, neither was any comment provided on the concomitant unwinding of the replication fork and changing conformation of the bubble. Overall, diagram A (Figure 7.3) presents with a range of difficulties fitting categories (i), (iv), (vi), (ix) and (x) (Table 7.2)

Diagram B (Figure 7.3) was generated in response to free-response probe 2 (3.5.8). It gives a hint that nascent cDNA fragments may be synthesized from RNA primers although this is not illustrated with all nascent DNA. There is no indication of the direction or mechanism of DNA synthesis. Several proteins are associated with the diagram without any mention of function or full name but acronyms or abbreviations. These include "SSB", "dnaB", "Pol III" and "protein n" drawn alongside the "primosome". The diagram appeared bare, unaccompanied by notes or comment on the finer details concerning the role of enzymes, Mg^{2+}, dNTPs, ATP, type of supercoil and PPi release as nucleotides are incorporated in nascent DNA. It is apparent that students engage in superficial learning (Chi *et al*., 1981; Kozma and Russell, 1997 ; Schönborn *et al*., 2002a) and have a tendency to reflect on weak mental schemata (Thorndyke and Stasz, 1985) and hence provide answers that inadequately explain phenomena or processes (Barlex and Carre, 1985; Schönborn *et al*., 2002a; Cook *et al*., 2006; Beltramini *et al*., 2006; Rotbain *et al*., 2006). Also, the interpretation of molecular

Table 7.2 Categories of student difficulties with DNA replication and its representation using schematic models

Category	Nature of difficulty and incidence
i.	Failure to label strands 5 → 3' within replication intermediates **Incidence :** probe 1: $n_1 = 10/54$; $n_2 = 4/35$; $n_3 = 5/31$; $n_4 = 9/44$ probe 2: $n_1 = 11/54$; $n_2 = 2/35$; $n_3 = 4/31$; $n_4 = 8/44$
ii.	Confusion arising out of (1) in the direction of synthesis **Incidence :** probe 1: $n_1 = 10/54$; $n_2 = 4/35$; $n_3 = 5/31$; $n_4 = 9/44$ probe 2 : $n_1 = 11/54$; $n_2 = 2/35$; $n_3 = 4/31$; $n_4 = 8/44$
iii.	Poor differentiation of gaps and nicks **Incidence :** probe 1: $n_1 = 15/54$; $n_2 = 5/35$; $n_3 = 8/31$; $n_4 = 9/44$ probe 2: $n_1 = 12/54$; $n_2 = 6/35$; $n_3 = 7/31$; $n_4 = 8/44$
iv.	Failure to differentiate DNA strand terminology; template, primer, nascent DNA, leading and lagging strands and Okazaki Fragment **Incidence :** probe 1: $n_1 = 13/54$; $n_2 = 6/35$; $n_3 = 8/31$; $n_4 = 14/44$ probe 2: $n_1 = 15/54$; $n_2 = 5/35$; $n_3 = 10/31$; $n_4 = 12/44$
v.	Perceptions about DNA size and immobile templates **Incidence :** probe 1: $n_1 = 3/54$; $n_2 = 3/35$; $n_3 = 3/31$; $n_4 = 2/44$ probe 2: $n_1 = 3/54$; $n_2 = 3/35$; $n_3 = 2/31$; $n_4 = 2/4$

Table 7.2 continued

vi Failure to explain the semi-conservative mode of replication

Incidence : probe 1: $n_1 = 1/54$; $n_2 = 2/35$; $n_3 = 2/31$; $n_4 = 3/44$

probe 2: $n_1 = 2/54$; $n_2 = 2/35$; $n_3 = 2/31$; $n_4 = 2/44$

vii. Lack of comment on finer details concerning role of Mg^{2+}, dNTPs, ATP, type of supercoil and PPi release as nucleotides are incorporated in nascent DNA

Incidence : probe 1: $n_1 = 20/54$; $n_2 = 12/35$; $n_3 = 15/31$; $n_4 = 17/44$

probe 2: $n_1 = 25/54$; $n_2 = 12/35$; $n_3 = 12/31$; $n_4 = 14/44$

viii. Failure to integrate mechanisms with line drawings where needed

Incidence : probe 1: $n_1 = 10/54$; $n_2 = 3/35$; $n_3 = 8/31$; $n_4 = 11/44$

probe 2: $n_1 = 10/54$; $n_2 = 2/35$; $n_3 = 6/31$; $n_4 = 12/44$

ix. Poor knowledge of enzyme function

Incidence : probe 1: $n_1 = 10/54$; $n_2 = 4/35$; $n_3 = 5/31$; $n_4 = 14/44$

probe 2: $n_1 = 9/54$; $n_2 = 5/35$; $n_3 = 7/31$; $n_4 = 12/44$

x. Poor illustration of the transition of the structural form of replication intermediates

Incidence : probe 1: $n_1 = 16/54$; $n_2 = 2/35$; $n_3 = 4/31$; $n_4 = 8/44$

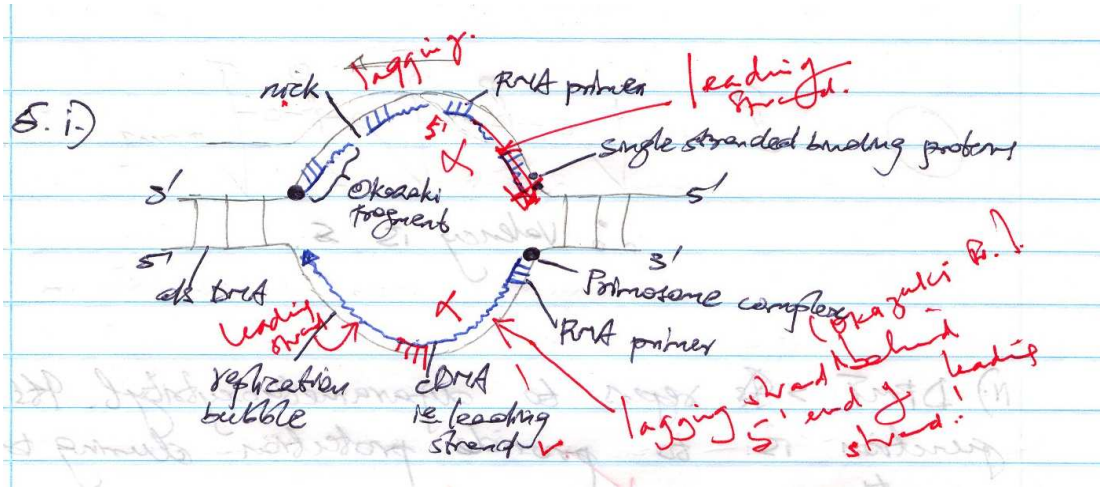
probe 2: $n_1 = 15/54$; $n_2 = 5/35$; $n_3 = 5/31$; $n_4 = 11/44$

xi. Use of acronyms and abbreviations without providing additional information or meaning of these

Incidence : probe 1: $n_1 = 5/54$; $n_2 = 0/35$; $n_3 = 5/31$; $n_4 = 6/44$

probe 2: $n_1 = 4/54$; $n_2 = 2/35$; $n_3 = 5/31$; $n_4 = 7/44$

A



B

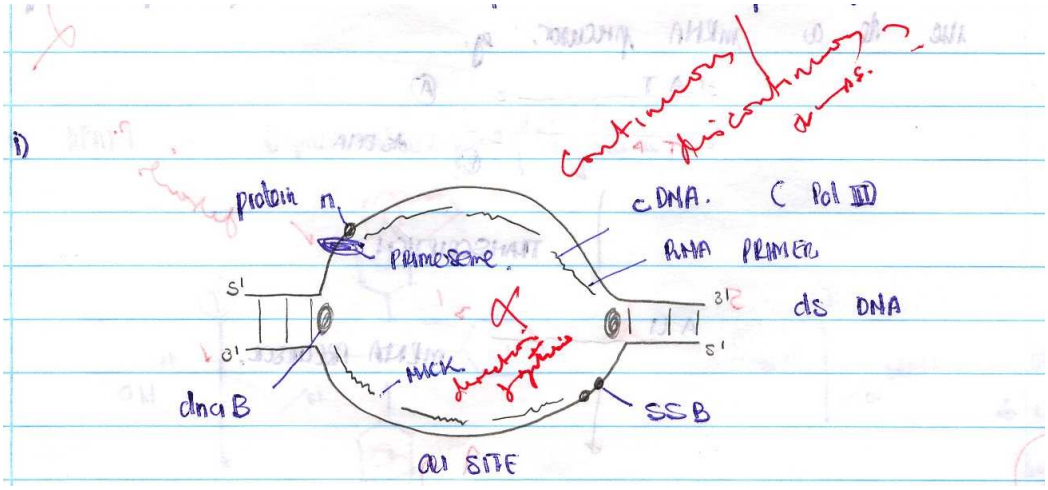


Figure 7.3 Students' diagrams illustrating DNA replication within a transition bubble feature.

mechanisms from a diagram or symbolic representation could be cognitively demanding (Scaife and Rogers, 1996; Kirschner, 2002; Treagust *et al*., 2003; Beltramini *et al*., 2006). Diagram B displayed difficulties fitting categories (i), (iv), (vi), (vii), (ix) and (xi) (Table 7.2).

Focused probes 1, 2 and 3 (3.5.8) yielded the same range of difficulties shown in Table 7.2; however, the incidences differed (Table 7.3). Notably, category (i) and (ii) difficulties, concerning 5→3′strand labelling and the direction of DNA synthesis, were eliminated in response to probes 2 and 3. Presumably, this tendency emerged as the diagrams forming part of these probes were appropriately labelled (3.5.8). In general, a marginal decrease in the incidence of difficulties was recorded when using the three focused probes. This might be attributed to an on-going, metacognitive learning process (Guterman, 2003; Georghiades, 2004); however, various student-generated diagrams (SGDs) and accompanying short notes confirmed the nature of difficulties first recorded when using the free-response probes (7.4.1). The depiction of inaccurate replication intermediates can be seen in randomly selected diagrams A-G (Figure 7.4). Diagram A of ΦX 174 RF DNA was reproduced by a student in response to focused probe 2, erroneously labelling the inner and outer parental strand templates as "+" and "-", respectively. Continuous cDNA synthesis begins at the ori site base position G ^{4305} (3.5.8) and not the outer strand template as shown. Further, the diagram fails to differentiate RNA primer and nascent cDNA strands as constituents of the discontinuous strand. Instead, both are collectively labelled as the Okazaki Fragment which is in fact the nascent cDNA strand. The definitions and roles of SSB, gpA, Rep and primosome proteins were not given as was the case in many other SGDs. Overall, many SGDs including this diagram, were accompanied by poor detail on the mechanism of DNA synthesis, the role of enzymes and their requirements to enable synthesis.

Table 7.3 Categories of student difficulties with DNA replication models using focused probes

---

Category **Nature of difficulty and incidence**

i. Failure to label strands 5 → 3' within replication intermediates

Incidence : probe 1: $n_1 = 11/54$; $n_2 = 4/35$; $n_3 = 4/31$; $n_4 = 7/44$

probe 2: $n_1 = 0/54$; $n_2 = 0/35$; $n_3 = 0/31$; $n_4 = 0/44$

probe 3: $n_1 = 0/54$; $n_2 = 0/35$; $n_3 = 0/31$; $n_4 = 0/44$

ii. Confusion arising out of (1) in the direction of synthesis

Incidence : probe 1: $n_1 = 5/54$; $n_2 = 2/35$; $n_3 = 2/31$; $n_4 = 3/44$

probe 2: $n_1 = 0/54$; $n_2 = 0/35$; $n_3 = 0/31$; $n_4 = 0/44$

probe 3: $n_1 = 0/54$; $n_2 = 0/35$; $n_3 = 0/31$; $n_4 = 0/44$

iii. Poor differentiation of gaps and nicks

Incidence : probe 1: $n_1 = 11/54$; $n_2 = 3/35$; $n_3 = 3/31$; $n_4 = 2/44$

probe 2: $n_1 = 3/54$; $n_2 = 3/35$; $n_3 = 3/31$; $n_4 = 2/44$

probe 3: $n_1 = 2/54$; $n_2 = 2/35$; $n_3 = 2/31$; $n_4 = 2/44$

iv. Failure to differentiate DNA strand terminology;
template, primer, nascent DNA , leading and lagging strands
and Okazaki Fragment

Incidence : probe 1: $n_1 = 10/54$; $n_2 = 4/35$; $n_3 = 5/31$; $n_4 = 4/44$

probe 2 : $n_1 = 11/54$; $n_2 = 5/35$; $n_3 = 6/31$; $n_4 = 6/44$

probe 3: $n_1 = 10/54$; $n_2 = 2/35$; $n_3 = 4/31$; $n_4 = 5/44$

v. Perceptions about DNA size and immobile templates

Incidence : probe 1: $n_1 = 3/54$; $n_2 = 2/35$; $n_3 = 2/31$; $n_4 = 2/44$

probe 2: $n_1 = 3/54$; $n_2 = 2/35$; $n_3 = 2/31$; $n_4 = 2/44$

probe 3: $n_1 = 3/54$ $n_2 = 2/35$; $n_3 = 3/31$; $n_4 = 2/44$

Table 7.3 continued

vi Failure to explain the semi-conservative mode of replication

Incidence : probe 1: $n_1 = 1/54$; $n_2 = 2/35$; $n_3 = 1/31$; $n_4 = 3/44$

probe 2: $n_1 = 2/54$; $n_2 = 2/35$; $n_3 = 1/31$; $n_4 = 2/44$

probe 3: $n_1 = 2/54$; $n_2 = 2/35$; $n_3 = 1/31$; $n_4 = 2/44$

vii. Lack of comment on finer details concerning role of Mg^{2+}, dNTPs, ATP, type of supercoil and PPi release as nucleotides are incorporated in nascent DNA

Incidence : probe 1: $n_1 = 21/54$; $n_2 = 12/35$; $n_3 = 14/31$; $n_4 = 19/44$

probe 2: $n_1 = 25/54$; $n_2 = 12/35$; $n_3 = 11/31$; $n_4 = 16/44$

probe 3: $n_1 = 21/54$; $n_2 = 11/35$; $n_3 = 10/31$; $n_4 = 12/44$

viii. Failure to integrate mechanisms with line drawings where needed

Incidence : probe 1: $n_1 = 10/54$; $n_2 = 4/35$; $n_3 = 6/31$; $n_4 = 6/44$

probe 2: $n_1 = 5/54$; $n_2 = 2/35$; $n_3 = 1/31$; $n_4 = 2/44$

probe 3: $n_1 = 0/54$; $n_2 = 0/35$; $n_3 = 0/31$; $n_4 = 0/44$

ix. Poor knowledge of enzyme function

Incidence : probe 1: $n_1 = 11/54$; $n_2 = 4/35$; $n_3 = 5/31$; $n_4 = 12/44$

probe 2: $n_1 = 10/54$; $n_2 = 4/35$; $n_3 = 8/31$; $n_4 = 11/44$

probe 3: $n_1 = 11/54$; $n_2 = 4/35$; $n_3 = 7/31$; $n_4 = 10/44$

x. Poor illustration of the transition of the structural form of replication intermediates

Incidence : probe 1: $n_1 = 17/54$; $n_2 = 4/35$; $n_3 = 4/31$; $n_4 = 9/44$

probe 2: $n_1 = 15/54$; $n_2 = 5/35$; $n_3 = 5/31$; $n_4 = 11/44$

probe 3: $n_1 = 11/54$; $n_2 = 2/35$; $n_3 = 4/31$; $n_4 = 8/44$

xi. Use of acronyms and abbreviations without providing additional information or meaning of these

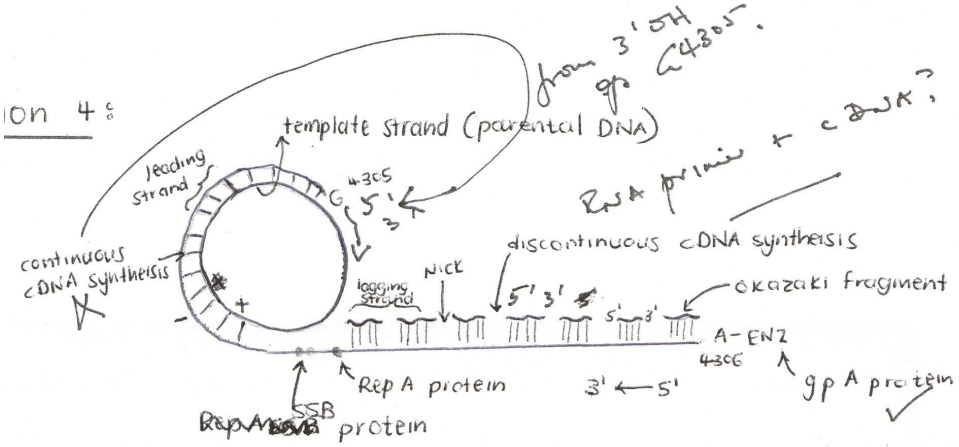
Table 7.3 continued

Incidence : probe 1: $n_1 = 5/54$; $n_2 = 0/35$; $n_3 = 5/31$; $n_4 = 6/44$

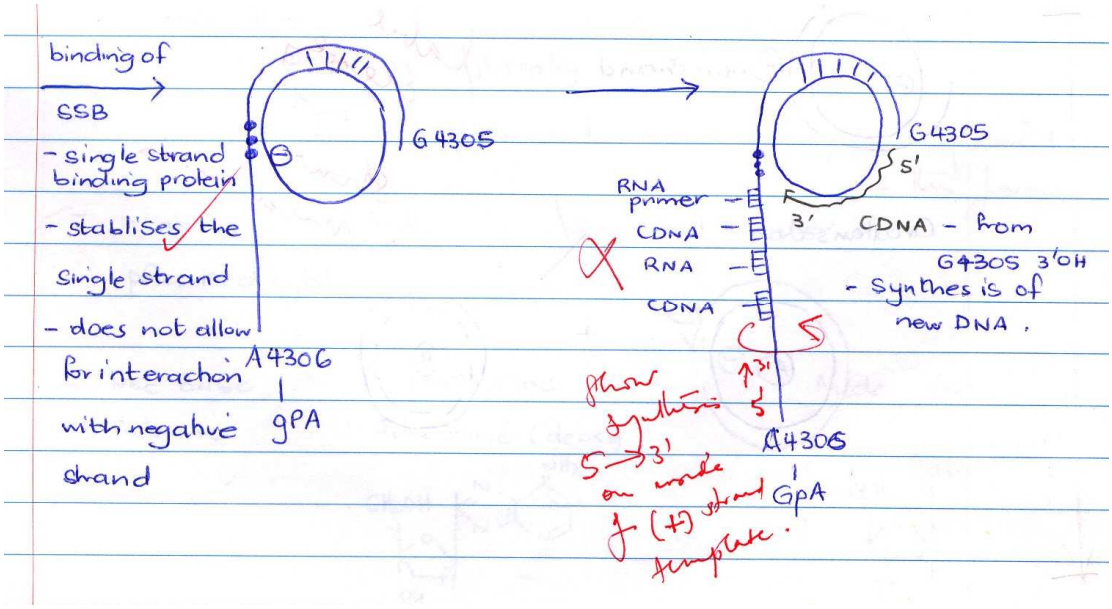
probe 2: $n_1 = 4/54$; $n_2 = 2/35$; $n_3 = 5/31$; $n_4 = 7/44$

probe 3: $n_1 = 5/54$; $n_2 = 2/35$; $n_3 = 4/31$; $n_4 = 8/44$

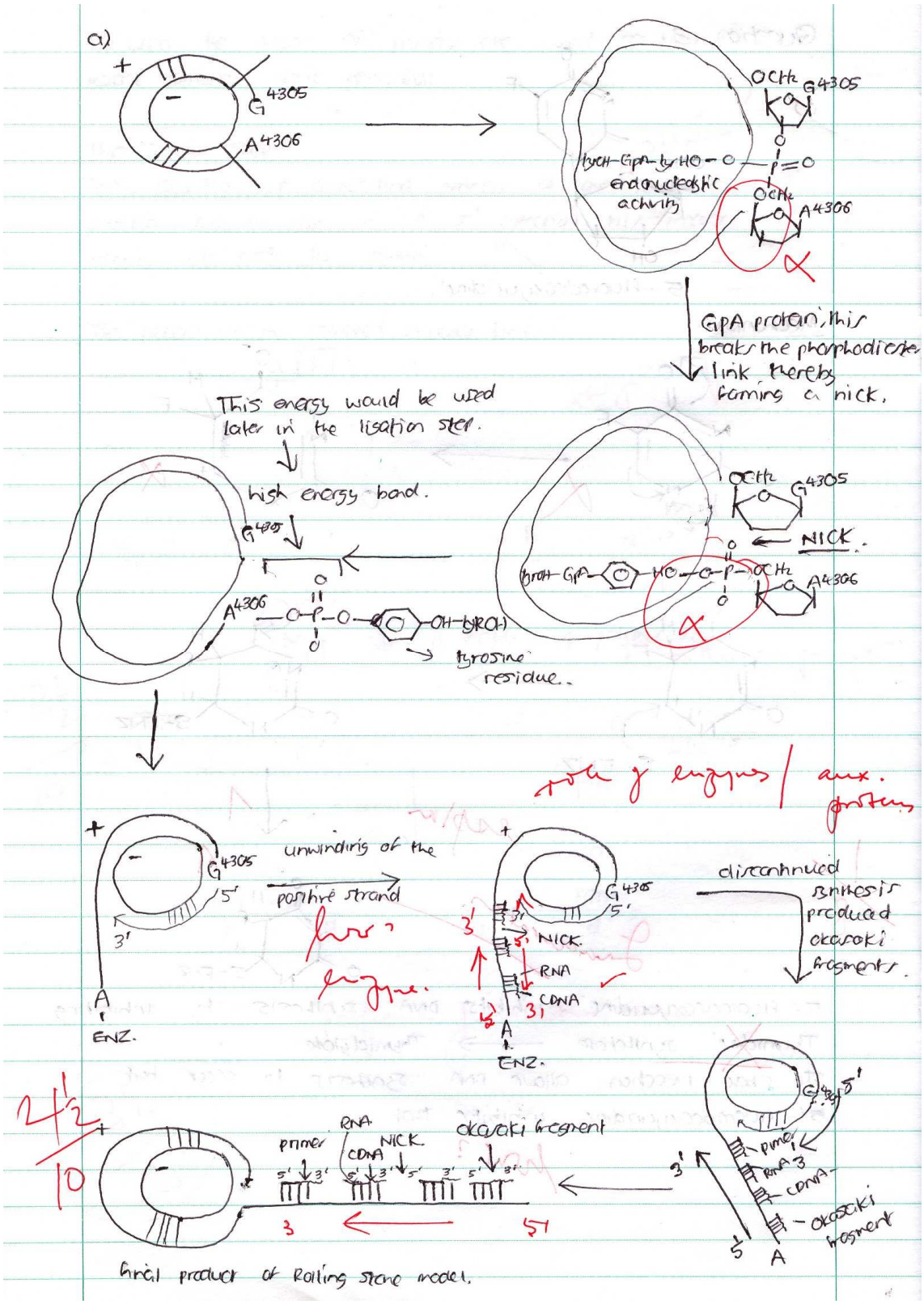
A


template strand (parental DNA)
leading strand
4305
5'
3'
continuous cDNA synthesis
lagging strand
NICK
discontinuous cDNA synthesis
okazaki fragment
A-ENZ
4306
Rep A protein
SSB protein
3' ← 5'
gp A protein

