UNIVERSITY OF KWAZULU-NATAL

A COMPUTATIONAL PERSPECTIVE OF INFLUENZA A VIRUS TARGETS: NEURAMINIDASE AND ENDONUCLEASE

2016

ASHONA SINGH

205504216

A COMPUTATIONAL PERSPECTIVE OF INFLUENZA A VIRUS TARGETS: NEURAMINIDASE AND ENDONUCLEASE

ASHONA SINGH

205504216

2016

A thesis submitted to the School of Pharmacy and Pharmacology, Faculty of Health Science, University of KwaZulu-Natal, Westville, for the degree of PhD Pharmaceutical Chemistry.

This is a thesis in which the chapter are written as a set of discrete research publications, with an overall introduction and final summary. Typically these chapters will have been published in internationally recognised, peer-reviewed journals.

This is to certify that the contents of this thesis are the original work of Ms Ashona Singh.

As the candidate's supervisor. I have approved this thesis for submission.

Supervisor:

Signed: _____ Name: **Prof. Mahmoud Soliman** Date: _____

Abstract

Through the ages the viruses have plagued mankind claiming the lives of millions, pre-dating any advancements in the medicinal sciences. One such pathogenic virus is influenza A, which has been implicated in the 1918-Spanish flu, the 2006-avian flu outbreak and the 2009-swine flu pandemic. It is a highly sophisticated species, alluding efforts to thwart the spread of disease and infection. One of the main reasons influenza has survived this long is simple evolution. Natural mutation within the genome of virions expressed in proteins, enzymes or molecular structure render us unable to predict or take preventative measures against possible infection. Thus, research efforts toward the competitive inhibition of biological pathways that lead to the spread of disease, have become attractive targets.

The influenza A virus has a number of chemotherapeutic targets, such as:

1) The surface antigens, hemagglutinin and neuraminidase,
2) RNA-dependent RNA polymerase, and
3) The M2 proton channel.

Influenza RNA polymerase is composed of three large segments encoding polymerase acidic protein (PA), polymerase basic protein 1 (PB1) and polymerase basic protein 2 (PB2). The PA protein is an N-terminal domain subunit which contains the endonuclease activity. The influenza virus is incapable of synthesizing a 5'-mRNA cap, so it has adapted a cap-snatching mechanism whereby the PB2 subunit binds to the 5'-end of host mRNA, after which 10-14 nucleotides downstream the PA-subunit (aka PA_N) cleaves the strand forming a primer for viral mRNA synthesis which is catalysed by the PB1 subunit. Influenza target identification is based primarily on evidence suggesting sequence conservation of each entity and its selective expression in the virus and not the host.

In this thesis two enzymatic targets were investigated, the PA protein of RNA polymerase and neuraminidase. The studies focussed on using computational tools to:

1) provide insight into the mechanism of drug-resistance,
2) describe the conformational structure of the protein in the presence of point mutations and in complex with an inhibitor,
3) determine the essential binding pharmacophoric features to aid the design of new drug therapies.

An array of computational techniques were employed in the studies, such as: molecular dynamics (MD) simulation, structure-based and ligand-based *in silico* screening, principal component analysis, radius of gyration analysis, binding free energy calculations and solvent-accessible surface area analysis.

The first study (Chapter 5) determined the mechanism of drug-resistance in influenza A neuraminidase as a consequence of antigenic variations. Two distinct mutations in the enzyme sequence that were investigated are H274Y and I222K. The active site residues of neuraminidase are conserved among the subtypes of influenza A. However, it was discovered that the occurrence of resistance to the drug oseltamivir, in the H1N1 species was different to the H5N1 virus. Although both systems shared a loss in hydrophobicity of the active site, the conformational distortion of the active site pocket distinguished the enzyme of the two viral entities, from one another.

The discoveries made in the first study laid the foundation for the second study (Chapter 6), which was based on the *in silico* design and screen of potential neuraminidase inhibitors. As a result 10 characteristic molecular scaffolds were suggested as potential inhibitors. The pharmacophore design was constructed with consideration to the new conformational structure of the active site pocket.

Chapter 7 is the third study of this thesis. The active site pocket enclosing the endonuclease activity of the PA subunit was investigated. Using molecular dynamics simulations and post-dynamic analyses, a description of the protein conformation was offered. Subsequently, a pharmacophore was proposed as a potential scaffold to which endonuclease inhibitors may be modelled upon.