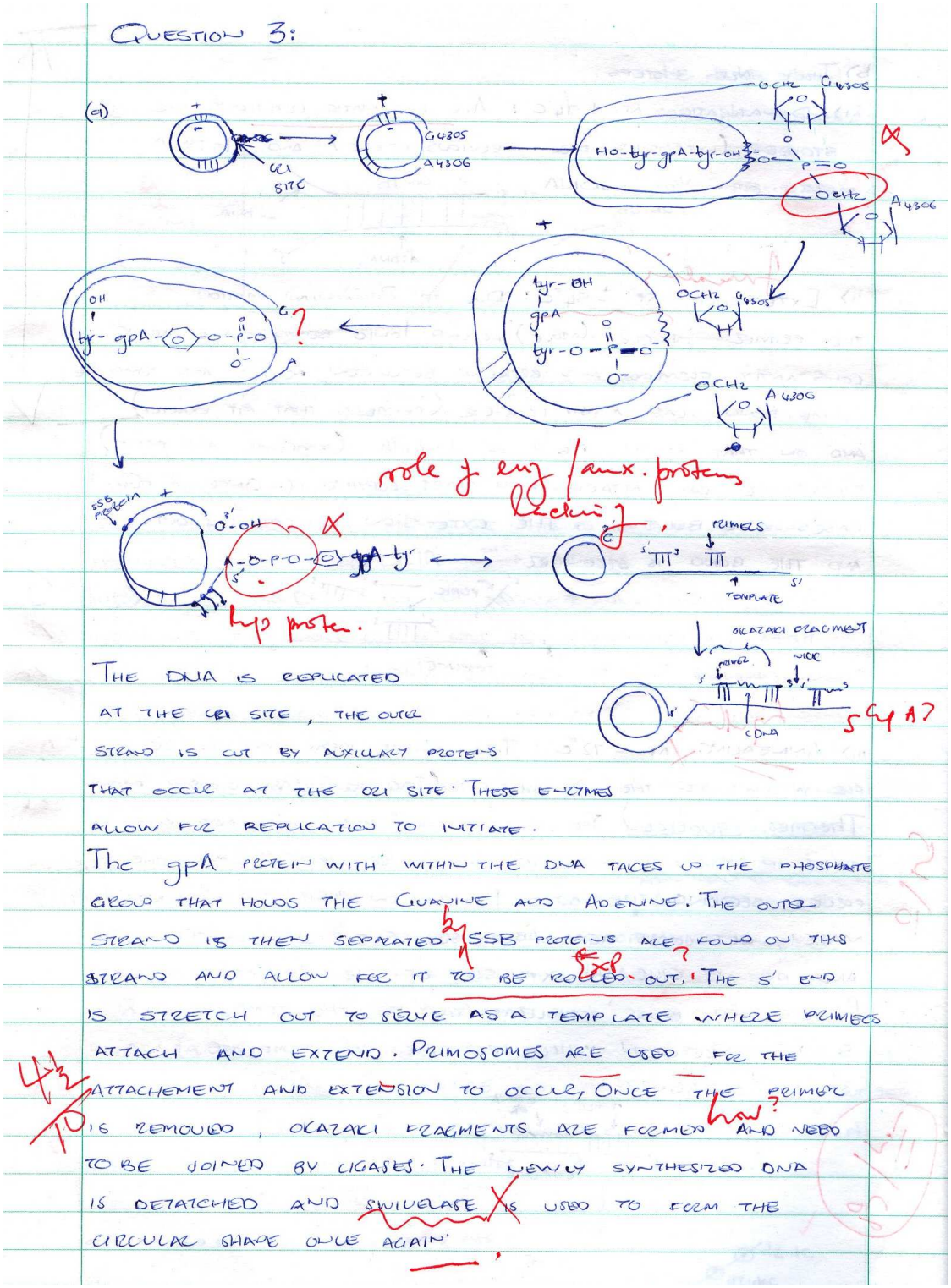
B


binding of
SSB
- single strand binding protein
- stablises the single strand
- does not allow for interaction with negative strand
A4306
gPA
G4305
RNA primer
CDNA
RNA
CDNA
3'
5'
CDNA - from G4305 3'OH
- Synthesis of new DNA.
A4306
GpA

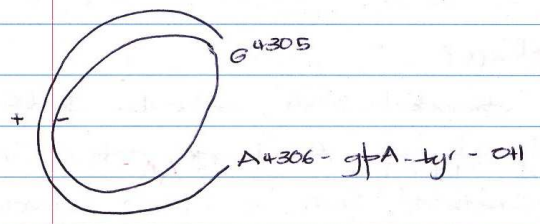
C


a)
G4305
A4306
endonucleolytic activity
GpA protein, this breaks the phosphodiester link thereby forming a nick.
This energy would be used later in the ligation step.
high energy bond.
NICK
tyrosine residue.
unwinding of the parental strand
ENZ.
discontinued synthesis produced okazaki fragments.
RNA
CDNA
primer
okazaki fragment
final product of Rolling Stone model.

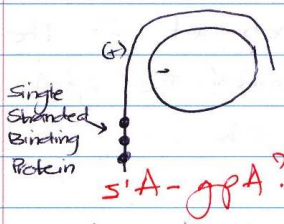
D


QUESTION 3:
(a)
ORI SITE
role of enz / aux. proteins
helicase
TEMPLATE
OKAZAKI FRAGMENT
RNA PRIMER
NICK
DNA
THE DNA IS REPLICATED AT THE ORI SITE, THE OUTER STRAND IS CUT BY AUXILLARY PROTEINS THAT OCCUR AT THE ORI SITE. THESE ENZYMES ALLOW FOR REPLICATION TO INITIATE.
The gpA PROTEIN WITH WITHIN THE DNA TAKES UP THE PHOSPHATE GROUP THAT HOLDS THE GUANINE AND ADENINE. THE OUTER STRAND IS THEN SEPARATED BY SSB PROTEINS ARE FOUND ON THIS STRAND AND ALLOW FOR IT TO BE ROLLED OUT. THE 5' END IS STRETCH OUT TO SERVE AS A TEMPLATE WHERE PRIMERS ATTACH AND EXTEND. PRIMOSOMES ARE USED FOR THE ATTACHEMENT AND EXTENSION TO OCCUR, ONCE THE PRIMER IS REMOVED, OKAZAKI FRAGMENTS ARE FORMED AND NEED TO BE JOINED BY LIGASES. THE NEWLY SYNTHESIZED DNA IS DETATCHED AND SWIVELASE IS USED TO FORM THE CIRCULAR SHAPE ONCE AGAIN.

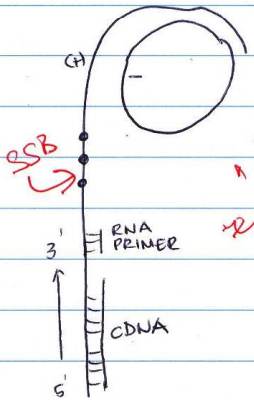
E



Once the ~~the~~ phosphodiester bond is broken, single stranded binding protein attaches to the (+) strand.

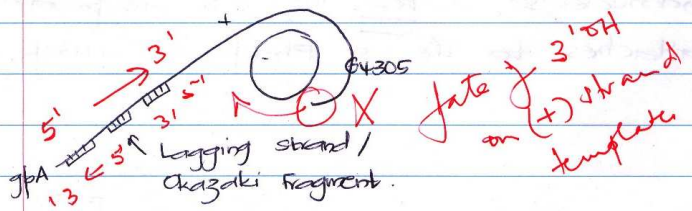


This protein stabilizes the (+) strand and prevents it from pairing with the ~~regn~~ (-) strand.

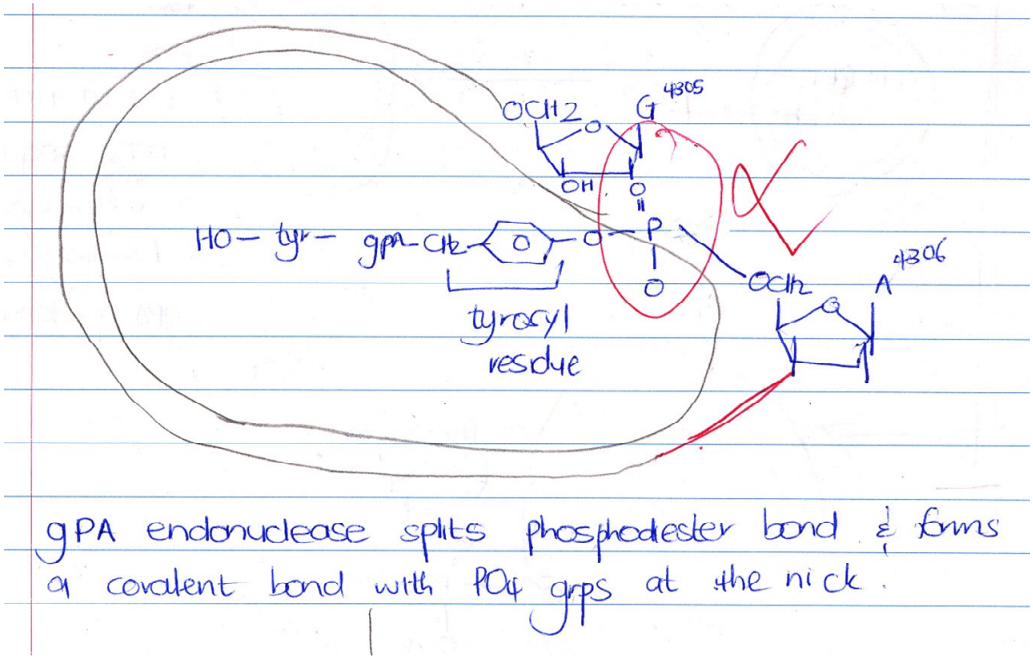


role of RepA?
- gene involvement?
- DNA primase
- DNA polymerase

With the addition of DNA dependant DNA polymerase and Mg^{2+}, nascent DNA can now be synthesized.



F



G

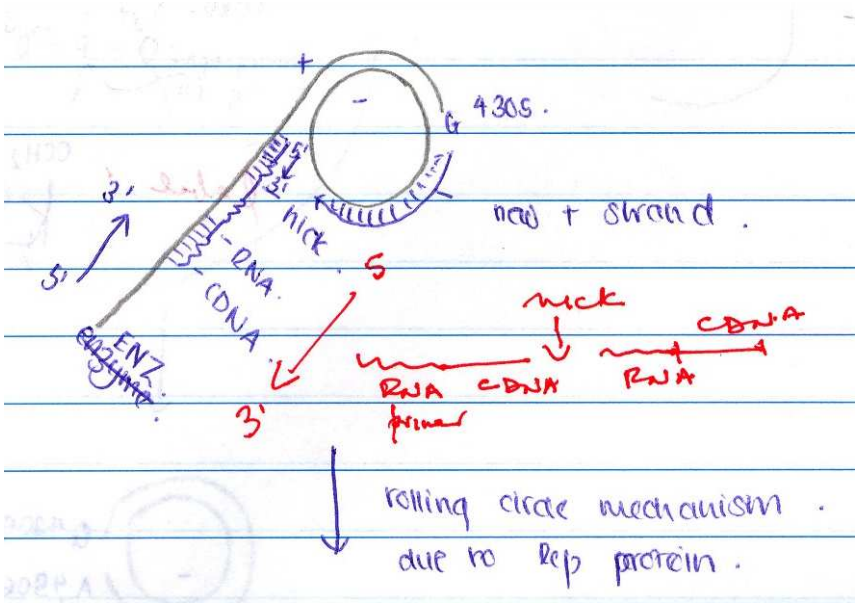


Figure 7.4 Students' illustration of replication intermediates of ΦX174 RF DNA.

Discontinuous strand synthesis was generally poorly explained and characterised by imprecise labelling of the RNA primer, nicks, nascent cDNA strand or Okazaki Fragment (Diagrams A, B, C, D, E and G; Figure 7.4). For example, diagram B shows no link between nascent cDNA fragment and RNA primer, rather these are shown incorrectly as alternate nucleic acid fragments on the reverse side of the "+" parental DNA template. Diagram C shows the placement of the RNA primer at the 5′ and 3′ ends of separate discontinuous, nascent DNA strands while diagram E shows a detachment of the RNA primer from nascent cDNA (Figure 7.4). There appeared to be difficulty integrating molecular mechanisms with line drawings. For example, this is evident in the endonucleolytic cleavage reaction (3.5.8; Table 3.7), facilitated by gpA protein in ΦX 174 RF DNA replication, where gpA attaches covalently via an active tyrosyl group to the 5′ phosphate moiety of an A residue at position 4306. Erroneous depictions of both endonucleolytic cleavage mechanism and covalent attachment of gpA protein can be seen in diagrams C, D and F (Figure 7.4). In contrast, diagrams A, B, E and G provide inadequate detail of this reaction (Figure 7.4). The replication bubble model (Figure 3.6) and its structural feature illustrated in focused probe 3 (3.5.8) provided lesser scope to integrate molecular mechanisms with the line drawings. Coinciding with this limitation, no students showed the category (viii) difficulty (Table 7.3) in response to focused probe 3.

Diagram C (Figure 7.4), generated in response to focused probe 1, presents transition features of ΦX 174 RF DNA, making up a flow diagram, finally labelled as "the Rolling Stone" model instead of "Rolling Circle Model" (3.5.8). The "Rolling Stones" are a famous UK rock group of the 1970s. The word "rolling" might be the presumptive cue for this type of erroneous association (White and Gunstone, 1992) where unrelated "everyday life contexts" can be supplanted into mental schemata (Nurrenbern, 2001; Kasanda *et al*., 2005). Flow diagrams were largely inaccurate, lacking essential detail (Pittman, 1999; Ploetzner and Lowe, 2004) on molecular mechanisms which invariably impact on the structural form of the DNA shown in each sequential step (Hancock and Fossey, 2005; Beltramini *et al*., 2006). It does appear that complex replication models, bearing fine molecular detail, can be challenging to learn, perhaps overwhelming

(Kirschner, 2002; Lowe, 2003), difficult to visualize (Hegarty, 2004; Mathewson, 2005; Cook *et al*., 2006) or to reproduce in diagrammatic form (Cho *et al*., 1985; Kindfield, 1994). While students could draw models, some failed to explain the semi-conservative mode of DNA replication (Table 7.3). Some students (Table 7.3) perceived DNA to be large, inert structures, acting as immobile templates for small enzymes which are mobile (7.2.5).

The schematic replication diagrams presented as Figure 3.5 (Chapter 3) facilitated the interactive "think aloud" exercise. Owing to the length of the interviews, only a single interview will be shown in this section. It is shown in two parts, allowing for a recess which was taken by both interviewer and interviewee. The chosen interview illustrates the typical difficulties experienced by many other students with DNA replication and the underlying molecular information conveyed by features of schematic diagrams. The author provides integrated comment to break through the tedious reading of this rather lengthy interview; however, the in-depth questioning was necessary and the feedback interesting to analyze.

Part 1:

I: Alright, you will notice on this page that I have three diagrams labeled A ,B and C. [Interviewer presents diagrams shown as Figure 3.5] Let's talk about them. What are they?

S34: These are diagrams showing the replication of DNA.

I: Excellent. Now [pause] if you say that these are associated with DNA replication, do they represent different systems?

S34: Yes. The first one is the circular DNA ,ehh[pause] with the replication of ϕx 174 and B and C[pause], eh, they show the replication bubble, ehh of E.coli DNA.

I: Tell me in simple terms, what do you understand by term, DNA replication ?

S34: DNA replication is the [pause] duplication of the DNA. For instance if there is a plasmid [pause] to get two eh strands that are copied so to have another two daughter DNAs from the original one.

Student S34 correctly identifies diagrams A, B and C (Figure 3.5) as DNA replication intermediates of either ΦX 174 or *E.coli*, respectively. In addition student S34 is able to define DNA replication but expresses this in an awkward manner.

I: Okay, let us talk about strands of DNA. If I draw a straight line and label one end 5' and the other 3', now [pause] what does this mean to you?
S34: In terms of replication it indicates [pause] ehh the order in which the replication will will [stammer] contin' , ehh will go about.
I: My friend, let us talk about this line and not replication for a while [Interviewer points to straight line drawn on a page and marked 5′ and 3′]. What does this line mean? What are the components that this line represents?
S34: [Silence]
I: If we say that DNA runs 5 to 3' and I draw a straight line showing it, what does that line mean?
S34: [Silence]
I: If you look at diagram A, for instance, you can see these lines marked 5 to 3′[Interviewer points features in diagram A] and across we have these lines linking the two strands [I points to hydrogen bonds]
S34: Yes. Yes.
I: Now all I am trying to say is that if you label a line 5 to 3', what does it indicate? I mean what is it that's contained from here to there [Interviewer points at 5 and 3′ ends on straight line and in diagram A]?
S34: Contains? Ehh the information is the coding strand basically.
I: Err, I was not asking you what this strand is called. But what are the components of DNA? Double-stranded DNA is made up of two strands, not so?
S34: Yes.
I: What are these strands made up of?
S34: Strands are made up of bases, basically.
I: Which bases?
S34: Adenine, guanine, thymine and cytosine.
I: What is it that holds these bases?

S34: Mmmm!

I: If there are two strands of DNA, What is it that holds these bases? So, what are we looking at in this diagram [I points to hydrogen bonds in diagram A]?

S34: Those are hydrogen bonds. Yeah!

I: Now this line here [I points to label (e) in diagram A], what does this represent?

S34:[hesitation] I think it represents, ehh [pause] the direction of replication.

I: Is deoxyribose sugar a component of strand e?

S34: Yes.

I: How is it linked?

S34: It is linked to a phosphate.

I: To a phosphate or via a phoshodiester bond?

S34: I think through a…eh, a phosphate link.

I: So, these strands [I points to features (e), (a) and (b) in diagram A], do they indicate base-pairing?

S34: Yes, they would, like adenine to thymine and guanine to cytosine but fragment (a) is an RNA primer where the short DNA forms. There you don't have thymine but uracil will bond adenine in fragment (e).

S34 shows great difficulty explaining the molecular detail or symbolic representation (Treagust *et al*., 2003; Ferk *et al*., 2003) of a line marked 5→3′ and associates strand labels with the direction of DNA replication. The questions are focused on the molecular structure of a DNA polynucleotide chain represented by a line marked 5→3′.

I: In diagram A, you will notice that we have a circle which is indicated as minus, that's the inner strand. On the outside, we have plus [Interviewer points to the strand]. Okay?

S34: Yes. Yes.

I: Now,…basically I have indicated here that the negative strand is copied, as you can see [pointing to the strand] from a particular point but there is also synthesis opposite the plus strand. Okay, you can see that.

S34: Yes.

I: Replication begins at a specific site. What is that site called?

S34: The origin of replication.

I: Okay. Now in this case, [pause] as you can see it is between a G and A base pair. I have not marked this but this would be position 4305 [I points to G] and 4306 [I points to A], okay. Here you see an enzyme attached at the 5' end which I have marked "d" . What is this enzyme called?

S34: I don't know.

I: Do you know how it attaches here [I points at A at the 5′end] at this point?

S34: [pause]Can't remember!

I: Can't remember? Okay, now I have indicated to you that the G base you are looking at here and the A are consecutive bases. This is at position 4305 and this at 4306 but here on this diagram you can see that A is on the linear stretch but the G is around the circle [Interviewer points to diagram A]. Okay? Surely this enzyme [Interviewer points to feature (d)] does something and there are other factors that contribute to this kind of structure. For instance, I will give you a clue. There is an arrow here pointing to (g). What does it tell you? What is it that has happened here?

S34: It shows the…unwinding of the strand.

I: Of the plus strand?

S34: Yes.

I: What enzyme unwinds the plus strand?

S34: Helicases.

I: A helicase?

S34: Yes.

I: Is there a specific name for this helicase?

S34: I can't remember the name.

I: Can't remember the name? Okay.

Does a helicase use a particular substance to carry out its function?

S34: I think it uses ATP.

I: Okay. Let's say per base pair that it separates or unwinds, how much of ATP does it use?

S34: I am not sure.

I: Not sure? Eh this marked in "f", what do think this is? These round globular[pause] right here [I points to the feature].

S34: SSB binding protein.

I: SSB binding protein! What does it do?

S34: They prevent the strand from [pause] eh since they are separating from ehh re-binding.

I: In here, you will notice I have marked various fragments "a", "b", and there is a kind of space here as "c". What do you think are "a", "b" and "c"?

S34: "a" is the RNA strand.

I: RNA strand? Okay, followed by "b"?

S34: cDNA

I: And "c" ?

S34: "c" is a nick.

I: Now, what is it that inserts or synthesizes "a"? What enzyme? What enzyme synthesizes "a" or the RNA?

S34: Polymerase.

I: Are you sure?

S24: I think it is polymerase three (III).

I: Polymerase three? And "b"?

S34: I think also polymerase three.

I: Also polymerase three! You say "c" is a nick. What is a nick?

S34: A nick is a gap between these fragments, the Okazaki Fragments

I: Let's take a break my friend.

This student is unable to name enzymes implicated in the replication process involving ΦX 174 RF DNA. This includes Rep protein, DNA Pol III, DNA primase and gpA protein. Generally, poor detail of enzyme requirements was provided. The molecular basis of enzyme function such as those linked to Rep protein or gpA protein (3.5.8) were inadequately explained. Student S34 is unable to differentiate a nick from a gap (Table 3.7).

Part 2

I: Let us now take a look at diagrams B and C. You told me earlier that these were about a replication bubble.

S34: Yes. These show E. coli replication.

I: Good. Tell me more about the formation of the bubble. By this I mean what proteins play a role in the bubble formation?

S34: I know that protein n, SSB and many other auxiliary proteins bind at the ori site to form the eye or bubble.

I: Can you be more specific?

S34: Well, SSB and n protein, I mean protein n eh, [pause] bind outer strands (i) and (b) and they, also the auxiliary proteins, form the bubble.

I: Does the bubble remain fixed in its size or shape?

S34: No. Helicases unwind it at the forks [S34 correctly points out the replication forks].