It is my belief that the impact of the results derived from the above mentioned studies would greatly contribute to the development of new and effective anti-influenza drugs.

DECLARATION- I PLAGIARISM

I, Ashona Singh, declare that

1. The research reported in this thesis, except where otherwise indicated, is my original research.

2. This thesis has not been submitted for any degree or examination at any other university.

3. This thesis does not contain other persons' data, pictures, graphs or other information, unless specifically acknowledged as being sourced from other persons.

4. This thesis does not contain other persons' writing, unless specifically acknowledged as being sourced from other researchers.   Where other written sources have been quoted, then:

    a. Their words have been re-written but the general information attributed to them has been referenced.
    b. Where their exact words have been used, then their writing has been placed in italics and inside quotation marks, and referenced.

5. This thesis does not contain text, graphics or tables copied and pasted from the internet, unless specifically acknowledged, and the source being detailed in the thesis and in the References sections.

    A detailed contribution to publications that form part and/or include research presented in this thesis is stated (include publications submitted, accepted, in *press* and published).

Signed: _____

DECLARATION- II LIST OF PUBLICATIONS

PUBLISHED

1. Singh, A., Mhlongo, N. and Soliman, M. E. (2015). *Anti-cancer Glycosidase Inhibitors from Natural Products: A Computational and Molecular Modelling Perspective.* Anti-Cancer Agents of Medicinal Chemistry, 15 (8): 933-946. (**Impact Factor** = 2.4)

   Contributions:
   a. <u>Singh, A.</u>: Main contributor, contributed by literature search, compilation and writing of the manuscript
   b. Mhlongo, N.: contributed to the manuscript as a post-submission editor
   c. Soliman, M.E.: supervisor

2. Singh, A. and Soliman, M.E (2015). *Understanding the cross-resistance of oseltamivir to H1N1 and H5N1 influenza A neuraminidase mutations using multidimensional computational analyses.* Drug Design, Development and Therapy, 9, 4137-4154. (**Impact Factor** = 3.0)

   Contributions:
   a. <u>Singh A.</u>: contributed, to the project by performing experimental work, data analysis and interpretation and manuscript preparation and writing
   b. Soliman, M. E.: supervisor

SUBMITTED

1. Singh, A., Chetty, S. and Soliman, M. E (2015). *Per-residue Free Binding Energy Profiled Pharmacophore Modelling as an Enhanced Approach in Drug Discovery: A Case Study for in silico Screening and Validation of Potential Influenza A Neuraminidase Inhibitors*. Submitted to the Journal of Molecular Biosystems (**Impact Factor** = 3.2)

   Contributions:
   a. Singh, A.: contributed, to the project by performing experimental work, data analysis and interpretation and manuscript preparation and writing
   b. Chetty, S.: co-supervisor with academic contribution
   c. Soliman, M. E.: supervisor

2. Singh, A., Chetty, S. and Soliman, M. E (2015). A molecular dynamics description of the conformational binding of a potential influenza A endonuclease inhibitor. Submitted to the European Journal of Medicinal Chemistry (**Impact Factor** = 3.4)

   Contributions:

   a. Singh, A.: contributed, to the project by performing experimental work, data analysis and interpretation and manuscript preparation and writing
   b. Chetty, S.: co-supervisor with academic contribution
   c. Soliman, M. E.: supervisor

Acknowledgements

"If I have seen further it is by standing on the shoulders of Giants."

I dedicate this thesis to Jeven, my little scientist. Thank you for being my companion through the good and the bad, and making me believe I could do anything. I am blessed as you constantly inspire me to be and do more.

I wish to give a further thanks to:

Lord let this be my time!

List of Abbreviations

Δ	A Change
HA	Hemagglutinin
LB-VS	Ligand-based virtual screening
MD	Molecular dynamics
MM	Molecular Mechanics
NA	Neuraminidase
NMR	Nuclear magnetic resonance
ns	nanosecond
PA	Polymerase acidic protein
PCA	Principal component analysis
PE	Potential energy
QM	Quantum Mechanics
Rg or ROG	Radius of gyration
RMSD	Root mean square deviation
RMSF	Root mean square fluctuation
RNA	Ribonucleic acid
RNP	Ribonucleoprotein
SASA	Solvent accessible surface area
SB-VS	Structure-based virtual screening
TM	Transmembrane
WT	Wild-type

List of Amino Acids

Amino Acid	Three letter code	Single letter code
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic Acid	Asp	D
Glutamic Acid	Glu	E
Glutamine	Gln	Q
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Lysine	Lys	K
Serine	Ser	S
Tyrosine	Tyr	Y

Table of Contents

Chapter 4- Published work

Anti-cancer Glycosidase Inhibitors from Natural Products: A Computational and Molecular Modelling Perspective

Chapter 5- Published work

Understanding the cross-resistance of oseltamivir to H1N1 and H5N1 influenza A neuraminidase mutations using multidimensional computational analyses

Results and discussion

Chapter 6- Submitted manuscript

Per-residue Free Binding Energy Profiled Pharmacophore Modelling as an Enhanced Approach in Drug Discovery: A Case Study for *in silico* Screening and Validation of Potential Influenza A Neuraminidase Inhibitors

Chapter 7- Submitted manuscript

A molecular dynamics description of the conformational binding of a potential influenza A endonuclease inhibitor

Chapter 8- General conclusions

Chapter 1

1. Introduction

1.1 Background and Rationale of Study

The spread of seasonal influenza viral infections causes annual epidemics with more than 20,000 reported deaths. Recent outbreaks of the influenza A virus; subtype H5N1 and pandemic H1N1, have caused concern because of their higher mortality rates [1-5]. International health bodies have invested significant research in effective management strategies for the influenza virus which include: prevention strategies *i.e.* seasonal flu vaccinations [6] and curative chemoprophylaxis active against circulating subtypes of influenza- oseltamivir (Tamiflu®) [7], zanamivir (Relenza®) [8], laninamivir (Inavir®) [9] and peramivir (Rapivab®) [10].

The influenza virus has demonstrated a remarkable ability to mutate rapidly. Mutagenesis occurs by antigenic drifts and antigenic shifts. In 2009, an H1N1 virus with a new combination of genes emerged with a rapid rate of infection, resulting in a pandemic claiming approximately 284,500 lives. The current drug treatments were ineffective against the outbreak. The increase in reported incidences of resistance of H1N1 and H5N1 to current chemotherapies poses a threat to the human population, making us vulnerable to genetically divergent influenza A viruses [11-13].

An extensive database of approximately 870 crystallised proteins (www.rcsb.org) or protein fragments of the influenza A virus [14-16], has assisted in isolating potential targets for the inhibition by a chemotherapeutic. High profile targets include [16, 17]:

1. surface antigens: hemagglutinin (HA) and neuraminidase;
2. RNA dependent RNA polymerase: polymerase acidic protein (PA), polymerase basic protein 1 (PB1) and polymerase basic protein 2 (PB2); and
3. the M2 ion channel. Our study focusses on the enzyme neuraminidase of H1N1 and H5N1 subtypes and the PA endonuclease fragment of the RNA polymerase of H1N1.

Influenza A neuraminidase and endonuclease have been reported to have a high success rate of inhibition when treated with selective chemoprophylaxes. Both enzymes, according to

literature have a largely conserved active site residue profile amongst the influenza A subtypes [18, 19]. However, recent pandemics *i.e.* the 2009-*p*H1N1, expressed a genetic mutation at residue position 274 in the glycoprotein, neuraminidase. The point mutation of histidine (H) to tyrosine (Y), observed a more potent and resilient H1N1 strain comparable to the 2006 highly pathogenic H5N1 avian flu [20, 21]. Resistance of *p*H1N1 and pathogenic H5N1 to the available drug therapies emphasizes the need for further research for new, innovative and effective drug treatments.

Endonuclease, however, has been an untapped focal point. The influenza virus requires transcription and replication mechanistic tools [22-24], by selectively inhibiting the PA subunit we prevent the propagation and spread of viral infection. There are no current FDA-approved drug candidates targeting the heterotrimeric polymerase complex [17, 25]. The favourability and robustness of the PA subunit against evolutionary mutagenesis enable us to design and develop a potential chemotherapy [26]. This study embarks on investigation of the resistance mechanisms behind neuraminidase in the presence of evolutionary mutations as well as characterising the binding profile of endonuclease. We propose potential new drug candidates for the inhibition of neuraminidase.