I: In doing so, what is it that happens to the bubble?

S34: The unwound DNA becomes part of of [stammer] the template strand which is copied during replication.

I: Excellent. Now take a careful look at diagrams B and C. Is there a difference you feel that is worth commenting on?

S34: Well, mmmm not really. No. They are the same.

I: Now take a look at DNA synthesis within the bubble. Is there something different in the two diagrams? Perhaps you want to tell me more about cDNA synthesis.

S34: Agh! No. Diagram B must show cDNA synthesis in a 5 to 3′direction. The arrow on the leading strand (d) is correct but not the 5 and 3′label. It is correct in diagram C.

I: What is it that contributes to the leading and lagging strands of DNA synthesis?

S34: As unwinding takes place, more of the bubble is exposed as the template. So, the leading strand grows ahead once it starts from the RNA primer but the lagging strand needs a new primer therefore it is short, eh sorry, I mean discontinuous.

I: What are these discontinuous strands called?

S34: Okazaki Fragments.

I: Is the Okazaki Fragment DNA, RNA or a combination of the two. Let us look at diagram C. Is it (g), (e) or (g) + (e)?

S34: Well, all Okazaki Fragments have the RNA primer, so it must be (g)+ (e).

I: What enzymes are implicated in the synthesis of the Okazaki Fragment?

S34: It must be DNA polymerase.

I: But, you just said to me that the Okazaki Fragment contains RNA.

S34: I think I made a mistake.

I: What is it?

S34: I think the Okazaki Fragment is the DNA part not the RNA because a different polymerase makes the RNA.

I: Which?

S34: Eh [pause] I forget the name. It may be RNA polymerase.

I: Can you tell me more about the feature marked (f) in diagram C?

S34: It is a nick or like a gap in the DNA separating the Okazaki Fragment.

I: Take a look at the strands labelled (c) and (i) in diagram C. What are these strands called and what is the fate of each strand [pause], say at the end of replication?

S34: Those are parental template strands. Each one is linked to new cDNA.

I: Very good. But tell me more about this mechanism. What does it tell you about the mode of replication?

S34: It is semi-conservative because each template is associated with the new DNA.

I: Thank you for giving me this input.

S34 is unable to give a precise definition of the actual constituent of a discontinuous strand named "the Okazaki Fragment". In this regard, there was a tendency to incorrectly associate both the RNA primer and adjoining cDNA as constituents of the Okazaki Fragment. In addition, S34 erroneously indicates that both components are synthesized by DNA polymerase. When challenged, S34 subsequently indicates that RNA primer synthesis is due to the action of an RNA polymerase and that the RNA is not part of the Okazaki Fragment. The latter is correct; however, there was no mention of the role of DNA primase which is in fact implicated in RNA primer synthesis (3.5.8). At first glance, S34 is unable to tell the difference between diagrams B and C until directed to a flaw (3.5.8; Figure 3.5) in the labelling of cDNA synthesis within the replication eye or bubble of diagram B. Subtle changes, even to salient features of a diagram, may not be

clearly visible or apparent to the learner (Ploetzner and Lowe, 2004; Crisp and Sweiry, 2006) or have little impact on the learner who is trying to process larger amounts of information on the diagram in a short time (Mayer, 2003; Schnotz and Bannert, 2003). It is reasonable to assume that the diagrams carried too much detail and perhaps was an overwhelming burden to this student's cognitive capacity (Lewalter, 2003; Kozma, 2003). The role of the parental template strands and their fate in semi-conservative DNA replication (Meselson and Stahl, 1958) was satisfactorily explained.

7.5 Discussion

At the inception of this work, five categories of difficulties (Table 7.1), concerning nucleic acid-strand nomenclature and function, have been reported previously in the literature (Gupthar and Anderson, 2003). Subsequently, Fossey and Hancock (2005) reported on the similar difficulties in a qualitative manner, covering fewer examples of student input than the current survey. Both studies, however, revealed that a poor understanding of nucleic acid-strand nomenclature and function contributed largely to the difficulties that students had experienced with the basic mechanism of DNA replication and transcription. Recently, Cook *et al*. (2006) reported on students' inability to interpret the replication process because of their limited knowledge about DNA-strand function, despite their ability to focus visually on salient replication features in the same way as experts do. The various difficulties reported in this study (Table 7.1) are classified at level 3 or "partially established" in terms of the four-level classification scheme provided by Grayson *et al*. (2001). Recently, much has been written about the nature of students' difficulties with replicative DNA synthesis (Fossey and Hancock, 2005; Cook *et al*., 2006; Beltramini *et al*., 2006; Rotbain *et al*., 2006). The nature of students' difficulties, presented in Tables 7.2 and 7.3, are similar to those reported in the literature although the current study involves the use of different replication models and reports a wider range of student difficulties. Many of the difficulties tabulated (Tables 7.2 and 7.3) are supported by students' illustrations of the replication process (Figures 7.3 and 7.4). The difficulties are currently classified as "partially established", level 3 difficulties in accordance with

the four-level classification scheme provided by Grayson *et al*. (2001). Diagram interpretation can be influenced positively or negatively by the content or conceptual knowledge a learner already knows (Henderson, 1999; Cook *et al*., 2006). At times, however, information gained from a diagram can be learnt more effectively than when presented in textual form (Kindfield, 1994; Crisp and Sweiry, 2006) and can be enhanced by tasks relating to the analyses and interpretation of diagrams (Barlex and Carre, 1985; Henderson, 1999). Models of DNA replication intermediates may be constructed using atomic space fill-type kits (Beltramini *et al*., 2006) or drawn using coloured pencils (Fossey and Hancock, 2005) to highlight different or salient features; however, an understanding of the molecular basis of these representations is essential (Treagust *et al*., 2003; Ferk *et al*., 2003; Cook *et al*., 2006; Rotbain *et al*., 2006).

CHAPTER 8

DEVELOPMENT AND PRELIMINARY TESTING OF A REMEDIATION MODEL

8.1 Development of a remediation model

Based on teaching experience, the author deems that the following factors impact on the development of a remediation model and elaborates on the influence of each (8.1.1, 8.1.2, 8.1.3, 8.1.4 and 8.1.5) with regard to his own configuration of such a model:

(i) The purpose or need for a model.

(ii) The representation and extraction of conceptual knowledge.

(iii) Conceptualization patterns and expression of knowledge.

(iv) Remedial strategy.

(v) Re-assessment of understanding.

Teachers may have other "notions" which factor into model construction such as the entities constituting the model, its uniqueness and nature, basis for its accreditation, its ability to make predictions and the time span concerning its use (Justi and Gilbert, 2003). Thus the author is aware that researchers consider a range of factors which may influence the design and purpose of different remediation models. In this chapter, the author discusses the factors (8.1.1, 8.1.2, 8.1.3, 8.1.4 and 8.1.5) that have influenced the design of his original model (Figure 8.1). In addition, the author shows the potential of his model in the remediation of students' difficulties with the use of symbolism in Molecular Biology.

8.1.1 The need for a remediation model

A remediation model should serve the purpose of a pedagogic tool (Justi and Gilbert, 2002; 2003), enabling the teacher to assist students overcome a range of learning

difficulties as shown in this study (Chapters 4 to 7). Inappropriate or erroneous information transfer, rote learning, surface-level understanding, weak epistemic scaffolding, poor expression of knowledge influenced by constructivism, formation of alternate mental schemata, poor reasoning or differentiating ability, a lack of lateral thinking, poor mental imagery or visualization were commonly encountered using symbolic representations in the teaching of Molecular Biology (Chapters 4 to 7). In identifying and carefully assessing these difficulties from both verbal and written expressions (Black and William, 2006; Yore and Treagust, 2006), it is possible to establish "what the learner already knows" (Ausubel, 1968, cited by Schönborn and Anderson, 2006a) about symbolism versus the depth of information it actually depicts. This enables the offer of appropriate remedial assistance, ideally fostering metacognitive awareness (Guterman, 2003; Georghiades, 2004) and the promotion of further learning (Duit and Treagust, 2003). The identification of critical (Huitt, 1998; Mayer, 2003) or threshold knowledge (Meyer and Land, 2003; Cousin, 2006) that promotes conceptual understanding is vital in any remediation exercise (Hewson and Thorley, 1989; Vosniadou and Ioannides, 1998), especially those adopting Socratic dialogue-inducing methods (Hake, 2008). Often, fundamental knowledge or protoconcepts (Fisher, 1985) which underpin the more complex or supraordinate information (Novak, 1984; Herron, 1996) are poorly understood, as explained by principles of epistemic scaffolding (Bruner, 1960; 1973, cited by Cooper, 2005; Sandoval and Reiser, 2004; Sins *et al*., 2005), thus impeding independent, heuristic learning (Novak, 1984) or the ability to understand the more complex (White and Gunstone, 1992; Duit and Treagust, 2003). Both fundamental and supraordinate knowledge usually constitute the underlying information depicted by symbolic representations, making the learning of information from such knowledge domains absolutely essential. This type of learning serves as a pre-requisite tool which enables learners to decode or explain symbolic representations more fully (Kozma, 2003).

In accordance with Vygotskian principles, remediation exercises and assistance with particular problems can be beneficial to the learner (Chapter 1; 1.2). It is imperative that previously taught concepts be re-visited, perhaps in "a spiral manner" as suggested by

Bruner, (1960; 1973; cited by Cooper, 2005) to promote familiarity and to reinforce or consolidate the learning process (1.2; 1.3; 1.3.1; 1.4.1).

8.1.2 Representation and extraction of content knowledge

The scope and depth of content knowledge which relates to a module, particular topic or symbolic representation can be defined for the level of study. Content knowledge can be presented in a sequential and logical format, usually characterized by the introductory or foundation information which precedes the more detailed content. When content knowledge is transformed to teaching formats that enable better understanding by the learners, this is acknowledged by some researchers as "pedagogic content knowledge" (Shulman, 1986; Chen and Ennis, 1995; Darby, 2005). Such an understanding may be influenced by the prior learning of related information (Rhemtula and Rollnick, 2002), followed by the subsumption (Ausubel, 1968, cited by Schönborn and Anderson, 2006a) and integration of new, related knowledge into existing mental schemata (Carey, 1985; Pearsall *et al*., 1997; Vosniadou and Ioannides, 1998). In this regard, a Socratic teaching approach may be beneficial to both teacher and learner as the teacher establishes the extent of understanding (Guterman, 2003; Hake, 2008). Some learners have a tendency to reflect on information from prior learning which might be unrelated in terms of the content knowledge which applies to a particular topic or symbolic representation. This type of information can be extracted from an extraneous source (ES) (Figure 8.1) such as the world-wide web (www) or library (de Jong, 2000; Kasanda *et al*., 2005) and must be identified by the teacher. The extraneous source is depicted in the model as "[ES]" and the "+" sign (Figure 8.1) denotes that the [ES] is the additional source of information that learners assimilate apart from that disseminated by the structured "information bank".

Content knowledge can be represented generically in a remediation model, constituting an "information bank" (Figure 8.1) or framework which shows the:

i. Logical and sequential presentation of foundation supraordinate knowledge domains or sections, for example, as A, B and C, followed by

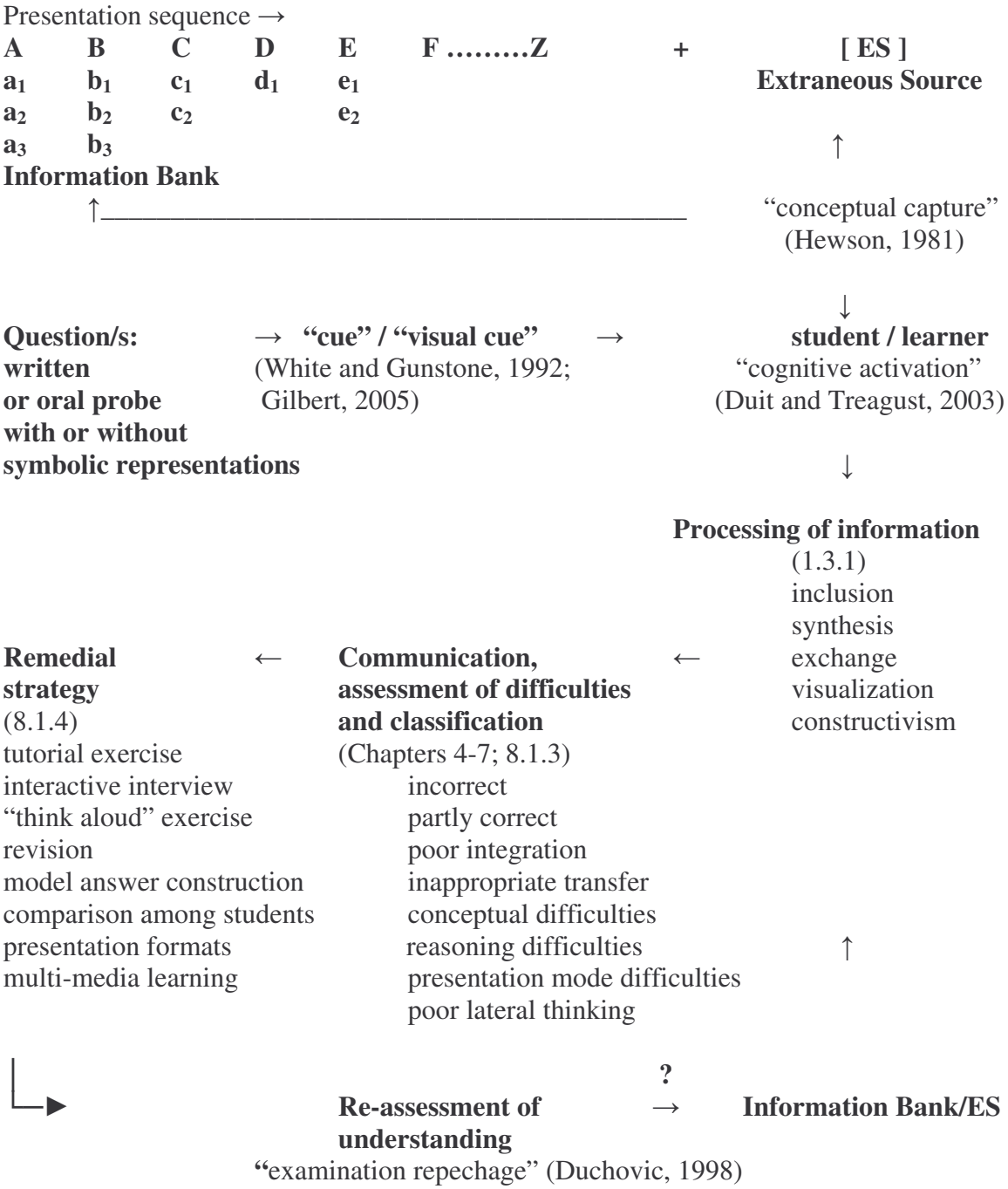
domains or sections of the more complex or supraordinate content knowledge or information, D, E , F, G or H etcetera. Knowledge domains may also constitute the representation of information using a range of symbolism as discussed earlier (Chapter 2) and the accompanying underlying information that is depicted by symbolic representations.

ii. Presentation of various subsections of subordinate concepts or protoconcepts as a, b, c (etcetera), which underpin the respective more complex or supraordinate content knowledge or information. Where applicable, a hierarchy of information comprising supraordinate (e.g. D→E→F→G→H), co-ordinate (equal in level or rank e.g. co-ordinate protoconcepts a_1, b_1, c_1, d_1 or e_1) and subordinate concepts (e.g. subordinate protoconcepts $a_3 \rightarrow a_2 \rightarrow a_1$) (Novak, 1984; Herron, 1996) could be depicted in a similar way [Key: → denotes increasing hierarchy]. Smaller to larger molecular constituents of complex structures can be described or illustrated hierarchically as shown in Figure 8.1.

Using the information above, the author illustrates the organisation of an "information bank" (Figure 8.2) concerning the content information which he presented as lectures to the students on monoclonal antibody production. Monoclonal antibody production is the subject of the remediation exercise in this study (8.2.1; 8.2.2).

8.1.3 Conceptualization patterns and expression of knowledge

Questions, in written format or interview protocol, provide a "cue" to the learner (White and Gunstone, 1992), initiating "cognitive activation" and interaction with mental schemata which have been influenced by the "information bank" or "ES" during the learning process (Figure 8.1). Such questions may be accompanied by a presentation of symbolic representations to learners so that these act as visual cues for information retrieval from an "information bank" or "ES". "Conceptual capture" (1.3.1) may be



? : Re-visit and guidance to content knowledge domain(s) if necessary.

Figure 8.1 Representation of a remediation model.

followed by processing of the knowledge involving "incorporation into mental schemata", "exchange through dissonance", "syntheses", "knowledge construction" or "visualization" (Chapter 1; 1.3.1). The expression of knowledge or communication may be viewed as an external representation of a complex internal and idiosyncratic mental process. Such representations may reveal knowledge expression in various ways, including the unexpected, which will require analyses. Some examples include:

i. Correct answers, reflecting the specific extraction of information from a content knowledge domain, D for example, from the "information bank". This may be integrated with knowledge derived from prior learning.
ii. Correct answers, reflecting the integration of information from co-ordinate knowledge domains of the "information bank" coupled with protoconcepts or information from a subordinate knowledge domain, for example $E + F + e_{1.}$
iii. Erroneous or inappropriate answers, reflecting unwanted information from the knowledge domain H instead of B, or perhaps from ES, the extraneous source.
iv. Partly correct answers reflecting the integration or coupling of information from knowledge domains B + C, with unwanted information from domain H.
v. Incorrect answers influenced by incorrect interpretation or poor differentiation of knowledge from correct knowledge domains.

Learners could reveal difficulties with conceptual aspects, reasoning ability and representation mode of the information as explained by Schönborn and Anderson (2006b) using Venn diagram logic to show the inter-relationship of these factors in knowledge expression. The appropriate remedial strategy may be offered once the nature of knowledge expression is understood.

8.1.4 Remedial strategy

Remedial strategies (1.6) may be adopted once the nature of learning difficulties is identified and classified (Grayson *et al*., 2001). Where "conceptual capture" (Hewson, 1981) and "the processing of information" (Duit and Treagust, 2003) have lead to inappropriate or erroneous information transfer or difficulties integrating or expressing information (Chapters 4 to 7), learners must be directed to learn (Darby, 2005) from the specific knowledge domain(s) of the information bank where correct information is presented. This pre-remediation strategy (8.2.2) could be adopted preceding a tutorial or POE exercise (Liew and Treagust, 1995; Kearney, 2004), problem-solving exercise (Taconis *et al*., 2001), class debate (Zohar and Nemet, 2002), personal consultation involving an interactive interview (Posner and Gertzog, 1982) or Socratic dialogue (Hake, 2008) or "think aloud" exercise (de Jong, 2000; Schönborn *et al*., 2002a), revision in class emphasizing model answer construction and a reflection on the learners' initial attempt at answering a question or circulation of correct answers written by class colleagues to enable comparison (Duchovic, 1998). Student group studies can be beneficial but must be monitored not to produce a social constructivist type of learning environment (Kearney, 2004). These initiatives must serve to improve the understanding and expression of content information in required contexts in order to be regarded as useful "remediation initiatives".

The learning of specific information from well-organised, logically presented knowledge domains (Novak, 1984; Wandersee *et al*., 1994; Herron, 1996) may be subject to "examination repechage" (Duchovic, 1998) where the learner is challenged on foundation or subordinate knowledge, protoconcepts, specific definitions, the ability to differentiate factual information or the use of information in specific contexts. Rote learning (Grayson, 1995) and surface-level understanding (Chi *et al*., 1981; Kozma and Russell, 1997) of subordinate knowledge undoubtedly contributes to weak epistemic scaffolding (Bruner, 1960; 1973, cited by Cooper, 2005); Sandoval and Reiser, 2004), characterised by an inability to understand supraordinate knowledge domains. The retrieval of information from unwanted domains, including the "ES" (Figure 8.1), should be pointed out, as well

as undesired forms of knowledge construction (Driver and Bell, 1986; Von Glasersfeld, 1992) or expression of alternate mental frameworks (Grayson, 1995; Taber, 1998). The recognition and learning of identical content knowledge in different presentation formats (Seufert, 2003), such as in different diagrams (Kindfield, 1994; Schönborn *et al*., 2002a; 2002b)), models (Huddle *et al*., 2000; Beltramini *et al*., 2006; Rotbain *et al*., 2006) or videos (Zahn *et al*., 2004) for example, are useful to learners as complex information could be made clearer or simpler, without compromise on the depth of information that has to be learnt (Kozma, 2003; Mathewson, 2005). Multi-media are generally known to present information in different modes or formats, potentially assisting learners overcome conceptual and reasoning difficulties, poor mental imagery or visualization which may arise using a single presentation format (Mayer, 2003; Parslow, 2004; Mathewson, 2005). Poor lateral thinking and the inability to reason (de Bono, 1967; Pata and Sarapuu, 2006) may result when there is poor assimilation (Thorndyke and Stasz, 1985) of content information. Often there is need to reflect on content information in a holistic manner, followed by a finer reflection on critical information (Huitt, 1998; Mayer, 2003) or vital threshold concepts (Meyer and Land, 2003; Cousin, 2006) as elements from specific knowledge domains (Huitt, 1998; Figure 8.1). A recombination of such elements or domain-specific knowledge is factored into analytical or deductive reasoning (McMillan and Schumacher, 1993; Pata and Sarapuu, 2006). Where domain-specific knowledge is problematic to understand, analogies (Treagust *et al*., 1992; Dagher, 1995; Orgill and Bodner, 2007) or a simplification of the facts (Pittman, 1999) may be adopted circumspectly to lessen the cognitive burden (Kirschner, 2002; Kozma, 2003); however, the substitution of irrelevant detail from such analogies must be pointed out for deletion or omission from all forms of communication (Hegarty, 2004; Ploetzner and Lowe, 2004; Orgill and Bodner, 2007). Learners may be guided to reflect on such knowledge domains metacognitively (Guterman, 2003; Georghiades, 2004) and be trained to work with critical information in a reasoning exercise or application (Kogut, 1996; Huitt, 1998; Mayer, 2003).

8.1.5 Re-assessment of understanding

The re-assessment of content information understanding is an important indicator of the efficacy of a remediation strategy. It is imperative to find out whether or not a re-test, involving the use of the same probe, will show an improved understanding of content knowledge, following the implementation of a remediation strategy. Duchovic (1998) recommends "examination repechage" (Figure 8.1) where learners are given a "second chance" at an examination, following close interaction with an instructor (Darby, 2005) or other members of a class to assist with corrections. Instructional dialogue from an instructor may also stimulate learning, as well as improve relations between learner and instructor. These are facets of "instructional" and "relational" pedagogy, respectively (Darby, 2005) which contribute to a supportive learning environment. Re-assessments are generally followed by relevant statistical analyses to verify the success of the remediation strategy. Unsuccessful attempts should be revisited and a different remedial strategy attempted. Referral to the relevant information bank domains, simplification of the facts and repeated appraisal of learning attempts may be necessary. Such problems may be peculiar to the understanding of specific content knowledge and may require a particular strategy (Chapter 1; 1.6) to alleviate the learning difficulty or aptitude to learn (Snow and Lohman, 1984). Recent work suggests that remediation models, which suggest conceptual change in particular ways, may be found deficient without the recognition of the thoughts and reflection of the mind as well as the socio-cultural influences that exist in the world outside the mind (Alexander, 2007).

8.2 Preliminary testing of a remediation model: A case study

8.2.1 Participants

A total of 30 students, comprising group n_5 (3.1), participated voluntarily in a remediation exercise, facilitated using guidelines of the newly developed model (8.1; Figure 8.1). In

their consultations with the writer, individuals or groups of up to 4 students drew attention to their poor performance in a written test covering, amongst other topics, the theory on gene marker symbolism and hybridoma selection (3.5.6.2). In their written test, students received the same focused probe on hybridoma selection as described (3.5.6.2). The consultations were viewed as a request for assistance by the students and a concern that this lecture topic did present difficulties in the learning process. As the author intended to help a cohort of weak students (Figure 8.3 A) with their difficulties, no control group of students was exempt from the pre-remediation exercise (8.2.2), specific intervention (8.2.4.1) and post-remediation test (8.2.5).

8.2.2 Pre-remediation exercise

Gene marker symbolism and the hybridoma fusion technology (3.5.6.2) were re-visited in a pre-remediation exercise to re-establish:

i. The level of understanding shown by each student.

ii. The nature of difficulties experienced by the students.

Students were supervised (as described below) to do an assignment, providing written answers to the focused probe (3.5.6.2), namely;

"Discuss the basis for hybridoma $Hgprt^{+}$ Tk^{+} selection using HAT medium. Why is the same medium detrimental to hybridoma progenitor cells of the phenotype, $Hgprt^{+}$ Tk^{-} or $Hgprt^{-}$ Tk^{+} " ?

The probe carried a maximum assessment mark of 10 for an answer which contained the correct propositional knowledge as described in section 3.5.6.2. Learners were guided to find such knowledge within appropriate domains of content knowledge (8.1.2) which resembled the "information bank" (Figure 8.2) constructed by the author. The knowledge domains define the content knowledge covered by the author during his lectures on monoclonal antibody production. It defines a sequence of presentation of supraordinate knowledge domains and their accompanying subordinate domains as shown (Figure 8.2).

The "knowledge within appropriate domains", as shown in the skeletal framework of Figure 8.2, could be found in their notes constructed during lectures, lecture supplements, relevant publications or the use of prescribed textbooks. The author ensured strict supervision that individual written contributions and were in fact a learner's own interpretation of facts and written effort, emanating from this guided search of information from relevant knowledge domains. The retrieval of inappropriate information or erroneous transfer of information from existing mental schemata, possibly derived earlier from unprescribed, extraneous sources (ES) (Figures 8.1 and 8.2) such as the world-wide web (www.), video or electronic media was also expected by the author. In this pre-remediation exercise, the students' answers were compared with the propositional knowledge (3.5.6.2) to assess the correctness of processed content knowledge from either an "information bank" and /or "ES", in response to the probe. The assessment breakdown, concerning anticipated structural components of written answers, is given in section 8.2.3.3.

8.2.3 Assessment of pre-remediation exercise

8.2.3.1 Extraction of information

The probe (8.2.2) did serve as a cue for most of the students of group n_5. Three students did not submit their written efforts for assessment and consequently received an assessment score (8.2.3.3) of zero out of ten. These students later admitted that they were unable to construct a written response despite receiving the assistance to retrieve information from relevant knowledge domains. The remaining twenty seven students presented answers that were largely incomplete and incorrect. There appeared to be evidence that the students did refer to information characterising knowledge domains D, E and F (Figure 8.2) to construct their answers but the level of understanding of the information within the subordinate domains of the above varied considerably. The students showed various categories of difficulties (8.2.3.2), possibly influenced by poor knowledge assimilation or "conceptual capture"

A: Introduction: Monoclonal antibody (mAb) production

- **a_1:** immune response: general aspects
- **a_2:** need
- **a_3:** monoclonal antibodies
- **a_4:** historical perspectives and basic strategy

B: The antigen

- **b_1:** immune response: primary and secondary responses
- **b_2:** use of adjuvants
- **b_3:** mode of administration
- **b_4:** properties

C: The animal

- **c_1:** B cell sensitization and harvest
- **c_2:** pre-immune serum
- **c_3:** heterozygosity and health

D: Cell types

- **d_1:** marker symbolism and selection
- **d_2:** mutations
- **d_3:** properties

+ [ES] information from extraneous sources

E: Cell fusion

- **e_1:** types of fusion products and stability
- **e_2:** techniques

F: Selection of the hybridoma

- **f_1:** propagation of hybridoma clones
- **f_2:** integrated role of biochemical pathways, media components and markers
- **f_3:** growth conditions
- **f_4:** HAT medium

G: Harvest and test of mAb specificity

- **g_1:** Immunological techniques
- **g_2:** Harvest protocol

Figure 8.2 Skeletal structure of knowledge domains and their respective subordinate domains constituting the information bank on the hybridoma cell fusion technology. A to G (upper case alphabetical order) indicates the sequence of presentation of supraordinate knowledge domains. Subordinate domains are indicated by lower case letters and subscript numbers. The hierarchical order of presentation of each subordinate domain follows a decreasing number order, e.g. $a_{4,}$ a_3 a_2 and $a_{1.}$

(Hewson, 1981) and the ensuing inability to "process" information correctly relative to the cue. It prompted a closer inspection of the notes compiled by the students which were found to be fairly accurate, however, accompanied by careless spelling errors and a variety of unique shorthand notation and use of acronyms or abbreviations. Examples, included "G-base" for guanine, "HX" or "hypoX" for hypoxanthine, "Hgprt" and "Tk" for the enzymes hypoxanthine guanine phosphoribosyl transferase and thymidine kinase respectively, "Am" for aminopterin, "PEG" for the fusogenic agent polyethylene glycol, "DHF reductase" and "Ts" for the enzymes dihydrofolate reductase and thymidylate synthase, respectively and "H-dome" for hybridoma. This inspection was not part of the remediation strategy; however, the author admits doing this as he was curious to find if indeed this information came from the students' compilation of notes.

8.2.3.2 Processing and expression of knowledge

In response to the probe (8.2.2), group n_5 students showed the following recurring categories of difficulties (Table 6.1) concerning the hybridoma cell fusion technique (6.2);

i. category i: Use of abbreviations and acronyms without definition or nomenclature as indicated above (8.2.3.1).
   Incidence: 18/30
   Classification of difficulty: Level 2 (Grayson *et al*. 2001).

ii. category ii: Synthesis of new words, vaguely resembling those encountered in the science of cell fusion technology. Examples included; "G-base" for guanine, "hypoX" or "hypozanthein" for hypoxanthine, "hetrokarones" for heterokaryon and "aminoptic acid" for aminopterin.
   Incidence : 15/30
   Classification of difficulty: Level 2 (Grayson *et al*. 2001).

iii. category iii: Erroneous substitution of nomenclature and related facts. Examples included; “the conversion of dUMP to dTMP via the enzyme thymidine kinase”, “the conversion of thymidine to thymidine monophosphate via the enzyme thymidylate synthase”, “the enzyme Hgprt converts HX to GMP using PRPP” and “Hgprt forms IMP using pyrophosphate and hypoxanthine or GMP using pyrophosphate and guanosine ”.

Incidence : 15/30

Classification of difficulty: Level 2 (Grayson *et al*. 2001).

iv. category iv: Poor understanding of the concept of a marker as reflected by $Hgprt^{+}$ Tk^{-} or $Hgprt^{-}$ Tk^{+} . Examples included the following ; (a) “B cells $Hgprt^{+}$ Tk^{-} will not grow the enzyme thymidine kinase (Tk^{-}) in HAT medium, likewise myelomas will not grow the enzyme $Hgprt^{-}$ ”. (b) “The enzyme thymidine kinase is not found in the Tk^{-} gene therefore cells which are Tk^{-} will not convert thymidine to TMP. $Hgprt^{-}$ cells do not have the enzyme Hgprt in its genes therefore they cannot convert HX to IMP”.

. Incidence : 19/30

Classification of difficulty: Level 3 (Grayson *et al*. 2001).

v. category v: Poor integration of knowledge and problems with reasoning. Examples which illustrate this category of difficulty include;

(a) “The hybridoma expresses both salvage enzymes therefore they will survive in HAT medium while the others will die since they express only one of the two enzymes”. (b) “B cells which are $Hgprt^{+}$ Tk^{-} will not survive in HAT because they cannot produce thymidine kinase and aminopterin blocks TMP formation from thymidine which results from phosphorylation of thymidine by thymidine kinase”.

Incidence : 27/30

Classification of difficulty: Level 2 (Grayson *et al*., 2001).

8.2.3.3 Assessment scores

The following is a breakdown of the marks allocated (score) to each of the anticipated structural components of written responses is indicated as follows:

Structural component	*Mark allocation*
Cell types and marker function	2/10
Fusion strategy and fusion products	2/10
Role of media components in selection of the hybridoma	2/10
Fate of cell types: integrated role of markers, media components and biochemical pathways	4/10
Total marks	10

Pre-remediation assessment of student scores (Table 8.1) is expressed using standard statistical indicators, viz., the mean, standard deviation, median mark, mode and distribution frequency (f) (Figure 8.3A) following simple guidelines described by Duncan *et al*. (1983). A mean and standard deviation of 2.833 and 1.422 were recorded, coupled with values of 3.5 and 4 for median and mode, respectively (Table 8.1). The distribution frequency (f) was asymmetrical and unimodal, showing the mode value of 4 and confined to a score dispersion range between 0 and 4.5 (Figure 8.3A). These values indicate a poor performance and a failure of the student group to obtain a minimum pass mark of 50% as indicated by the score dispersion range. The distribution frequency plot shows that 8 and 3 students of the group (n_5 =30) (3.1) obtained the higher end marks of 40 and 45 %, respectively. Microsoft Office Excel 2003 (Microsoft Corporation, USA) was used to process the raw data (Table 8.2) which enabled a comparison with post-remediation assessment scores (8.2.5.3).

Table 8.1 Raw data and statistical analyses on pre-remediation assessments

Student No.	Pre-remediation score	Statistical Indicator
1	4	**Mean:** 2.833
2	4	
3	1	
4	1	
5	2	
6	2	***SD:** 1.422
7	0	
8	1	
9	2	
10	4	
11	3	
12	3	**Median:** 3.5
13	3.5	
14	3.5	
15	3.5	
16	4.5	
17	4.5	**Mode**: 4
18	4.5	
19	0	
20	0	
21	2.5	
22	2.5	**Maximum**
23	4	**Score :** 10
24	4	
25	3.5	
26	3.5	
27	2	
28	4	
29	4	
30	4	

*SD: Standard deviation

8.2.4 Implementation of the remediation strategy

The remediation strategy adopted by the current model suggests the offer of one or more intervention measures (8.1.4) once the nature of the learning difficulty is established and classified (8.2.3.2). The model also emphasizes that students be guided to learn from specific knowledge domains constituting the "information bank". The offer of remediation involves specific intervention (8.2.4.1) to promote correct learning and the heuristic discovery of facts from relevant knowledge domains. While students may be shown the specific knowledge domains for information (Vygotsky, 1978, cited by Taylor, 1994), it is imperative that they discover the facts by themselves and incorporate these into mental schemata (Thorndyke and Stasz, 1985). These outcomes may be evaluated by a re-test (Duchovic 1998) involving the original probe and interviews using different questions to establish the understanding of protoconcepts (Fisher, 1985) or finer detail. Based on the author's experience, it is unwise to introduce the students to different probes without establishing fully whether or not the remediation strategy succeeds in alleviating the initial difficulties recorded with the original probe. Also the use of different probes too soon tends to be confusing to students as content information may not be fully understood or incorporated poorly into mental schemata. In the current remediation exercise, a series of questions concerning subordinate knowledge domains which focus on protoconcepts are given as a take-home assignment (Kogut, 1996) (8.2.4.1). This challenges the student to find answers which underpin more complex information required by the probe. The author places great emphasis that finding answers to such questions may not suffice to demonstrate the students' ability to synthesize the information in a form that is required by the probe. This synthesis also requires the meaningful learning of domain-specific knowledge. Following remediation, it is desirable that students show the ability to reflect on mental schemata which have integrated correct information from the knowledge domains of an "information bank". Cues of different probes should ideally elicit the appropriate communication of responses once meaningful learning has taken place (Shulman, 1986; Ametler and Pinto´, 2002). In this study, the author only resorts to the use different questions in interviews to evaluate the success or

failure of adopting certain interventions during remediation and to evaluate whether or not students could answer different oral probes which demanded finer details of their understanding of the topic. This also serves to interrogate the possibility that some learners could focus on providing an answer to a repeatedly used probe and not have the ability to give answers to other related questions.

8.2.4.1 Specific intervention

Further consultation with the learners, either in small groups or individually, was necessary once the categories of learning difficulties (8.2.3.2) were established. This enabled the offer of specific intervention measures as follows and the assessment of a metacognitive learning process (Georghiades, 2000; 2004). Each learner presented with a particular category of difficulty or most commonly with a combination of difficulties fitting various categories (8.2.3.2). Accordingly, the appropriate intervention measures had to be applied to assist with such variation in learning difficulties. The intervention measures, addressing each category of difficulty, are indicated as follows.

i. Category i: The use of abbreviations and acronyms without definition or nomenclature at first mention was pointed out to all learners as an unacceptable practice. Specific instruction emphasized the use of full names at first mention. Learners were further instructed that acronyms or abbreviations could be indicated alongside the full names in parenthesis and that the subsequent use of abbreviations or acronyms could be allowed where the definition of each is clearly indicated in earlier text or a key.

ii. Category ii: The syntheses of new words, and those incorrectly spelt, were highlighted in written responses for correction. Learners were also referred to specific knowledge domains (Figure 8.2) where the correct spelling would be found. In addition, the consultation of

glossaries in prescribed textbooks was suggested to assist in making the corrections and to reinforce the meaning of words.

iii. Category iii: The erroneous substitution of nomenclature and related facts were subject to challenge exercises where questions were designed to evoke a thought process that differentiates information. For example, the difference in the reactions driven by thymidine kinase and thymidylate synthase was determined in this manner. Instructional cues or commands such "tell me the difference", "what is the difference" (?) or "can you differentiate" (?) were incorporated into questions, instigating the re-examination of inappropriate or erroneous substitution and challenging the learner to provide correct information on each conceptual matter. For example, (a) "Can you differentiate the enzymes thymidine kinase and thymidylate synthase?" and (b) "Tell me the difference in the reactions facilitated by thymidine kinase and thymidylate synthase?". Learners were guided to consult the correct knowledge domains where the information would be found. Often the learning of protoconcepts, found in subordinate knowledge domains, assisted in the correction of substitution errors. For example, the identification of substrates required by the enzyme, hypoxanthine guanine phosphoribosyl transferase (Hgprt) is actually spelt out in the name of the enzyme itself. Learners were referred to the compounds hypoxanthine, guanine and phosphoribosyl pyrophosphate as substrates facilitating two different reactions (3.5.6.2) in the purine salvage pathway (Subordinate domain f_2, Figure 8.2) and the end products, GMP and IMP, to bring meaning and definition of the enzyme function. Substitute substrates such as guanosine and pyrophosphate (8.2.3.2) were pointed out for deletion using the basic name of the salvage enzyme as a cue and emphasizing the Chemistry involving substrate conversion to product.

iv. Category iv: The poor understanding of the concept of a marker as reflected by *Hgprt*$^{+}$ *Tk* $^{-}$ or *Hgprt* $^{-}$ *Tk*$^{+}$ required a special emphasis on the definition of several protoconcepts. Learners were directed to knowledge domains D (Cell types) and F (Selection of the hybridoma) and the respective subordinate domains d_3, d_2, d_1, f_4, f_3 f_2 and f_1 where information would be found on cell properties, mutations, gene marker symbolism and selection of markers, HAT medium components, growth conditions for hybridomas, biochemical pathways affecting hybridoma selection and propagation of hybridomas, respectively. The search for information from these subordinate domains is intended to promote the understanding of protoconcepts and the study of the inter-relationship between markers, biochemical pathways and media components which allow for hybridoma selection (3.5.6.2). In addition, learners were given a take-home assignment (Kogut, 1996) to search for information from the respective knowledge domains and were also encouraged to use their prescribed textbook in order to answer the following questions:

a) What is a gene?
b) What is a gene marker?
c) What is a gene notation?
d) Define *Hgprt*$^{+}$ *Tk* $^{-}$ or *Hgprt* $^{-}$ *Tk*$^{+}$.
e) Differentiate the abbreviation Hgprt from *Hgprt* $^{-}$ or *Hgprt*$^{+}$.
f) Differentiate the abbreviation Tk from *Tk* $^{-}$ or *Tk*$^{+}$.
g) What do mean by gene expression? Can you relate this to *Hgprt*$^{+}$ *Tk* $^{-}$or *Hgprt* $^{-}$ *Tk*$^{+}$?
h) What are complementary gene markers? Why are they used in the hybridoma fusion technique?

Students were encouraged to consult with the lecturer to assess the answers (Duchovic, 1998) or to assist with conceptual difficulties with the above. The understanding of foundation knowledge is of paramount

importance in “epistemic scaffolding” (Bruner, 1960; 1973, cited by Cooper, 2005; Sandoval and Reiser, 2004; Sins *et al*., 2005).

v. Category v: Poor knowledge integration (Thorndyke and Stasz, 1985) and reasoning difficulties, associated with the hybridoma selection process, necessitated learner assistance initially with domain-specific knowledge (Figure 8.2), followed by the use of a simple analogy (Treagust *et al*., 1992; Dagher,1995) where domain-specific knowledge (Huitt, 1998; Mayer, 2003) was recombined to show why an organism survives in a certain medium. These interventions are described as follows.

Firstly, the assistance with domain-specific knowledge included a referral to subordinate domains d_1, d_2 , d_3 , f_1, f_2 , f_3 and $f_{4;}$ (Figure 8.2), where information would be found to answer key study questions, for example;

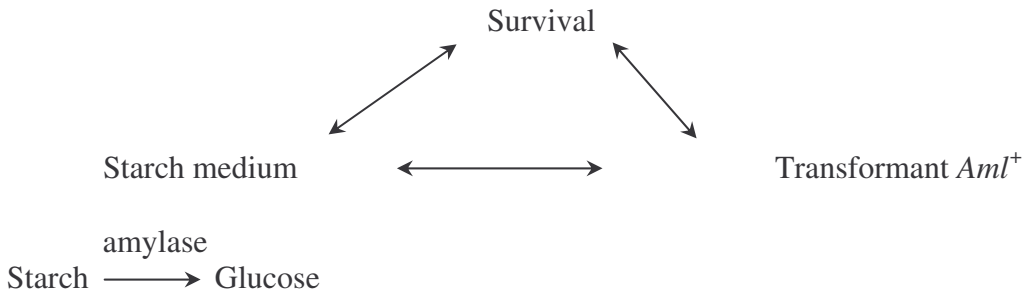
a) What are the main components of HAT medium?
b) Which component in HAT medium inhibits *de novo* synthesis of purine nucleotides and certain steps in pyrimidine nucleotide biosynthesis? List the inhibited reactions.
c) Name the cells implicated in the hybridoma fusion process and comment on how such cells overcome the inhibitions described in (b)?
d) Can you relate the gene markers used in the hybridoma fusion technology with certain enzymes?
e) Why does the hybridoma survive in HAT medium and not the parent or progenitor cell types?

Learners were encouraged to consult with the lecturer to assess the answers or to assist with conceptual difficulties as described above (Category iv). The interaction among students involving their discussion of answers was not monitored and difficult to assess as a remediation

strategy or factor influencing the understanding or misunderstanding of information in any way.

Secondly, learners were also exposed to reasoning using a recombination of domain-specific knowledge in predicting cell survival. The following simple analogy which presents domain-specific knowledge to do with medium components, gene markers and the ability of an organism to survive in a certain medium was used in this exercise.

The analogy: A certain species of a bacterium, *Bacillus sp.* is unable to grow on starch. It lacks an amylase gene which expresses the enzyme amylase that breaks starch into usable glucose. The bacterium's gene phenotype or gene marker is indicated as *Aml* $^{-}$. *Bacillus sp. Aml* $^{-}$ is transformed with a plasmid expressing an amylase gene. The transformant acquires the ability to break starch to glucose since it expresses the amylase gene. The transformant *Aml*$^{+}$ therefore survives in the starch medium. The transformant's ability to survive may be inter-related as follows:



After exposure to the analogy and reasoning process (Treagust *et al*., 1992; Orgill and Bodner, 2007), learners were also asked to reflect on their key study questions (Duchovic, 1998) and answers and submit a short essay explaining, the survival or death of the cells implicated in the hybridoma fusion technique, taking into account, the integrated role of HAT medium components, the cells' phenotype or gene markers and biochemical pathways (enzymology) affecting nucleotide metabolism in these cells. The

analogy was structured to promote reasoning (de Bono, 1967; Pata and Sarapuu, 2006) using a recombination of domain-specific knowledge in predicting cell survival.

An examination of the essays revealed that the integrated nature of domain-specific knowledge was considered in a holistic manner, followed by a finer reflection on critical information (Huitt, 1998; Mayer, 2003) from specific knowledge domains which enabled deductive or analytical reasoning (McMillan and Schumacher, 1993). The answers also reflected a reduction in difficulties belonging to categories (i), (ii), (iii) and (iv) although this was not quantified at this stage. A quantitative post-remediation assessment (8.2.5) followed approximately 4 weeks later where the exact probe (8.2.2) appeared in a re-test or "examination repechage" exercise as suggested by Duchovic (1998). Some of the students were also interviewed after the examination to establish personal difficulties encountered with the hybridoma fusion technique and to test their ability to answer other questions on the hybridoma fusion technique. The progress of the participants in the remediation exercise was followed to evaluate the potential of the remediation model and the extent of knowledge subsumption or integration of information into mental schemata.

8.2.5 Post-remediation assessment

8.2.5.1 Re-test

A re-test or "examination repechage" (Duchovic, 1998), involving the use of the same written probe (8.2.2), yielded encouraging results (8.2.5.3). The majority of students showed an improvement in their understanding of the hybridoma fusion technique. Domain-specific knowledge was better defined and students appeared to be more proficient at integrating factual information on markers, biochemical pathways and media components. Unfortunately, the difficulties were not eliminated totally among the group n_5 students. In comparison with the incidences recorded at pre-remediation (8.2.3.2), the incidence of difficulties belonging categories (i), (ii), (iii), (iv) and (v) showed decreased values of 5/30; 2/30; 10/30; 9/30 and 10/30, respectively. The author correlates the

reduction in difficulties with each remediation intervention and provides qualitative data which supports the use of the intervention.

The specific instruction emphasizing the use of full names at first mention and the need for a key that defined abbreviations were followed by 83 percent of the students. Learners provided the definition of all abbreviations in parenthesis at first mention. Some examples included "Hgprt (hypoxanthine guanine phosphoribosyltransferase)", "Tk (thymidine kinase)", "Ts (thymidylate synthase)" "HX (hypoxanthine)", "PRPP (phosphoribosylpyrophosphate)", "PEG" (polyethylene glycol"), "DHF (dihydrofolate)", and "am (aminopterin)". With regard to the category (ii) difficulty concerning spelling, only two students showed spelling errors. One of them spelt "hypoxanthine" in the form "hypozanthene" and the other spelt 'thymidylate synthase" in the form "thymidinylate synthase". Learners were referred to specific knowledge domains and glossaries in prescribed textbooks to assist with corrections and the meaning of words. Category (iii) difficulties concerned the erroneous substitution of nomenclature and related facts. Students were subject to challenge exercises where questions were designed to evoke a thought process that differentiates information. This strategy was less effective in alleviating the problem as 33 percent (n = 10/33) of the students continued to erroneously substitute nomenclature and equation-related detail. Previously, 50 percent of the students showed the category (iii) difficulty. Seven students erroneously substituted the enzyme thymidine kinase in the methylation reaction involving the conversion of dUMP to dTMP (3.5.6.2) while 3 other students gave the following incorrect information on the purine salvage pathways (3.5.6.2):

i. "Hypoxanthine guanine phosphoribosylpyrophosphate transferase converts HX to GMP using PRPP".
ii. "Hgprt (Hypoxanthine guanine pyrophosphate transferase) forms IMP using pyrophosphate and hypoxanthine or GMP using pyrophosphate and guanosine ".
iii. "IMP can be converted to first to AMP and then to hypoxanthine".

Students showed a better understanding of the concept of a marker as reflected by $Hgprt^{+}\ Tk^{-}$ or $Hgprt^{-}\ Tk^{+}$. The category (iv) difficulty with the concept of a marker

dropped from 63 to 30 percent as students were guided to work with protoconcepts from specific knowledge domains. In addition, the students were challenged to provide information on markers in a take-home assignment which posed several questions on markers. Four students wrongly indicated the markers $Hgprt^+$ and Tk^+ to represent the enzymes which facilitate the salvage reactions (3.5.6.2). Five other students wrongly indicated that the salvage enzymes would be found or not found in the marker gene itself. The superscripts "+" or "-" as shown by the marker notation were erroneously associated with the "presence or absence of the enzyme in the marker", respectively. Student expressions illustrating similar difficulties have already been shown (8.2.3.2). The category (v) difficulty concerning poor knowledge integration and reasoning difficulties associated with the hybridoma selection showed a pre-remediation incidence of 90 percent (n = 27/30). The guidance to learn from specific knowledge domains, the assignment of study questions and use of an analogy helped reduce the incidence to 33% (n =10/30). Despite the use of these interventions, some examples of the students' poor understanding of hybridoma selection can be seen from the following statements:

i. "The hybridoma will survive in HAT medium because it has the salvage enzymes in its genes to make purines and pyrimidines. But aminopterin in HAT medium can inhibit these enzymes as they are released by the genes to save the hybridoma from dying".

ii. "B cells ($Hgprt^+$ Tk^-) will not survive in HAT because they cannot produce thymidine kinase to make dTMP from dUMP and aminopterin in HAT medium blocks TMP formation by thymidine kinase. The hybridoma produces both hgprt and thymidine kinase therefore it lives in HAT medium".

Clearly the learning of conceptual knowledge from different representations of symbolism can pose a challenge to some students. It is likely that the learning of conceptual knowledge from symbolic representations imposes varying cognitive demands on different learners (Kozma and Russell, 1997; Kirschner, 2002; Kozma, 2003). Some learners show an understanding and the ability to express knowledge correctly when

assisted by a remedial exercise. Other students find the concepts difficult to understand and could well be affected by uneasy processing of knowledge in the LTM (1.3), a feature which characterises "germane cognitive load" (Cary, 1986; Chandler, 2004; 2.5.5).

8.2.5.2 Interviews

The following typical interview excerpts supported the fact that some students continued to have difficulties with the hybridoma fusion technique despite the adoption of specific intervention measures (8.2.4.1) during the remediation exercise. In addition, the interviews searched the students' ability to provide fine detail to several other questions which covered the knowledge domains they were referred to when given the original probe. The objective was to search meaningful learning and adaptation of answers to questions other than that to the original probe.

I: How did you find the test?
S35: Not too bad. I could have done better.
I: Did you manage to answer the question on hybridomas?
S35: Yes, I did. But not too well. I found it tough to re-call the information. It is complex to learn.
I: Did the other sections also present a problem?
S35: The volume of work made it difficult to do well.
I: Hope you do well enough to pass.
S35: Thank you, Sir.
I: I re-call that you had earlier trouble with the salvage pathways.
S35: Yes. I still find it difficult to remember the information on pyrimidines.
I: Okay. I will have a look at your written attempt.

Student 35 admits that it was "tough to re-call the information" and reflects on the coursework as "complex to learn". On examination of the student's script, a score of 6

out of 10 was attained for the hybridoma question. In comparison, S35's pre-remediation score was 2 out of 10. The re-test revealed recurring category (i) difficulties where typical (8.2.3.2) abbreviations were used without definition. In addition, student S35 substituted the enzyme thymidine kinase in the thymidylate synthase-driven reaction and failed to mention the inhibition of dihydrofolate reductase by aminopterin. Overall, the salvage pathway affecting pyrimidine biosynthesis was incorrectly described. This student required further consultation with domain specific knowledge and to learn the correct information. Information "tough to re-call" might be indicative of poor knowledge subsumption or integration into mental schemata (1.3.1).

The interview with student S36 tests the student's understanding of the concept of a marker and the meaning of its symbolic form in terms of expression.

I: Good to see you smiling. How was the test?
S36: I managed quite well. But it could have been a little shorter. It did demand fast writing. Yeah, no real complaints.
I: Did you answer the question on hybridomas and HAT selection?
S36: Yes. It is not as difficult as I first thought. But then, I really appreciate the help you gave me.
I: Tell me [Interviewer looks at the written effort of S36], what is the difference between Hgprt$^+$ and Hgprt ?
S36: It is the salvage enzyme Hgprt but we show a plus when the enzyme is produced.
I: What does Hgprt stand for?
S36: Hypo [pause]hypoxanthine guaninephosphoribo, its tough to remember.
I: But you did describe its full function.
S36. Yes. Yes. It is hypoxanthine guaninephosphoribosyl transferase.
I Okay. How do you show the gene encoding the enzyme?
S36: Hgprt$^+$! Like the enzyme it produces.
I: How is the enzyme produced?
S36: The gene releases it.

The student's pre-remediation score for the hybridoma question was 3.5 out of 10. This student's post-remediation score was 8 out of 10. The student showed minor, recurring difficulties belonging to categories (i) and (iv) (8.2.3.2). No definition was given for "Hgprt" in the purine salvage reaction and statements indicating that "*Hgprt*$^{+}$" (the marker) converts hypoxanthine to IMP in the presence of phosphoribosylpyrophosphate. Poor differentiation was made between gene marker and the encoded enzyme. In addition, S36's written statement, "the hybridoma *Hgprt* $^{+}$ *Tk*$^{+}$ will survive in HAT medium because it produces both salvage enzymes *Hgprt* $^{+}$ and *Tk*$^{+}$" further substantiates the claim that a category (iv) difficulty prevails in this case. This student required further remediation on the difference between gene marker and encoded enzyme as revealed by the interview. The concept of "the expression of a gene" requires re-visiting and consolidation.

The following interview with student S38 also reveals difficulties that fit categories (i), (iii), (iv) and (v) despite S38's participation in the remediation exercise.

I: Tell me, what is a marker?
S38: It is a gene that we show for a property.
I: What do you mean by this? Can you relate this to the hybridoma fusion technique?
S38: We have Hgprt $^{+}$ *Tk*$^{+}$ *which show us the salvage enzymes are with the gene.*
I: What are these enzymes? Name them, please.
S38: Hypo[pause], aai, I can't remember and thymine kinase.
I: You did mention the word salvage. Tell me more about salvage enzymes and what they do.
S38: I am sorry Sir, I had the same trouble in the test. I am confused.
I: Surely you can give me some information.
S38 : Well thymine kinase must phosphorylate thymine.
I: Okay. Let us stop here.

The author reports that S38's submission of answers to the study questions was accurate in respect of gene markers, the meaning of different notations and use in the selection

process. Also, the hybridoma selection technique was correctly explained**.** However, his pre- and post-remediation scores were 2 and 4 out of ten, respectively. Cognizance should be taken that some students may be adept at retrieving information from knowledge domains, writing it correctly but show difficulty learning the information. Knowledge subsumption is actually poor despite written submissions and a compilation of answers from textbooks and lecture notes. This interview confirms poor understanding of markers, their nomenclature, function and use in the hybridoma fusion technique.

The author also reports that many students gave interviews (data not shown) that corresponded to the quality of their written efforts, analysed during the remediation exercise and the post-remediation test. Clearly, remediation was difficult to accomplish with all students although there were indications that the current interventions might be useful (8.2.5.1; 8.2.5.3).

8.2.5.3 Comparative assessment scores

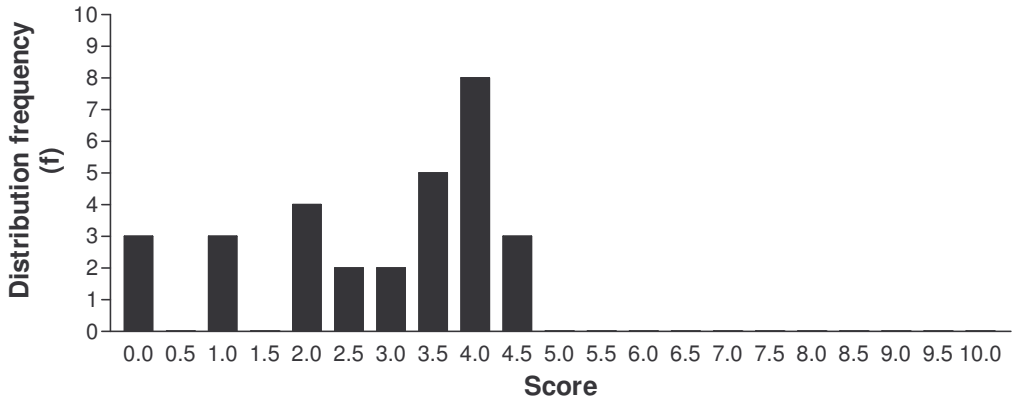
Comparative post-remediation assessment of student scores was performed exactly as described in section 8.2.3.3 using the students' marks generated in the re-test exercise. In comparison with pre-remediation scores, a mean and standard deviation of 6.45 and 1.392 were recorded, coupled with values of 6.5 and 5 for median and mode, respectively (Table 8.2). The distribution frequency (f) was asymmetrical and unimodal (Duncan *et al*., 1983), showing an increased mode value of 5 and a shift in the score dispersion range between 4 and 9 (Figure 8.3B). The distribution frequency plots were generated using GraphPad Prism Version 3.02 (GraphPad Software Inc., San Diego, CA, USA), purchased under licence and kindly loaned by Dr B. Masola, currently based at the University of KwaZulu-Natal, Department of Biochemistry, Durban Campus, SA.

Table 8.2 Raw data and statistical analyses on post-remediation assessments

Student No.	Post-remediation score	Statistical Indicator
1	6	
2	4	
3	5	**Mean:** 6.45
4	5	
5	6	
6	6	
7	4	***SD:** 1.392
8	5	
9	6	
10	8	
11	7.5	
12	7.5	**Median:** 6.5
13	8	
14	7	
15	6.5	
16	6.5	
17	5	**Mode:** 5
18	6.5	
19	5	
20	5	
21	6	**Maximum score:** 10
22	7	
23	9	
24	9	
25	8	
26	8	
27	5	
28	7	
29	7	
30	8	

*SD : Standard Deviation

A



B

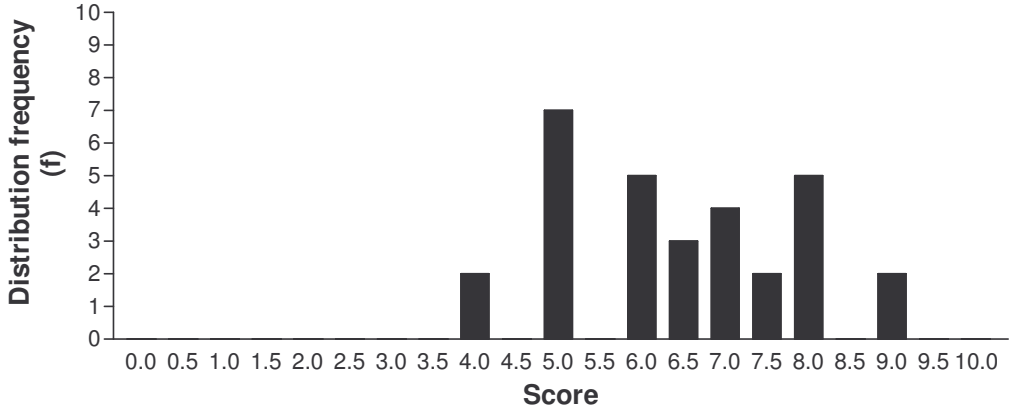


Figure 8.3 Distribution frequencies (f) of pre-remediation (A) and post-remediation (B) scores. [f = total number of students attaining a particular score]

In addition, the paired t-test was performed to establish the significance of the difference between two means recorded as "pre-remediation" and "post-remediation" values. The null (H_o) and alternative (H_1) hypotheses are formulated as follows;

H_o : The two means of the pre-remediation and post- remediation score are equal.

H_1 : The post-remediation mean score is greater than the pre-remediation mean score.

Computation (8.2.3.3) of the raw data (Tables 8.1 and 8.2) yielded a t-statistic value of 15.75571 (Table 8.3). This value exceeded the critical value of 2.04523, where $\alpha = 0.05$. Hence the H_o hypothesis concerning equality of the means is rejected. The alternative hypothesis H_1 is accepted at the 5% level of significance. In conclusion, the post-remediation mean score is significantly higher than the pre-remediation mean score. Microsoft Office Excel 2003 (Microsoft Corporation, USA)-facilitated output of statistical parameters and related values of the paired t-test are tabulated (Table 8.3).

Table 8.3 Output of statistical parameters of the computed paired t-test

	*Variable 1	*Variable 2
Mean	6.45	2.833333
Variance	1.937069	2.022989
Standard Deviation	1.391786	1.422318
Observations	30	30
Pearson Correlation	0.600969	
Hypothesized Mean	0	
Degrees of Freedom (df)	29	
t-Statistic	15.75571	
P(T‹ = t) one-tail	4.7E -16	
t-Critical one-tail	1.699127	
P (T‹ = t) two-tail	9.39E-16	
t-Critical two tail	2.04523	

*Variable 1: Post –remediation data

*Variable 2: Pre –remediation data

E: exponential

8.3 Summary and conclusion

Preliminary evaluation of the remediation model demonstrates potential in assisting learners understand the conceptual knowledge conveyed by symbolism in Molecular Biology. It identifies different categories of difficulties and offers a specific form of remedial intervention. It is of paramount importance to initially establish “what the learner knows” (Ausubel, 1968, cited by Schönborn and Anderson, 2006a) and then to offer guidance to learn (Vygotsky, 1978 [original work published in 1926], cited by Taylor, 1994) from specific knowledge domains. This strategy can prove to be useful in

lessening or eliminating inappropriate information transfer. Further, the understanding of protoconcepts (Fisher, 1985) or basic information is essential when attempts are made to understand the more complex information on a topic or symbolic representation (Sandoval and Reiser, 2004). The learning of smaller, vital amounts of domain-specific information could serve as a bridge between the learner's existing knowledge framework and what has to be learnt, thus promoting the subsumption of more complex knowledge as stated in Ausubel's assimilation theory (Ausubel, 1960; cited by Novak, 1984). It is imperative that learners reflect holistically on domain-specific knowledge, find meaningful inter-relationships, integrate or link information from relevant domains as required (Huitt, 1998).The meaningful understanding of such knowledge enables deductive or analytical reasoning (McMillan and Schumacher, 1993; Pata and Sarapuu, 2006). The current remediation model (Figure 8.1) allows for a heuristic (Novak, 1984), metacognitive learning process (Guterman, 2003; Georghiades, 2004) which has already demonstrated improvement in the students' scores. Instructor-student interactions (Darby, 2005; Oh, 2005) are very important in formative assessments (Black and William, 2006) where careful attention should be directed to specific learning difficulties, some less obvious to both the instructor and learner while others could manifest rather overtly. In this study, the author draws attention to a range of recurring learning difficulties (Chapters 4 to 8) which show low incidence and should not be ignored in remedial exercises.

CHAPTER 9

GENERAL DISCUSSION

The Introduction of this thesis poses four research questions. The answers are revealed succinctly by the author as follows, having described the detailed findings of this study in the various chapters.

1. **What is the nature of symbolic language used in Molecular Biology?** Symbolic language entails a visual communication of information depicted by, *inter alia*, drawings, diagrams, models, graphs, machine-generated images, molecular structures, formulae, abbreviations, animations, video and computer images as shown in Chapter 2. The content analysis presented in this study reveals that the expression of symbolic language can be diverse and rather complex, affecting many entities, processes and phenomena in Molecular Biology. Symbolic language is indeed a challenging task to decipher as Molecular Biology entails the study of the macroscopic, microscopic and sub-microscopic or molecular elements of nature (Treagust *et al*., 2003; Takayama, 2005). Symbolic representations of these elements, especially the sub-microscopic or molecular, tend to enhance the mental visualization of entities, processes and phenomena associated with life. The interpretation or visual communication of symbolic representations requires an understanding of the underlying information depicted by such symbolism (Kozma, 2003; Gilbert, 2005). Current investigations described in this thesis indicate explicitly that students must be taught the meaning of symbolic language apart from the conceptual knowledge represented by various forms of symbolism. It is a daunting task to teach students the skills (Lord, 1985; Richardson and Richardson, 2002; Wu and Shah, 2004) required to interpret a diverse range of symbolic representations which depict conceptual knowledge in Molecular Biology. Unfortunately, experts tend to introduce great variation in symbolic representations which describe existing or established knowledge in Molecular Biology. It is likely that new discoveries could also be

represented in different, previously unseen formats hence this could place serious cognitive demands on the teaching and learning of new information in Molecular Biology (Kozma and Russell, 1997; Schnotz and Bannert, 2003; Lewalter, 2003).

2. **To what extent do Biochemistry students find the symbolic and visual language used in Molecular Biology difficult to understand?**

   Indeed symbolic and visual language can be difficult to a great extent for students to understand as diverse forms of symbolism are used in Molecular Biology (Chapter 2) and the underlying scientific information that they depict is complex. The depth of information required from students fits the level of study. As shown in Chapters 4 to 8 of this study, the information conveyed by symbolic representations in Molecular Biology can be difficult to communicate when the underlying information concerning these depictions are not fully understood (Kozma, 2003). Learning from symbolic representations requires the ability of learners to skilfully analyse the mode or form of representation, coupled with reasoning and understanding of the conceptual information they depict. The author emphasises the understanding of domain-specific (Wandersee *et al*., 1994) or threshold information (Meyer and Land, 2003) associated with various forms of symbolism. It is essential that protoconcepts (Fisher, 1985) be understood as these underpin the more complex information represented by various forms of symbolism. The lack of standardization in the use of symbolic representations also demands learning from different symbolic formats (Seufert, 2003). Students are exposed to such incredible diversity in this regard that learning from diverse symbolic representations does become a cognitively demanding task. Equally important is the information that learners bring to the fore (Cheng *et al*., 2001) when analysing symbolic representations. This is a form of bootstrapping (Cheng *et al*., 2001) which makes learning from symbolic representations more meaningful.

3. **If so, what is the nature of such difficulties, and**

   Inappropriate or erroneous information transfer, rote learning, surface-level understanding, weak epistemic scaffolding, constructivism, formation of alternate mental schemata, poor reasoning or differentiating ability, a lack of lateral thinking, poor mental imagery or visualization were commonly encountered using symbolic representations in the teaching of Molecular Biology (Chapters 4 to 7).

4. **What steps could be taken to remediate the difficulties for improved understanding and learning of Molecular Biology?**

   The identification and assessment of learning difficulties can be analysed from both verbal and written expressions. It is important to establish the extent of prior learning and the depth of information that a learner knows. This enables the offer of appropriate remedial assistance which should ideally culminate in metacognitive awareness (Georghiadis, 2000; 2004), critical thinking (Kogut, 1996) and the promotion of further learning. The identification of critical (Huitt, 1998; Mayer, 2003) or threshold knowledge (Meyer and Land, 2003; Cousin, 2006) that must be attained in order to promote conceptual understanding is vital in any remediation exercise. Protoconcepts which underpin supraordinate information (Novak, 1984; Herron,1996) must be understood, as explained by principles of epistemic scaffolding (Bruner, 1973, cited by Cooper, 2005; Sandoval and Reiser, 2004), or else the ability to understand the apparently more complex information presents with difficulty (Kirschner, 2002; Schnotz and Bannert, 2003). Re-visiting the underlying information depicted by symbolism may be beneficial to consolidate understanding and the development of skill in decoding symbolism. The communication of factual information depends critically on such an understanding as shown by the remediation exercise in this study. The development of mental models concerning the sub-microscopic element is important especially when learners are challenged to demonstrate their understanding of the unseen. Symbolic representations are simply an abridged form of illustration; however, the meaning and depth of molecular information they depict about the sub-microscopic must be revealed clearly by learners.

Teaching such information in a lucid manner is currently a great challenge to many teachers.

The current investigations provide unusual qualitative data, indicating new evidence of student difficulties with the following topics (bulleted) and particular implications for the teaching of the symbolic language of Molecular Biology.

- **Polysemic symbols "Δ" and ": :" and their meaning in various scientific contexts**

There are indications that some students define such symbols in an idiosyncratic manner, drawing from daily life experiences (Kasanda *et al*., 2005) and relating definitions erroneously. As these symbols have particular meanings in specific contexts, these ought to be stressed by the teacher and differentiated clearly when used in different contexts. As explained in this study, the symbol, "**::**" can be used to show in tandem gene fusion and transposition.

- **The differentiation of the phosphite and phosphodiester groups in oligonucleotides on the basis of valency and the role of an established shorthand notation used to show the 3,5 phosphodiester link between nucleotides**

Shorthand notations present an abridged format representing more complex information. The interpretation of notations, such as the form used to differentiate the phosphite and phosphodiester groups in oligonucleotides, requires an understanding of valency and VSEPR theory to differentiate such groups. It is imperative that teachers present this theory lucidly and ascertain the students' understanding of it prior to the use of the shorthand notation. As shown in this study, students do not understand the molecular basis of bonding and simply count "lines" associated with "letters" to explain bonding ability or valency of an atom (Barnea and Dori, 1996; Bowen, 1998; Barak and Dori, 2005).

- **Symbolic plasmid form, functional plasmid domains and gene expression**

Conformational plasmid forms and their functional domains are indicated by a range of symbolism (Chapter 3). Teaching varied plasmid structure requires an integrated understanding of the molecular structure of DNA, the factors that induce different conformations and the symbolic representation of this information. Shorthand notation, diagrams or maps are commonly used in the teaching of plasmid structure and function. Conformational plasmid forms and their functional domains require differentiation, an explanation of their role in a host or transformant and an emphasis of processes associated with gene expression which is linked to specific plasmid domains. It is a common misconception among students that "genes contain the entity they encode". Transcription and translation of plasmid-borne genes are generally forgotten as processes linked to "gene expression" (Rotbain *et al*., 2006). These concepts must be emphasised during teaching.

- **Practical aspects of restriction mapping**

Restriction mapping presents a challenge to students mainly in terms of procedural knowledge or their understanding of the practical aspects of mapping. The teaching of relevant theory that assists in the understanding of the technique is essential. This will include DNA structure, properties of restriction endonucleases, cleavage conditions and patterns, electrophoresis and visualization of restricted DNA. Problem-solving exercises can promote reasoning ability (Pata and Sarapuu, 2006) and will assist students in the construction of restriction maps (Szerberényi, 2002; Walsh *et al*., 2007) using fragment size and overlap patterns derived from an electrophoretogram.

- **Gene markers and phenotypic expression**

Symbolic notations of gene markers can also pose a challenge to learn as shown in this study. It must be differentiated from marker DNA which some students associate with nucleic acid size determination in an electrophoresis run. Phenotypic expression tends to be an elusive concept as students can believe that visible traits such as eye colour or gene-encoded entities are found in the gene themselves. Teachers should regard concepts such as "gene markers" and "phenotypic expression" as important protoconcepts (Fisher,

1985) which require clear definition and linkage to show how gene markers influence or determine phenotypic expression.

- **The selection of hybridomas and transformants**

The understanding of protoconcepts such as gene markers and phenotypic expression is essential when trying to explain the selection of hybridomas and transformants. The interplay of symbolic representations takes the form of gene markers, media components and biochemical pathways when explaining the selection of hybridomas. Likewise transformants can express both genome- or plasmid-encoded traits which affect survival or selection in certain media. Students may be introduced to simpler analogies involving such interplay of factors; however, it is essential to promote the ability to work with domain-specific information to do with gene markers, biochemical pathways or media constituents. A recombination of such information should ideally promote reasoning with regard to the influential factors affecting cell selection.

- **Nucleic acid-strand nomenclature and function**

As shown in this study, nucleic acid strand nomenclature can be complicated owing to the use of synonyms and a lack of standardisation in this regard. This can be potentially confusing to the learner. Some students tend to reflect on the labelling of DNA replication intermediates and inappropriately transfer such labels when labelling nucleic acid strands. Teaching strand nomenclature requires an integration of function which clearly differentiates each label (Cook *et al*., 2006; Rotbain *et al*., 2006). The use of diagrams or models which show such labels and corresponding strand function (Figure 3.4) makes the learning of nucleic acid strand nomenclature more meaningful.

- **The mechanism of DNA replication**

DNA replication is a complex process, often depicted in textbooks by sequential steps involving structural change to the molecule. The process may be taught using well-labelled diagrams or models (Patrick *et al*., 2005; Cook *et al*., 2005; Fossey and Hancock, 2005; Beltramini *et al*., 2006; Rotbain *et al*., 2006); however, it is of paramount importance that learners are able to explain the molecular mechanisms associated with

the process. Certainly symbolic representations may enhance the visual perspective (Gilbert, 2005) about the process but it is essential that learners reflect on the underlying information that explains the structural change to the molecule. Teaching requires an emphasis on the enzymology and auxiliary protein interaction which facilitate the transition of structural intermediates from one form to another.

The author places great emphasis on the meaningful learning of domain-specific knowledge (Wandersee *et al*., 1994) to effect conceptual development and understanding. Protoconcepts (Fisher, 1985) which underpin the more complex information must be well-understood. Remediation may be difficult to achieve (Chapter 8) but requires careful assessment of the range of difficulties that symbolism presents when used to convey conceptual knowledge. As Molecular Biology advances, it is likely that experts will continue to use new and diverse forms of symbolic representations to explain their findings. The explanation of futuristic Science is likely to develop a symbolic language that will impose great teaching challenges and unimaginable learning difficulties to new generation teachers and learners, respectively. Our future work on symbolism will focus on the development of methods to improve the understanding of related pedagogical content knowledge. This will also entail the development of courses on symbol use and integration of tutorials during teaching. May the current work be a useful reference to those researchers who wish to pursue the characterisation of learning difficulties associated with new forms of symbolism and offer remedial assistance to learners.

REFERENCES

Ahmed, A., Sesti, F., Ilan, N., Shih, T.M., Sturley, S.L. and Goldstein, A.N. (1999) A molecular target for viral killer toxin: *TOK1* potassium channels. Cells, **99**, 283-291.

Akers, H.A. and Smith, E. (1987) Problems with phosphate abbreviations. Biochemical Education, **15**, (2), 78-79.

Alexander, P.A. (2007) Bridging cognition and socioculturalism within conceptual change research: Unnecessary foray or unachievable feat? Educational Psychologist, **42**, (1), 67-73.

Ametler, J. and Pinto´, R. (2002) Students' reading of innovative images of energy at secondary school level. International Journal of Science Education, **24**, (3), 285-312.

Amir, R. and Tamir, P. (1994) Detailed analysis of misconceptions as a factor for developing remedial instruction: the case of photosynthesis. The American Biology Teacher, **56**, (2), 94-100.

Anderson, T.R. and Grayson, D.J. (1994) Improving students' understanding of carbohydrate metabolism in first-year Biochemistry at tertiary level. Research in Science Education, **24**, 1-10.

Anderson, L.W. and Krathwohl, D.R. (Editors) (2001) A taxonomy for learning, teaching and assessing. A revision of Blooms' Taxonomy of Educational Objectives. Complete Edition, New York, Longman.

Anderson, T.R. and McKenzie, J.M. (2007) Conceptual and Reasoning Difficulties (CARD) Resource. In: http:/www.card.unp.ac.za, Accessed: September, 2007.

Ariatti, M. (2007) Personal communication. University of KwaZulu-Natal, South Africa.

Armstrong, F.B. (1989) Biochemistry, 3rd ed., Oxford University Press, UK.

Armstrong, K.A., Som, T., Volkert, F.C., Rose, A. and Broach, J.R. (1989) Propagation and expression of genes in yeast using two micron circle vectors. In: Barr *et al*., (eds), Yeast Genetic Engineering, Butterworth Publishers, Chapter 9, pp.165-192.

Atherton, J.S. (2003a) Learning and teaching: Learning curve [On-Line] UK: http://dmu.ac.uk./~jamesa/learning/lerncrv.htm, Accessed: October 2004.

Atherton, J.S. (2003b) Learning and teaching: Bloom's Taxonomy [On-Line] UK: http://dmu.ac.uk/~jamesa/learning/bloomtax.htm, Accessed: October 2004.

Ausubel, D.P. (1960) The use of advance organisers in learning and retention of meaningful information. Journal of Educational Psychology, **51**, 26.

Ausubel, D.P. (1968) Educational Psychology: A cognitive view. Holt, Rinehart and Wilson, New York.

Bahl, C.P., Wu, R., Brousseau, R., Sood, A.K., Hsiung, H.M. and Narang, S.A. (1978) Chemical synthesis of versatile adaptors for molecular cloning. Biochem. Biophys. Res. Comm., **81**, 695-703.

Baker, T.A. and Kornberg, A. (1988) Transcriptional activation of initiation of replication from the *E.coli* chromosomal origin: An RNA-DNA hybrid near oriC. Cell, **55**, 113-123.

Baker, T.A., Funnell, B.E. and Kornberg, A. (1987) Helicase action of dnaB protein during replication from the *Escherichia coli* chromosomal origin in vitro. Journal of Biological Chemistry, **262**, 6877-6885.

Baker, T.A., Sekimizu, K., Funnell, B.E. and Kornberg, A. (1986) Extensive unwinding of the plasmid template during staged enzymatic initiation of DNA replication from the origin of *Escherichia coli* chromosome. Cell, **45**, 53-64.

Banet, E. and Ayuso, E. (2000) Teaching genetics at secondary school. A strategy for teaching about location of inheritance information. Science Education, **84**, (3), 313-351.

Barak, M. and Dori, Y.J. (2005) Enhancing undergraduate students' Chemistry understanding through project-based learning in an IT environment. Science Education, **89**, 117-139.

Barayani, J. and Pin, C. (2001) A parallel study on bacterial growth and inactivation. Journal of Theoretical Biology, **210**, 327-336.

Barlex, D. and Carre, C. (1985) Visual communication in Science. Cambridge Science Education Series, Cambridge University Press, London, pp.1-103.

Barnea, N. and Dori, Y.J. (1996) Computerized molecular modelling as a tool to improve Chemistry teaching. J. Chem. Inf. Comput. Sci., **36**, 629-636.

Barr, P.J., Blake, A.J. and Valenzuela, P. (1989) Yeast Genetic Engineering, Butterworth Publishers, MA, USA.

Bechtel, W. (1998) Representations and cognitive explanations: Assessing the dynamicists' challenge in Cognitive Science. Cognitive Science, **22**, (3), 295-318.

Beckwith, J. (2005) Harvard Medical School, USA. personal communication.

Beltramini, L.M., Araujo, A.P.U., de Oliviera, T.H.G., dos Santos Abel, L.D., da Silva, A.R. and dos Santos, N.F. (2006) A new three-dimensional educational model kit for building DNA and RNA molecules. Biochemistry and Molecular Biology Education, **34**, (3), 187-193.

Birk, J.P. and Kurtz, M.J. (1999) Effect of experience on retention and elimination of misconceptions about molecular structure and bonding. Journal of Chemical Education, **76**, (1), 124-128.

Black, P. and William, D. (2006) Inside the Black Box: Raising standards through classroom assessments. Electronic version from Phi Delta Kappa International, In: http://www.pdkintl.org/kappen/kbla9810.htm. Accessed: June 2006.

Bloom, B.S. (1956) Ed. Taxonomy of educational objectives, the classification of educational goals- Handbook 1, Cognitive Domain, Longmans, Green, Toronto.

Bodemer, D., Ploetzner, R., Feuerlein, I. and Spada, H. (2004) The active integration of information during learning with dynamic and interactive visualizations. Learning and Instruction, **14**, 325-341.

Bohinski, R.C. (1987) Modern Concepts n Biochemistry. Allyn and Bacon Inc., MA.

Bolivar, F., Rodriguez, R.L., Greene, P.J., Betlach, M.C., Heyneker, H.L., Boyer, H.W., Crosa, J.H. and Falkow, S. (1977) Construction and characterization of new cloning vehicles. 2. Multi-purpose cloning system. Gene, **2**, (2), 95-113.

Botha, P.N. (1991) A neurological, physiological and genetic orientation. In: Kapp, J.A. (ed.). Children with problems – an orthopedagogical perspective. J.L. van Schaik, Pty (Ltd) Publishers, Pretoria, SA, Chapter 12, pp. 201-255.

Bowen, C.W. (1998) Item design considerations for computer-based testing of student learning in Chemistry. Journal of Chemical Education, **75**, (9), 1172-1174.

Brown, T.A. (1986) Gene cloning: An introduction. Van Nostrand Reinhold (UK) Co. Ltd., UK.

Bruner, J. (1960) The process of Education. Harvard University Press, Cambridge, MA

Bruner, J. (1973) Going beyond the information given. Norton, NY.

Cairns, J. (1963a) The bacterial chromosome and its manner of replication as seen by autoradiography. Journal of Molecular Biology, **6**, 208-213.

Cairns, J. (1963b) The chromosome of *Escherichia coli*. Cold Spring Harbor Symp. Quant. Biol., **28**, 43-46.

Caldwell, B., Rohlman, C. and Benore-Parsons, M. (2004) A curriculum skills matrix for the development and assessment of undergraduate Biochemistry and Molecular Biology laboratory programs. Biochemistry and Molecular Biology Education, **32**, (1), 11-16.

Campbell, M.K. (1995) Biochemistry, 2nd ed., Saunders College Publishing/Harcourt Brace College Publishers, New York.

Campbell, P.N. and Smith, A.D. (1993) Biochemistry Illustrated, 3rd ed., Churchill Livingstone, UK.

Caramazza, A., McCloskey, M. and Green, B. (1981) Naive beliefs in sophisticated subjects: Misconceptions about trajectories of objects. Cognition, **9**, 117-123.

CARD : http://www.unp.ac.za/generalcon.asp, University of KwaZulu-Natal, Science Education Research Group, Pietermaritzburg, SA..

Carey, S. (1986) Cognitive Science and Science Education. American Psychologist, **41**, (10), 1123-1130.

Carusi, E.A. (1992) Its time we replaced 'high energy phosphate group' with 'phosphoryl group'. Biochemical Education, **20**, (3), 145-147.

Chandler, P. (2004) The crucial role of cognitive processes in the design of dynamic visualizations. Learning and Instruction, **14**, 353-357.

Chapman, B.S. (2001) Emphasizing concepts and reasoning skills in introductory college molecular cell biology. International Journal of Science Education, **23**, (11), 1157-1176.

Chen, A, and Ennis, C.D. (1995) Content knowledge transformation: Relationship between content knowledge and curricula. Teaching and Teacher Education, **11**, (4), 389-401.

Chen, S.Y. and Raffan, J. (1999) Biotechnology: students' knowledge and attitudes in the UK and Taiwan. Journal of Biological Education, **34**, 17-23.

Cheng, P.C.-H., Lowe, R.K. and Scaife, M.(2001) Cognitive science approaches to understanding diagrammatic representation. Artificial Intelligence Review, **15**, 79-94.

Chi, M.T.H., Feltovich, J. and Glaser, R. (1981) Categorization and representation of Physics problems by experts and novices. Cognitive Science, **5**, 121-152.

Chiras, D.D. (1992) Teaching critical thinking skills in the Biology and Environmental Science classrooms. The American Biology Teacher, **54**, (8), 464-468.

Cho, H-H., Butler-Kahle, J. and Nordland, F.H. (1985) An investigation of high school biology textbooks as sources of misconceptions and difficulties in Genetics and some suggestions for teaching Genetics. Science Education, **69**, (5), 707-719.

Christopherson, J.T. (1997) The growing need for visual literacy at the university. Paper presented at the Visionquest: Journeys toward visual literacy. 28th Annual Conference of the International Visual Literacy Association, Cheyenne,WY.

Cohen, L., Manion, L. and Morrison, K. (2000) Meta-analysis in educational research, In: Research Methods in Education, 5th ed., Routledge Falmer, London, pp. 220-225.

Cohen, S.N., Chang, A.C.Y., Boyer, H.W. and Helling, R.B. (1973) Construction of biologically functional bacterial plasmids in vitro. Proc. Natl. Acad. Sci. USA., **70**, 3240-3244.

Colin, P. (2002) Reading images in Optics: Students' difficulties and teachers' views. International Journal of Science Education, **24**, (3), 313-332.

Collis, K.F., Jones, B.L., Sprod, T., Watson, J.M. and Fraser, S.P. (1998) Mapping development in students' understanding of vision using a cognitive model. International Journal of Science Education, **20**, (1), 45-66.

Conn, E.E., Stumpf, P.K., Bruening, G. and Doi, R.H. (1987) Outlines of Biochemistry, 5th ed., John Wiley and Sons, NY.

Cook, P.M., Carter, G. and Wiebe, E.N. (2006) Visual representations of DNA: A comparison of salient features for experts and novices, Proceedings of the NARST 2006 Annual Meeting, April 3-6, San Francisco, CA, USA. (8 pages).

Cooper, G.M. and Hausman, R.E. (2004) The Cell: A Molecular Approach, 3rd edition, ASM Press, Washington, D.C.

Cooper, S. (2005) Jerome Bruner (1915-) Constructivism and discovery learning. http://www.konnections.net/lifecircles/Bruner.htm, Accessed: May 2005.

Cousin, G. (2006) An introduction to threshold concepts. Planet, December 2006, (17), 4-5.

Crenshaw III, E.B., Russo, A.F., Swanson, L.W. and Rosenfeld, M.G. (1987) Neuron-specific alternative RNA processing in transgenic mice expressing a metallothionein-calcitonin fusion gene, Cell, **49**, 389-398.

Crisp, V. and Sweiry, E. (2006) Can a picture ruin a thousand words? The effects of visual resources in exam questions. Educational Research, **48**, (2), 139-154.

.

Dagher, Z. (1995) Review of studies on the effectiveness of instructional analogies in science teaching. Science Education, **79**, (3), 295-312.

Danna, K. and Nathans, D. (1971) Specific cleavage of SV40 DNA by restriction endonucleases of *Haemophilus influenzae*. Proc. Natl. Acad. Sci. USA., **68**, (12), 2913-2917.

Danna, K., Sacks, G.H. Jr. and Nathans, D. (1973) Studies of SV40 DNA. vii. A cleavage map of the SV40 genome. Journal of Molecular Biology, **78**, (2), 363- 376.

Darby, L. (2005) Science students' perceptions of engaging pedagogy. Research in Science Education, **35**, (4), 425-445.

Dawson, V. and Schibeci, R. (2003) Western Australian school students' understanding of Biotechnology. International Journal of Science Education, **25**, (1), 57-69.

DebBurman, S.K. (2002) Learning how scientists work: experiential research projects to promote cell biology learning and scientific process skills. Cell Biology Education, **1**, 154-172.

de Bono, E. (1967) New Think: The use of lateral thinking in the generation of new ideas. Basic Books, New York.

de Jong, O. (2000) Crossing the borders: chemical education research and teaching practice. University Chemistry Education, **4**, (1), 29-32.

Demerec, M., Adelberg, E., Clark, A. and Hartman, P. (1966) A proposal for a uniform nomenclature in bacterial genetics, Genetics, **54**, (1), 61-76.

Demerec, M., Adelberg, E., Clark, A. and Hartman, P. (1968) A proposal for a uniform nomenclature in bacterial genetics. Journal of General Microbiology, **50**, (1), 1-14.
diSessa, A. (1993) Toward an epistemology of Physics. Cognition and Instruction, **10**, 105-225.

Dressler, D. and Wolfson, J. (1967) The rolling circle for ΦX174 DNA replication. III. Synthesis of supercoiled duplex rings. Proc. Natl. Acad. Sci. USA., **67**, (1), 456-463.

Driver, R. and Bell, B. (1986) Students' thinking and the learning of Science: A constructivist view. School Science Review, **67**, 443-456.

Du, Q.(1999) Ribozyme Enzymology.
In:http://academic.brooklyn.cuny.edu/chem./zhuang/QD/toppage1.htm, Accessed June, 2007.

Duchovic, R.J. (1998) Teaching college General Chemistry: Techniques designed to communicate a conceptual framework. Journal of Chemical Education, **75**, (7), 856-857.

Duncan, R., Knapp, R.G. and Miller III, M.C. (1983) Introductory Biostatistics for the Health Sciences, 2nd edition, John Wiley and Sons, New York.

Duit, R.(2007) Bibliography-STCSE, Students' and teachers' conceptions and science education. In: http://www.ipm.uni-kiel.de/aktuell/stce/stce.html, Accessed: August 2007.

Duit, R. and Treagust, D.F. (2003) Conceptual change: a powerful framework for improving science teaching and learning. International Journal of Science Education, **25**, (6), 671-688.

Dutkiewicz, T. (1982) The symbolism of a chemical reaction and the teaching of Biochemistry. Biochemistry Education, **10**, (1), 25.

du Toit, L. (1997) Education for all. University of South Africa Press, OSN 421-Q/98-2000, p.36.

Dykxhoorn, D.M. and Lieberman, J. (2006) Silencing viral infection. Research in Translation: PLoS Medicine, **3**, (7) e242. In: http:// medicine.plosjournals.org/per/serv/?request=get-document&doi=10.1371/journal.pmed.0030242&ct=1, Accessed August 2007.

Eisenberg, S., Griffith, J. and Kornberg, A. (1977) ΦX174 cistron A protein is a multifunctional enzyme in DNA replication. Proc. Natl. Acad. Sci. USA., **74**, (8), 3189-3202.

Elliot, W.H. and Elliot, D.C. (2005) Biochemistry and Molecular Biology, 3rd edition, Oxford University Press, UK.

Feinman, R.D. (2004) Oxidation - reduction calculations in the Biochemistry course. Biochemistry and Molecular Biology Education, **32**, (3), 161-166.

Ferk, V., Vrtacnik, M. and Blejec, A. (2003) Students' understanding of molecular structure representations. International Journal of Science Education, **25**, (10), 1227-1245.

Fisher, K. (1992) Improving high school genetics instruction. In: Smith, M.U. and Simmons, P.E. (eds.), Teaching Genetics: Recommendations and research proceedings of a national conference, Cambridge, MA., USA, 18-21 March 1992, pp.24-28.

Fisher, K.M. (1985) A misconception in Biology : amino acids and translation. Journal of Research in Science Teaching, **22**, 53-62.

Fisher, K.M., Wandersee, J.H. and Moody, D.E. (2000) Student misconceptions in Biology. In: Fisher, K.M., Wandersee, D.E. and Moody, D.E. (eds.), Mapping Biology Knowledge, Kluwer Academic Publishers, Dordrecht, pp.55-75.

Fossey, A. and Hancock, C. (2005) DNA replication and transcription: An innovative teaching strategy. Biochemistry and Molecular Biology Education, **33**, (6), 411-415.

Fox, M. (1996) Breaking the genetic code in a letter by Max Delbruck. The Journal of College Science, March / April, 324-325.

Funnell, B.E., Baker, T.A. and Kornberg, A. (1986) Complete enzymatic replication of plasmids containing the origin of the *Escherichia coli* chromosome. Journal of Biological Chemistry, **261**, 5616-5624.

Funnell, B.E., Baker, T.A. and Kornberg, A. (1987) In vitro assembly of a prepriming complex at the origin of the *Escherichia coli* chromosome. Journal of Biological Chemistry, **262**, 10327-10334.

Garcia-Vallve, S. (2004) Contribution of each complex of the mitochondrial respiratory chain in the generation of the proton-motive force. Biochemistry and Molecular Biology Education, **32**, (1), 17-19.

Gardner, E.J., Simmons, M.J. and Snustad, D.P. (1984) Genetic engineering and the future. In: Principles of Genetics, 8th ed., John Wiley and Sons, NY, Chapter 24, pp.604-649.

Garlikov, R. (2003) The Socratic Method: Teaching by asking instead of by telling. http://www.garlikov.com/Soc_Meth.html, Accessed: 17 May 2005.

Garnett, P.J. and Hackling, M.W. (1995) Students' alternative conceptions in Chemistry: A review of research and implications for teaching and learning. Studies in Science Education, **25**, 69-95.

Garnett, P.J. and Treagust, D.F. (1992a) Conceptual difficulties experienced by senior high school students of Electrochemistry: Electric circuits and oxidation reduction equations. Journal of Research in Science Teaching, **29**, 121-142.

Garnett, P.J. and Treagust, D.F. (1992b) Conceptual difficulties experienced by senior high school students of Electrochemistry: Electrochemical (Galvanic) and electrolytic cells. Journal of Research in Science Teaching, **29**, (10) 1079-1099.

Garnett, P.J., Garnett, P.J. and Hackling, M.W. (1985) Students' alternative conceptions in Chemistry : A review of research and implications for teaching and learning. Studies in Science Education, **25**, 69-95.

Garrett, R.H. and Grisham, C.M. (1995) Biochemistry, Intl. ed., Sanders College Publishing, New York.

Garrett, R.H. and Grisham, C.M. (2005) Biochemistry, 3rd ed., Thomson Books/Cole , Belmont, CA, USA.

Geddis, A.N., Onslow, B., Beynon, C. and Oesch, J. (1993) Transforming content knowledge: Learning to teach about isotopes. Science Education, **77**, (6), 575-591.

Gene Proforma Version 33 database (2003) http://www.gmod.org/sop/curatorial/flybase/FB.gene.pro.notes, Accessed: September 2005.

Georghiades, P. (2000) Beyond conceptual change learning in Science Education: focussing on transfer, durability and metacognition. Educational Research, **42**, (2), 119-139.

Georghiades, P. (2004) From the general to the situated: three decades of metacognition. International Journal of Science Education, **26**, (3), 365-383.

Gilbert, J.K. (2005) Visualization in Science Education, Springer, The Netherlands.

Gilbert, W. and Dressler, D. (1968) DNA replication: The rolling circle model. Cold Spring Harbor Quant. Biol., **33**, 473-484.

Gilbert, J.K., Osborne, R.J. and Fensham, P. (1982) Childrens' Science and its implications for teaching. Science Education, **66**, 625-633.

Gorodetsky, M. and Keiny, S. (1995 Conceptual change and environmental cognition. International Journal of Science Education, **17**, (2), 207-217.

Goulian, M. and Kornberg, A. (1967) Enzymatic synthesis of DNA. 23. Synthesis of circular replicative form of phage ΦX174 DNA. Proc. Natl. Acad. Sci. USA, **58**, (4), 1723-1730.

Gowin, D.B. (1981) Educating, Cornell University Press, Ithaca, NY.

Grayson, D. (1996) Concept substitution: A strategy for promoting conceptual change. In: Treagust, D.F., Duit, R. and Fraser, B.J. (eds.), Teaching and learning in Science and Mathematics , Teachers College Press, NY, pp.152-161.

Grayson, D.J. (2004) A teaching strategy for helping students disentangle related Physics concepts. American Journal of Physics, **72**, (8), 1126-1133.

Grayson, D.J. (1995) Science Education research and implications for university science instruction. South African Journal of Science, **91**, 168-172.

Grayson, D.J., Anderson, T.R. and Crossley, L.G. (2001) A four-level framework for identifying and classifying students' conceptual and reasoning difficulties. International Journal of Science Education, **23**, (6), 611-622.

Greenbowe, T.J. (1994) An interactive multi-media software program for exploring electrochemical cells. Journal of Chemical Education, **71**, 555-557.

Grosslight, L., Unger, C., Jay, E. and Smith, C. (1991) Understanding models and their use in Science: Conceptions of middle and high school students and experts. Journal of Research in Science Teaching, **28**, 799-822.

Gupthar, A.S. (1983) Chemical and biochemical investigations into nucleic acids and nucleic acid components. B.Sc. Honours thesis. University of Durban-Westville, Durban, SA.

Gupthar, A.S. and Garnett, H.M. (1987) Hybridization of *Pichia stipitis* with its presumptive imperfect partner, *Candida shehatae*, Current Genetics, **12**, 199-204.

Gupthar, A.S. (1987) Construction of a series of *Pichia stipitis* strains with increased DNA contents, Current Genetics, **12**, 605-610.

Gupthar, A.S. (1989) Hybridization studies of D-xylose fermenting yeasts. Ph.D. thesis, University of the Witwatersrand, Johannesburg, SA.

Gupthar, A.S. (1992) Segregation of altered properties in fusions between *Saccharomyces cerevisiae* and the D-xylose fermenting yeasts *Candida shehatae* and *Pichia stipitis*. Canadian Journal of Microbiology, **38**, 1233-1237.

Gupthar, A.S. (1996) The use of false statement-type questions in Biochemistry tests. Biochemical Education, **24**, (2), 98.

Gupthar, A.S. and Anderson, T.R. (2003) Student difficulties associated with DNA-strand nomenclature and function. Proc. 11th Annual Conference of the Southern Association Research in Mathematics, Science and Technology Education (SAARMSTE), Putsoa *et al*., (Eds.) Volume 1, Waterford Kamhlaba United World College, Swaziland, pp.361-365.

Guterman, E. (2003) Integrating written metacognitive awareness guidance as a "psychological tool" to improve student performance. Learning and Instruction, **13**, (6), 633-651.

Hake, R. (2008) Research in Physics Education. In: http://www.physics.indiana.edu/~hake/, Accessed: March 2008.

Hameed, H., Hackling, M.W. and Garnett, P.J. (1993) Facilitating conceptual change in chemical equilibrium using a CAI strategy. International Journal of Science Education, **15**, (20), 221-230.

Hancock, C. (2006) Identification and remediation of student difficulties with quantitative genetics. Ph.D. thesis, Awarded October 2006, University of KwaZulu-Natal, Pietermaritzburg, SA.

Hand, B.M. and Treagust, D.F. (1988) Application of conceptual conflict strategy to enhance student learning of acids and bases. Research in Science Education, **18**, 53-63.

Harrison, A.G., Grayson, D.J. and Treagust, D.F. (1999) Investigating a grade 11 students' evolving conceptions of heat and temperature. Journal of Research in Science Teaching, **36**, (1), 55-87.

Hayton, W.L. (1990) Symbol-and-arrow diagrams in teaching pharmacokinetics, American Journal of Pharmaceutical Education, **54**, 290-292.

Hegarty, M. (2004) Dynamic visualization and learning: Getting to the difficult questions. Learning and Instruction, **14**, (3), 343-351.

Henderson, G. (1999) Learning with diagrams. Australian Science Teachers' Association, **15**, 17-25.

Herron, J.D. (1996) The Chemistry Classroom: Formulas for successful teaching. American Chemical Society, Washington, D.C., p.121-123.

Hewson, P.W. (1981) A conceptual change approach to learning Science. European Journal of Science Education, **3**, 383-396.

Hewson, P.W. (1982) A case study of conceptual change in special relativity: The influence of prior knowledge in learning. European Journal of Science Education, **4**, 61-78.

Hewson, P.W. and Hewson, M.G.A'B. (1984) The role of conceptual conflict in conceptual change and the design of Science instruction. Instructional Science, **13**, 1-13.

Hewson, P.W. and Thorley, N.R. (1989) The conditions of conceptual change in the classroom. International Journal of Science Education, **11** (special issue), 541-553.

Hill, R., Stanisstreet, M. and Boyes, E. (2000) What ideas do students associate with "Biotechnology" and "Genetic Engineering"? School Science Review, **81**, 77-83.

Horton, H.R., Moran, L.A., Ochs, R.S., Rawn, J.D. and Scrimgeour, K.G. (1996) Principles of Biochemistry, 2nd ed., Prentice Hall Inc., New Jersey.

Huddle, P.A., White, M.W. and Rogers, F. (2000) Using a teaching model to correct known misconceptions in Electrochemistry. Journal of Chemical Education, **77**, 104-110.

Huitt, W. (1998) Critical thinking: An overview – Educational Psychology Interactive, http://chiron.valdosta.edu.whuitt/col/cogsys/critthnk.html, Accessed: August 2005.

Hull, T.G. (2003) Students' use of diagrams for the visualization of biochemical processes. M.Sc. thesis, University of KwaZulu-Natal, Pietermaritzburg, SA.

Ikeda, J-E., Yudelevich, A. and Hurwitz, J. (1976) Isolation and characterization of the protein coded by gene A of bacteriophage.v.phiX174 DNA. Proc. Natl. Acad. Sci. USA, **78**, (8), 2669-2673.

Ikeda, J-E., Yudelevich, A., Shimamoto, N. and Hurwitz, J. (1979) Role of polymorphic forms of the bacteriophage phiX174 DNA cleavage. Journal of Biological Chemistry, **254**, (19), 9416-9428.

Jablonowski, D., Fichtner, L., Martin, V.J., Klassen, R., Meinhardt, F., Stark, M.J.R. and Schaffrath, R. (2001) *Saccharomyces cerevisiae* cell wall chitin, the *Kluyveromyce lactis* zymogen receptor. Yeast, **18**, 1285-1299.

Johnson, P. (2000) Developing students' understanding of chemical change: What should we be teaching? Chemistry Education: Research and Practice, **1**, (1), 77-90. http://www.uoi.gr/cerp

Jones, L.L., Jordan, K.D. and Stillings, N.A. (2005) Molecular visualization in Chemistry Education: The role of multidisciplinary collaboration. Chemistry Education Research and Practice, **6**, (3), 136-149.

Jones, M.G., Minogue, J., Oppewal, T., Cook, M.P. and Broadwell, B. (2006) Visualizing without vision at the microscale: Students with visual impairment explore cells with touch. Proceedings of the NARST 2006 Annual Meeting, April 3-6, San Francisco, CA, USA.

Justi, R.S. and Gilbert, J.K. (2002) Modelling teachers' views on the nature of modelling, and implications for the education of modellers. International Journal of Science Education, **24**, (4), 369-387.

Justi, R.S. and Gilbert, J.K. (2003) Teachers' views on the nature of models. International Journal of Science Education, **25**, (11), 1369-1386.

Kadonaga, J.T., Carner, K.R., Masiarzft, F.R. and Tjian, R. (1987) Isolation of cDNA encoding transcription factor sp1 and functional analysis of the DNA binding domain. Cell, **51**, 1079-1090.

Kang, S., Scharmann, L.C., Noh, T. and Koh, H. (2005) The influence of students' cognitive and motivational variables in respect of cognitive conflict and conceptual change. International Journal of Science Education, **27**, (9), 1037-1058.

Karp, G. (1996) Cell and Molecular Biology: Concepts and experiments, John Wiley and Sons, New York.

Kasanda, C., Lubben, F., Gaoseb, N., Kandjeo-Marenga, U., Kapenda, H. and Campbell, B. (2005) The role of everyday contexts in learner-centred teaching: The practice in

Namibian secondary schools. International Journal of Science Education, **27**, (15/16), 1805-1823.

Kearney, M. (2004) Classroom use of multimedia-supported predict-observe-explain tasks in a social constructivist learning environment. Research in Science Education, **34**, (4), 427-453.

Kindfield, A.C.H. (1994) Biology diagrams: Tools to think with. The Journal of the Learning Sciences, **3**, (1), 1-36.

Kinnear, J.D. (1992) Teaching Genetics: Recommendations and research. In: Smith, M.U. and Simmons, P.E. (eds.), Proceedings of a national conference, Cambridge, MA, 18-21 March, pp.44-55.

Kirschner, P.A. (2002) Cognitive load theory: Implications of cognitive load theory on the design of learning. Learning and Instruction, **12**, 1-10.

Klein, J. (1990) Immunoglobulins and their receptors. In: Immunology, Chapter 6, Blackwell Scientific Publishers, Oxford, pp.102-160.

Kogut, L.S. (1996) Critical thinking in General Chemistry. Journal of Chemical Education, **73**, (3), 218-221.

Köhler, G. and Milstein, C. (1975) Continuous cultures of fused cells secreting antibody of predefined specificity. Nature, **256**, 495-497.

Kong, X.P., Onrust, R., O' Donnell, M. and Kuriyan, J. (1992) Three dimensional structure of the beta subunit of *E.coli* DNA polymerase III holoenzyme: A sliding clamp. Cell, **69**, 425-437.

Kornberg, A. and Baker, T.A. (1992) DNA Replication, 2nd edition, W.H.Freeman and Co., USA.

Kornberg, A., Scott, J.F. and Bertsch, L.L. (1978) ATP utilization by rep protein in the catalytic separation of DNA strands at a replication fork. Journal of Biological Chemistry, **253**, (9), 3298-3304.

Kozma, R. (2003) The material features of multiple representations and their cognitive and social affordances for science understanding. Learning and Instruction, **13**, 205-226.

Kozma, R. and Russell, J. (1997) Multimedia and understanding: Expert and novice responses to different representations of Chemistry phenomena. Journal of Research in Science Teaching, **34**, (9), 949-968.

Kuhn, T. S. (1962) The structure of scientific revolutions. University of Chicago Press, Chicago, USA.

Lambrecht, M.G., Bauer, F.F., Marmur, J. and Pretorius, I.S. (1996) *Muc 1*, a mucin-like protein that is regulated by *Mss 10*, is critical for pseudohyphal differentiation in yeast. Proc. Natl. Acad. Sci. USA, **93**, 8419-8424.

Larkin, J., McDermott, J., Simon, D. and Simon, H. (1980) Expert and novice performance in solving Physics problems. Science, **208**, 1335-1342.

Lemon, K.P. and Grossman, A.D. (1998) Localization of bacterial DNA polymerase: Evidence for a factory model of replication. Science, **282**,1516-1519.

Letsinger, R.L. and Lunsford, W.B. (1976) Synthesis of thymidine oligonucleotides by phosphite triester intermediates. Journal of the American Chemical Society, **98**, 3655-3661.

Letsinger, R.L., Finnan, J.L., Heavner, G.A. and Lunsford, W.B. (1975) Phosphite coupling procedure for generating internucleotide links. Journal of the American Chemical Society, **97**, 3278-3279.

Leung, S., Mellish, C. and Robertson, D. (2001) Basic gene grammars and DNA-ChartParser for language processing of *Escherichia coli* promoter DNA sequences. Bioinformatics, **17**, (3), 226-236.

Lewalter, D. (2003) Cognitive strategies for learning from static and dynamic visuals. Learning and Instruction, **13**, 177-189.

Lewin, B. (1994) Genes V, Oxford University Press, UK.

Lewis, J. and Kattman, U. (2004) Traits, genes, particles and information: Re-visiting students' understanding of Genetics. International Journal of Science Education, **26**, (2), 195-206.

Lewis, J. and Wood-Robinson, C. (2000) Genes, chromosomes, cell division and inheritance. Do students see any relationship? International Journal of Science Education, **22**, (2), 177-195.

Lewis, J., Leach, J. and Wood-Robinson, C. (2000) All in the genes? – Young peoples' understanding of the nature of genes. Journal of Biological Education, **34**, 74-79.

Liew, C. and Treagust, D.F. (1995) A predict-observe-explain teaching sequence for learning about students' understanding of heat and expansion of liquids. Australian Science Teachers' Journal, **41**, 68-71.

Lifton, R.P., Goldberg, M.C., Karp, R.W. and Hogness, D.S. (1978) The organization of the histone genes in *Drosophila melanogaster*: functional and evolutionary implications. Cold Spring Harbor Symp. Quant. Biol., **42**, 1047-1051.

Ligon, J.M., Bolen, P.L., Hill, D.S., Bothast, R.J. and Kurtzman, C.P. (1989) Physical and biological characterization of linear DNA plasmids of the yeast *Pichia inositovora*. Plasmid, **21**, (3), 185-194.

Lincoln, Y.S. and Guba, E.G. (1985) Naturalistic Inquiry. SAGE Publications, Newbury Park, pp.268-273.

Lock, R. and Miles, C. (1993) Biotechnology and Genetic Engineering: Students' knowledge and attitudes. Journal of Biological Education, **27**, 267-272.

Lodish, H., Berk, A., Matsudaira, P., Kaiser, C.A., Kreiger, M., Scott, M.P., Zipursky, S.L. and Darnell, J. (2004) Molecular Cell Biology, 5th edition, W.H. Freeman and Co., New York.

Longden, B. (1982) Genetics – are these inherent learning difficulties? Journal of Biological Education, **16**, 135-140.

Lord, T.R. (1985) Enhancing the visuo-spatial aptitude of students. Journal of Research in Science Teaching, **22**, 395-405.

Losick, R. and Pero, J. (1981) Cascade of sigma factors. Cell, **25**, 582-584.

Losick, R. and Shapiro, L. (1998) Bringing the mountain to Mohammed. Science, **282**, 1430-1431.

Lowe, R. (1988) "Reading" scientific diagrams: Characterising components of skilled performance. Research in Science Education, **18**, 112-122.

Lowe, R. (1993) Diagrammatic information: techniques for exploring its mental representation and processing. Information Design Journal, **7**, (1), 3-17.

Lowe, R. (2003) Animation and learning: Selective processing of information in dynamic graphics. Learning and Instruction, **13**, (2), 125-139.

Lowe, R. (2004) Interrogation of a dynamic visualization during learning. Learning and Instruction, **14**, (3), 257-274.

Maki, H., Maki, S. and Kornberg, A. (1988) DNA polymerase III holoenzyme of *Escherichia coli*. Journal of Biological Chemistry, **263**, 6570-6578.

Malloy, S. (2001) Genetic Nomenclature. http://www.sci.sdsu.edu/~smalloy/Microbial Genetics/topics/mutations/nomenclature-v3.pdf, Accessed: August 2005.

Mammino, L. (2001) Physical quantities and their changes: Difficulties and perceptions by Chemistry students. Journal of SAARMSTE, **5**, 29-40.

Marbach-Ad, G. (2001) Attempting to break the code in student comprehension of genetic concepts. Journal of Biological Education, **35**, 185-189.

Marchant, R. and Marchant, E. M. (1999) GM plants: Concepts and issues. Journal of Biological Education, **34**, 5-12.

Martinez-Garcia, M.V., Gil-Quilez, M.J. and Osada, J. (2003) Genetic Engineering: A matter that requires further refinement in Spanish secondary school textbooks. International Journal of Science Education, **25**, (9), 1147-1168.

Mathew, H.R., Freedland, R. and Miesfield, R.L. (1997) Biochemistry: A short course, Wiley Liss, New York.

Mathews, M.R. (2004) Thomas Kuhn's impact on Science Education: What lessons can be learnt? Science Education, **88**, (1), 90-118.

Mathews, C.K. and van Holde, K.E. (1990) Biochemistry, Benjamin/Cummings, Menlo Park, CA.

Mathewson, J.H. (2005) The visual core of Science: Definition and applications to Education. International Journal of Science Education, **27**, (5), 529-548.

Mayer, R.E. (2003) The promise of multimedia learning: Using the same instructional design methods across different media. Learning and Instruction, **13**, (2), 125-139.

McClintock, B. (1951) Chromosome organisation and genic expression. Cold Spring Harbor Symp. Quant. Biol., **16**, 13-47.

McClintock, B. (1956) Controlling elements and the gene. Cold Spring Harbor Symp. Quant. Biol., **21**, 197-216.

McCracken, D.A., Martin, V.J., Stark, M.J. and Bolen, P.L. (1994) The linear plasmid-encoded toxin produced by the yeast *Pichia acaciae*: Characterization and comparison with the toxin of *Kluyveromyces lactis*. Microbiology, **140**, 425-431.

McKee, T. and McKee, J.R. (1996) Biochemistry: An introduction, Wm C. Brown Publishers, London.

McKloskey, M. (1983) Intuitive Physics. Scientific American, **248**, 122-130.

McMillan, J.H. and Schumacher, S. (1993) Research in Education, 3rd ed., Harper Collins College Publishers, New York.

Menger, F.M., Zana, R. and Lindman, B. (1998) Portraying the structure of micelles. Journal of Chemical Education, **75**, 115.

Meselson, M. and Stahl, F.W. (1958) The replication of DNA in *Escherichia coli*. Proc. Natl. Acad. Sci. USA, **44**, 671-682.

Messing, J. and Vieira, J. (1982) A new pair of M13 vectors for selecting either DNA strand of double-digest restriction fragments. Gene, **19**, (3), 269-276.

Meyer, E. and Land, R. (2003) Threshold concepts and troublesome knowledge: Linkage to ways of thinking and practicing within the discipline. Occasional Report 4, May 2003, ETL Project, University of Edinburgh, Coventry and Durham, UK.

Novak, J.D. (1984) Applications of advances in learning theory and philosophy of Science to the improvement of Chemistry teaching. Journal of Chemical Education, **61**, (7), 607-612.

Nurrenbern, S.C. (2001) Piaget's theory of intellectual development revisited. Journal of Chemical Education, **78**, (8), 1107-1110.

Nakhleh, M.B. (1994) Symposium: What is research in Chemistry Education? Chemical Education research in the laboratory environment. Journal of Chemical Education, **71**, (3), 201-205.

Novak, J.D. (1977) A Theory of Education. Ithaca, NY, Cornell University.

Novick, R.(1998) Contrasting lifestyles of rolling-circle phages and plasmids. TIBS, **23**, 434-438.

Nes, W.R. and McKean, W.L. (1977) Analytical procedures in Biochemistry of steroids and other isopentenoids, Chapter 3, University Park Press, Baltimore, USA, p.119.

O' Donnell, M.E. and Kornberg, A. (1985) Dynamics of DNA polymerase III holoenzyme of *Escherichia coli* in replication of a mutiprimed template. Journal of Biological Chemistry, **260**, 12875-12883.

Oh, P.S. (2005) Discursive roles of the teacher during class sessions for students presenting their Science investigations. International Journal of Science Education, **27**, (15/16), 1825-1851.

Oka, A., Sugimoto, K., Takanami, M. and Hirota, Y. (1980) Replication origin of the *Escherichia coli* K-12 chromosome: The size and structure of the minimum segment carrying the information for autonomous replication. Molecular and General Genetics, **178**, (1), 9-20.

Okazaki, R., Okazaki, T., Sakabe, K., Sugimoto, K. and Sugino, A. (1968) Mechanism of DNA chain growth. i. Possible discontinuity and unusual secondary structure of newly synthesized chains. Proc. Natl. Acad. Sci. USA, **59**, 598-605.

Orgill, M. and Bodner, G. (2007) Locks and Keys: An analysis of Biochemistry students' use of analogies. Biochemistry and Molecular Biology Education, **35**, (4), 244-254.

Ozkaya, A.R. (2002) Conceptual difficulties experienced by prospective teachers in Electrochemistry: Half-cell potential, cell potential, and chemical and electrochemical equilibrium in galvanic cells. Journal of Chemical Education, **79**, 735-738.

Partington, J.R. (1951) General and Inorganic Chemistry, 2nd ed., Macmillan and CO., Ltd., London.

Patrick, M.D., Carter, G. and Wiebe, E.N. (2005) Visual representations of DNA replication: Middle grades students' perceptions and interpretations. Journal of Science Education and Technology, **14**, 353-365.

Paivio, A. (1986) Mental representations: A dual coding approach, Oxford University Press, Oxford.

Parslow, G.R. (2004) Multimedia in Biochemistry and Molecular Biology Education. Biochemistry and Molecular Biology Education, **32**, (4), 280-281.

Pata, K. and Sarapuu, T. (2006) A comparison of reasoning processes in a collaborative modelling environment: Learning about Genetics problems using virtual chat. International Journal of Science Education, **28**, (11), 1347-1368.

Pearsall, N.R., Skipper, J.E.J. and Mintzes, J.J. (1997) Knowledge restructuring in the Life Sciences: A longitudinal study of conceptual change in Biology. Science Education, **81**, 193-215.

Perkins, D.N. and Salomon, G. (1989) Are cognitive skills context-bound? Educational Researcher, **18**, (1), 16-25.

Perini, L. (2005) Explanation in two dimensions: Diagrams and biological explanation. Biology and Philosophy, **20**, 257-260.

Peterson, R.F. and Treagust, D.F. (1989) Grade-12 students' misconceptions of covalent bonding and structure. Journal of Chemical Education, **66**, (6), 459-460.

Peterson, R.F., Treagust, D.F. and Garnett, P. (1989) Development and application of a diagnostic instrument to evaluate grade -11 and -12 students' concepts of covalent bonding and structure following a course of instruction. Journal of Research in Science Teaching, **26**, (4), 301-314.

Pfundt, H. and Duit, R. (1994) Bibliography : Students' Alternative Frameworks, 4th ed., Kiel, Germany IPN.

Pines, A.L. and West, L.H.T. (1986) Conceptual understanding and science learning: An interpretation of research within a sources of knowledge framework. Science Education, **70**, 584-604.

Pintrich, P.R., Marx, R.W. and Boyle, R.A. (1993) Beyond cold conceptual change: The role of motivational beliefs and classroom contextual factors in the process of conceptual change. Review of Educational Research, **63**, 167-199.

Pittman, K.M. (1999) Student generated analogies: Another way of knowing? Journal of Research in Science Teaching, **36**, (1), 1-22.

Ploetzner, R. and Lowe, R. (2004) Dynamic visualizations and learning. Learning and Instruction, **14**, (3), 235-240.

Plummer, D.T. (1987) An introduction to Practical Biochemistry, 3rd edition, McGraw-Hill Book Co., London.

Pon, R.T. (1987) An improved iodine / water oxidation reagent for automated oligonucleotide synthesis. Nucleic Acids Research, **15**, (17), 7203.

Posner, G.J. and Gertzog, W.A. (1982) The clinical interview and measurement of conceptual change, Science Education, **66**, (2), 195-209.

Posner, G.J., Strike, K.A., Hewson, P.W. and Gertzog, W.A. (1982) Accommodation of a scientific conception: Toward a theory of conceptual change. Science Education, **66**, 211-227

Pretorius, I.S. (2000) Tailoring wine yeast for the new millennium: Novel approaches to the ancient art of winemaking. Yeast, **16**, (8), 675-729.

Pribnow, D. (1975) Nucleotide sequence of an RNA polymerase binding site at an early T_7 promoter. Proc. Natl. Acad. Sci. (USA), **72**, 784-788.

Race, P. (1996) The Art of Assessing, In: http://www.city.londonmet.ac.uk/deliberations, Accessed: June 2005.

Rao, A.S. (1996) Terminology in microbial metabolism. Biochemical Education, **24**, (1), 61-62.

Reynolds, B. (1965) Learning and teaching of the practice of social work, 2nd edition, Russell and Russell, New York.

Rhemtula, M. and Rollnick, M. (2002) The influence of the initial aspects of a course on pedagogical content knowledge on a group of teachers and teacher educators. In: Malcolm, C. and Lubisi, C. (eds.), Proceedings of the Tenth Annual Meeting of the Southern African Association for Research in Mathematics, Science and Technology Education, University of Natal, Durban, 22-26 January 2002, section III, pp.335-340.

Richardson, D.C. and Richardson, J.S. (2002) Teaching molecular 3-D literacy. Biochemistry and Molecular Biology Education, **30**, (1), 24-26.

Rieber, L.P., Tzeng, S-C. and Tribble, K. (2004) Discovery learning, representation and explanation within a computer-based simulation: Finding the right mix. Learning and Instruction, **14**, (3), 307-323.

Rotbain, Y., Marbach-Ad, G. and Stavy, R. (2006) Effect of bead and illustrations models on high school students' achievement in Molecular Genetics, **43**, (5), 500-529.

Sadler, T.D. and Zeidler, D.L. (2004) The morality of socioscientific issues: Construal and resolution of genetic engineering dilemmas. Science Education, **88**, (1), 4-27.

Sadler, T.D. and Zeidler, D.L. (2005a) Patterns of informal reasoning in the context of socioscientific decision-making. Journal of Research in Science Teaching, **42**, (1), 112-138.

Sadler, T.D. and Zeidler, D.L. (2005b) The significance of content knowledge for informal reasoning regarding socioscientific issues: Applying Genetics knowledge to Genetic Engineering issues. Science Education, **89**, (1), 71-93.

Salomon, G. (2005) Symbol systems. In: http://tip.psychology.org/salomon.html, Accessed: May 2005.

Salomon, G. and Perkins, D.N. (1989) Rocky roads to transfer: Rethinking mechanisms of a neglected phenomenon. Educational Psychologist, **24**, (2), 113-142.

Sandoval, W. and Reiser, B.J. (2004) Explanation-driven inquiry: Integrating conceptual and epistemic scaffolds for scientific inquiry. Science Education, **88**, 345-372.

Sanger, F., Coulson, A.R., Friedmann, T., Air, G.M., Barrell, B.G., Brown, N.L., Fiddes, J.C., Hutchinson, C.A. III, Slocombe, P.M. and Smith, M. (1978) The nucleotide sequence of bacteriophage phiX174. Journal of Molecular Biology, **125**, (2), 225-246.

Scaife, M. and Rogers, Y. (1996) External cognition: How do graphical representations work? Int. J. Hum. Comput. Studies, **45**, 185-213.

Schaller, H., Gray, C. and Herrmann, K. (1975) Nucleotide sequence of an RNA polymerase binding site from the DNA of bacteriophage fd. Proc. Natl. Acad. Sci.(USA), **72**, 737-744.

Schnotz, W. and Bannert, M. (2003) Construction and interference in learning from multiple representations. Learning and Instruction, **13**, 141-156.

Schönborn, K.J. and Anderson, T.R. (2006a) The importance of visual literacy in the education of biochemists. Biochemistry and Molecular Biology Education, **34**, (2), 94-102.

Schönborn, K.J. and Anderson, T.R. (2006b) A model of factors determining students' ability to interpret external representations in Biochemistry. Proceedings of the NARST 2006 Annual Meeting , April 3-6, San Francisco, CA, USA.

Schönborn, K.J., Anderson, T.R. and Grayson, D.J. (2002a) Student difficulties with the interpretation of a textbook diagram of Immunoglobulin G (IgG). Biochemistry and Molecular Biology Education, **30**, (2), 93-97.

Schönborn, K.J., Anderson, T.R. and Grayson, D.J. (2002b) Developing and testing a model of the factors affecting student interaction with scientific diagrams. In: Malcolm, C. and Lubisi, C. (eds.), Proceedings of the Tenth Annual Meeting of the Southern African Association for Research in Mathematics, Science and Technology Education, University of Natal, Durban, 22-26 January 2002, section III, pp.377-383.

Scism, A.J. (1996) The Biochemistry Student Companion, Prentice Hall, New Jersey.

Seeman, N. (2004) Nanotechnology and the double helix. Scientific American, **290**, (6), 64-75.

Seethaler, S. and Linn, M. (2004) Genetically-modified food in perspective: An inquiry-based curriculum to help middle school students make sense of tradeoffs. International Journal of Science Education, **26**, (14), 1765-1785.

Seufert, T. (2003) Supporting coherence formation learning from multiple representations. Learning and Instruction, **13**, 227-237.

Sharp, P.A., Sugden, B. and Sambrook, J. (1973) Detection of 2 restriction endonuclease activities in *Haemophilus parainfluenzae* using analytical agarose-ethidium bromide electrophoresis. Biochemistry, **12**, 3055-3063.

Sherman, F. and Lawrence, C.W. (1974) *Saccharomyces*. In: King, R.C. (ed.), Handbook of Genetics, Plenum Press, New York, London, Vol. 1, pp.359-393.

Shulman, L.S. (1986) Those who understand: Knowledge growth in teaching. Educational Researcher, **15**, (2), 4-14.

Silhavy, T.J. (2005) Princeton University, Biotechnology Department, personal communication.

Simonneaux, L. (2001) Role-play or debate to promote students' argumentation and justification on an issue in animal transgenesis. International Journal of Science Education, **23**, 903-927.

Simonneaux, L. (2000) A study of pupils' conceptions and reasoning in connection with "microbes", as a contribution to research in Biotechnology Education. International Journal of Science Education, **22**, 619-644.

Sins, P.H.M., Savelsbergh, E.R. and van Joolingen, W.R. (2005) The difficult process of scientific modelling: An analyses of novices' reasoning during computer-based modelling. International Journal of Science Education, **27**, (14/18), 1695-1721.

Sloutsky, V.M., Kaminski, J.A. and Heckler, A.F. (2005) The advantage of simple symbols for learning and transfer. Psychonomic Bulletin and Review, **12**, (3), 508-513.

Snedecor, G.W. and Cochran, W.G. (1980) Statistical Methods, The Iowa State University Press, USA.

Snow, R.E. and Lohman, D.F. (1984) Toward a theory of cognitive aptitude for learning from instruction. Journal of Educational Psychology, **76**, (3), 347-376.

Southern, E.M. (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol., **98**, 503-517.

Srinivasan, K.R. (1995) Use of an RK2-based tri-parental mating system with *Rhodococcus* 124 and B264-1. Bug Journal, **2**, 223-231.

Staines, N.A. (1983) Monoclonal Antibodies. In: Wrigglesworth, J.M. (ed.), Biochemical Research Techniques, Chapter 6, John Wiley and Sons Ltd., pp. 178-205.

Staver, J.R. and Lumpe, A.T. (1993) A content analysis of the presentation of the mole concept in Chemistry textbooks. Journal of Research in Science Teaching, **30**, (4), 321-337.

Stewart, J. (1985) Cognitive Science and Science Education. European Journal of Science Education, **7**, (1), 1-17.

Stewart, J. and Atkin, J. (1982) Information processing psychology: A promising paradigm for research in Science teaching. Journal of Research in Science Teaching, **19**, 321-332.

Stewart, J. and Dale, M. (1989) High school students' understanding of chromosome / gene behaviour during meiosis. Science Education, **73**, 501-521.

Stokstad, E. (2001) Reintroducing the Intro Course. Science, **293**, 1608-1610.

Stryer, L. (1995) Biochemistry, 4th ed., W.H. Freeman and Co., New York.

Stylianidou, C. and Boohan, R. (1998) Understanding why things happen: Case-studies of pupils using an abstract picture language to represent the nature of changes. Research in Science Education, **28**, (4), 447-462.

Sundberg, M.D. (2002) Assessing student learning. Cell Biology Education, **1**, 11-15.

Sweller, J. and Chandler, P. (1994) Why some material is difficult to learn. Cognition and Instruction, **12**, (3), 185-233.

Szeberényi, J. (2002) Problem-based learning: Restriction Mapping. Biochemistry and Molecular Biology Education, **30**, (4), 258-259.

Szeberényi, J. (2003) Problem-based learning: The role of a chaperone in protein synthesis . Biochemistry and Molecular Biology Education, **31**, (2), 137-138.

"Teachnology" (2005) The Web Portal for Educators. In: http://tip.psychology.org /bruner.html, Accessed June 2005.

Taber, K.S. (1998) An alternative conceptual framework from Chemistry Education. International Journal of Science Education, **20**, (5), 597-608.

Taber, K.S. (2003) Mediating mental models of metals: Acknowledging the priority of the learners' prior learning. Research in Science Education, **87**, 732-758.

Taconis, R., Ferguson-Hessler, M.G.M. and Broekkamp, H. (2001) Teaching Science problem solving: An overview of experimental work. Journal of Research in Science Teaching, **38**, 442-468.

Takayama, K. (2005) Visualizing the science of genomics. In: Gilbert, J.K. (Ed.), Visualization in Science Education, Chapter 11, Springer, The Netherlands, pp.217-252.

Talbot, C. (2001) Some common biochemical mistakes. School Science Review, **82**, (301), 91-92.

Tamir, P. (1989) Some issues related to the use of justifications to multiple choice answers. Journal of Biological Education, **23**, 4, 285-292.

Tanaka, T. and Letsinger, R.L. (1982) Syringe method for stepwise chemical synthesis of oligonucleotides. Nucleic Acids Research, **10**, (10), 3249-3260.

Taylor, T.R. (1994) A review of three approaches to cognitive assessment, and a proposed integrated approach based on a unifying theoretical framework. South African Journal of Psychology, **24**, (4), 184-193.

The Concise Oxford Dictionary (1984) Sykes, J.B. (Ed.), 7th edition, University Press, Oxford, pp.160-161.

The Flybase Reference Manual C (2005) Using flybase on the web. In: http://fbserver.gen.cam.ac.uk:7081/docs/JK/refman-C.html, Accessed: August 2005.

The Plant Cell, journal website, http://www.plantcell.org/misc/ifora.shtml, Accessed: August 2005.

The Transformer Protocol (2001) Version 1, Teachers guide to symbol use. In: www.ncbe.reading.ac.uk/NCBE/MATERIALS/DNA/Trans TG.pdf, Accessed: June 2005.

Thorndyke, P.W. and Stasz, C. (1985) Individual differences in procedures for knowledge acquisition from maps. Cognitive Psychology, **12**, 137-175.

TIP Theories (2005) Lateral thinking (de Bono). In: http://tip.psychology.org/debono.html, Accessed: May 2005.

Tomaska, L. (2000) Increasing student involvement in lectures by error-prone talks. Biochemistry and Molecular Biology Education, **28**, 242-243.

Treagust, D.F. (1988) Development and use of diagnostic tests to evaluate students' misconceptions in Science. International Journal of Science Education, **10**, (2), 159-169.

Treagust, D.F., Chittleborough, G. and Mamiala, T.L. (2003) The role of submicroscopic and symbolic representations in chemical explanations. International Journal of Science Education, **25**, (11), 1353-1368.

Treagust, D.F., Duit, R. and Fraser, B.J. (1996) Overview: Research on students' preinstructional conceptions - the driving force for improving teaching and learning in Science and Mathematics. In: Treagust, D.F., Duit, R. and Fraser, B.J. (eds.), Improving teaching and learning in Science and Mathematics, Teachers College Press, New York, pp.1-14.

Treagust, D.F., Duit, R., Joslin, P. and Lindauer, I. (1992) Science teachers use of analogies: Observations from classroom practice. International Journal of Science Education, **14**, (4), 413-422.

Tsui, C.-Y. and Treagust, D.F. (2003) Genetics reasoning with multiple external representations. Research in Science Education, **33**, 111-135.

Turner, P.C., McLennan, A.G., Bates, A.D. and White, M.R.H. (1997) Instant notes in Molecular Biology, Bios Scientific Publishers Limited, UK.

Tversky, B. (2003) Spatial schemas in depictions. In: Gattis, M. (ed) Spatial schemas and abstract thought, MIT Press, Cambridge.

Vadasz, P. and Vadasz, A. (2005) Predictive modelling of micro-organisms: LAG and LIP in monotonic growth. International Journal of Food Microbiology, **102**, 257-275.

Valentine, R.C. and Green, N.M. (1967) Electron microscopy of an antibody-hapten complex. J. of Mol. Biol., **27**, 615-617.

van Mansfield, A.D.M., van Teeffelen, H.A.A.M., Baas, P.D. and Jansz, H.S. (1986) Two juxtaposed tyrosyl-OH groups participate in phiX174 gene A protein catalysed cleavage and ligation of DNA. Nucleic Acids Research, **14**, 4229-4238.

Van Zele, E., Lenaerts, J. and Wieme, W. (2004) Improving the usefulness of concept maps as a research tool for Science Education. International Journal of Science Education, **26**, (9), 1043-1064.

Venville, G. and Donovan, J. (2007) Developing year 2 students' theory of Biology with concepts of the gene and DNA. International Journal of Science Education, **29**, (9), 1111-1131.

Venville, G. and Treagust, D.F. (1997) Analogies in Biology education: A contentious issue. The American Biology Teacher, **59**, 282-287.

Venville, G. and Treagust, D.F. (1998) Exploring conceptual changes in Genetics using a multi-dimensional framework. Journal of Research in Science Teaching, **35**, (9), 1031-1055.

Verstrepen, K.J., Derdelinckx, G., Delvaux, F.R., Winderickx, J., Thevelein, J.M., Bauer, F.F. and Pretorius, I.S. (2001) Late fermentation expression of *FLO 1* in *Saccharomyces cerevisiae*. American Society of Brewing Chemists, Inc., Publication J-2001-0314-02R, pp.69-75.

Vieira, J. and Messing, J. (1982) The pUC plasmids, an M13 mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. Gene, **19**, 259-268.

Virtual Lab Book (2005) Understanding plasmid DNA. In: http://www.cofc.edu/~dellis/virtuallabbook/PreLabReadings/Exb.html, Accessed: May 2005.

Vlassis, J. (2004) Making sense of the minus sign or becoming flexible in 'negativity'. Learning and Instruction, **14**, 469-484.

Voet, D. and Voet, J.G. (2004) Biochemistry, 3rd edition, John Wiley and Sons Inc., USA.

Vogelezang, M. J. (1987) Development of the concept "chemical substance" – some thoughts and arguments. International Journal of Science Education, **9**, 519-528.

Von Glasersfeld, E. (1992) A constructivist view of learning and teaching. In: Duit, R., Goldberg, F. and Niederrer, H. (eds.), Research in Physics learning: Theoretical issues and empirical studies. Kiel, Germany: IPN, University of Kiel, pp.29-39.

Vosniadou, S. and Ioannides, C. (1998) From conceptual change to Science Education: A psychological point of view. International Journal of Science Education, **20**, 1213-1230.

Vygotsky, L.S.(1978) Mind in society: the development of higher psychological processes. Harvard University Press, Cambridge, MA.

Wallace, J.D. and Mintzes, J.J. (1990) The concept map as a research tool: Exploring conceptual change in Biology. Journal of Research in Science Teaching, **27**, 1033-1052.

Walsh, L., Shaker, E. and De Stasio, E.A. (2007) Using restriction mapping to teach basic skills in the Molecular Biology laboratory. Biochemistry and Molecular Biology Education, **35**, (3), 199-200.

Wandersee, J.H., Mintzes, J.J. and Novak, J.D. (1994) Research on alternative conceptions in Science. In: Gabel, D. (ed.), Handbook of research on Science teaching and learning, Macmillan, New York, pp.177-210.

Watson, J.D. and Crick, F.H.C. (1953) A structure for deoxyribose nucleic acid. Nature, **171**, 737-738.

Watson, J.D., Hopkins, N.H., Roberts, J.W., Steitz, J.A. and Weiner, A.M. (1987) Molecular Biology of the Gene, 4th ed., Benjamin/Cummings, Menlo Park, CA.

Weaver, R.F. (1999) Molecular Biology, International edition, The McGraw-Hill Companies Inc., USA.

Wheeler, A.E. and Hill, D. (1990) Diagram-Ease:Why some students misinterpret diagrams. The Science Teacher, **57**, 59-63.

White, A., Handler, P. and Smith, E.L. (1959) Nucleic Acids and Nucleoproteins. In: Principles of Biochemistry, 3rd edition, Chapter 10, McGraw-Hill Inc., New York, pp.162-186.

White, R.T. and Gunstone, R.F. (1992) Probing Understanding, Falmer Press, London.

Wikipedia Encyclopedia (2007a) http://en.wikipedia.org/wiki/Argonauts, Accessed: August 2007.

Wikipedia Encyclopedia (2007b) http://en.wikipedia.org/wiki/Genetic_code, Accessed: May 2007.

Winberg, C. (1997) How to research and evaluate. In: Calitz, H.(ed), The Teaching and Learning series, Juta and Co., Ltd., Cape Town, SA, pp.14-29.

Wood, W.B. (2002) Genesis of Biochemistry: A problems approach. Cell Biology Education, **1**, 16-17.

Wu, H.-K. and Shah, P. (2004) Exploring visuospatial thinking in Chemistry learning. Science Education, **88**, 465-492.

Yarroch, W.L. (1985) Student understanding of chemical equation balancing. Journal of Research in Science Teaching, **22**, 449-459.

Yore, L.D. and Treagust, D.F. (2006) Current realities and future possibilities: Language and science literacy – empowering research and informing instruction. International Journal of Science Education, **28**, (2)-(3), 291-314.

Zahn, C., Barquero, B. and Schwan, S. (2004) Learning with hyperlinked videos - design criteria and efficient strategies for using audiovisual hypermedia. Learning and Instruction, **14**, (3), 275-291.

Zamore, P.D. (2001) RNA Interference: listening to the sound of silence. Nature Structural Biology, **8**, (9), 746-740.

Zohar, A. and Nemet, F. (2002) Fostering students' knowledge and argumentation skills through dilemmas in Human Genetics. Journal of Research in Science Teaching, **39**, (1), 35-62.

Zubay, G.L., Parson, W.M. and Vance, D.E. (1995) Principles of Biochemistry, Wm. C. Brown Publishers, Iowa, USA.

Zurawski, R. (1998) Making the most of exams: procedures for item analysis. The National Teaching and Learning Forum, **7**, (6), 1-4.