This study utilises multidimensional computational techniques, such as: molecular dynamics (MD) simulations which is a useful tool in defining the ligand binding affinities and characterising the binding landscape, as well as *in silico* virtual screening methods to identify possible inhibitors [27]. Post-dynamic evaluations provide the insight into the mechanism of resistance and the specific binding mode of drug candidates. The techniques combined in this thesis incorporate: root mean square deviation (RMSD), potential energy, root mean square fluctuation (RMSF), principal component analysis (PCA), residue interaction network (RIN) profiling, solvent accessible surface area (SASA), molecular mechanics generalised Born surface area (MM-GBSA), quasi-harmonic entropy estimations, estimation of total hydrogen bonds, radius of gyration (Rg), secondary structure analysis and average volume estimation of the three-dimensional protein.

1.2 Aim and Objectives

The primary focus of the thesis is to investigate specific classes of influenza virus glycosidase targets. This includes the surface antigen neuraminidase and RNA dependent endonuclease. There are three critical goals:

1. To determine, differentiate and distinguish the mechanism of resistance of neuraminidase in influenza A, 2009-pandemic H1N1 (2009-*p*H1N1) and H5N1 subtypes in the presence of mutation/s H274Y and I222K.

   In order to accomplish this the following objectives were outlined:

   1.1 Estimation of the free binding energies of the drug-protein complexed of wild type and mutant 2009-*p*H1N1 and H5N1 enzyme species.

   1.2 Identify, characterise and compare the binding landscape of wild type and mutant 2009-*p*H1N1 and H5N1 subtypes.

2. To propose potential neuraminidase scaffolds for new and innovative prophylactic chemotherapies *via* the aid of *in silico* technique ligand-based virtual screening (LB-VS) and structure-based virtual screening (SB-VS).

   The study was performed by focussing on the following objectives:

   2.1 Identify current active drug targets of influenza neuraminidase.

   2.2 Establish an average cumulative scaffold of current actives using just the ligand three-dimensional structure for LB-VS.

   2.3 Establish an average cumulative scaffold of current actives using the three-dimensional spatial orientation of ligands in the protein active site for SB-VS.

3. To identify the binding mode of the polymerase endonuclease of H1N1 and isolate potential drug therapies to selectively inhibit enzyme activity, using LB-VS and SB-VS.

   This was achieved using the following objectives:

   3.1 Characterise the active site *i.e.* amino acid residues responsible for the endonuclease activity.

   3.2 Highlight essential pharmacophoric features for future anti-influenza.

1.3 Novelty and Significance of the Study

Contributions have been made in isolating and characterising the protein structure of the influenza A neuraminidase [28-30] and the PA subunit [31, 32]. There have been no reported computational analyses which comparatively investigate the difference in the binding site of H1N1 and H5N1 neuraminidase to explain the contrast in susceptibility to the potent drug, oseltamivir. Computational investigation of influenza A endonuclease is limited and remains largely underdeveloped. There is minimal understanding with regard to the selective inhibition of the enzyme, the binding pocket responsible for enzyme activity and whether the presence of a single or multiple metal center/s is imperative to enzyme function.

This study would provide insight into understanding the mechanism of resistance and relative susceptibilities of genetically diverse H1N1 and H5N1 neuraminidase to available drug therapies. By characterising the altered active sites' structural and functional composition, new inhibitors may be developed. With the use of computational tools a refined perspective may be offered to highlight the endonuclease active site. The study will characterise amino acid residues implicated in enzyme activity as well as the co-ordination profile of metal centres. Defining the binding landscape will offer prospective design of selective and unique inhibitors with critical pharmacophores features.

1.4 Overview of the Thesis

The thesis is divided into eight chapters, including this one:

Chapter 1: Outlines briefly the rationale, aims and objectives of the study, the impact of the research and structure of the thesis.

Chapter 2: Provides insight into the targets of influenza A virus, current drug therapies and the limitations due to resistance. The chapter highlights the structure of neuraminidase and endonuclease, and the mode of action and inhibition of the enzymes.

Chapter 3: This chapter offers a general insight into computational chemistry. The specific molecular modelling tools and simulation techniques utilised in the study are highlighted.

Chapter 4: (Published work)

Has been presented in the format requested by the journal.

This is a review paper which initiated our investigation into influenza glycosidase enzymes. The paper is entitled "Anti-cancer Glycosidase Inhibitors from Natural Products: A Computational and Molecular Modelling Perspective". Published in the Journal of Anti-Cancer Agents of Medicinal Chemistry (IF = 2.4). The journal article specifically addresses natural products used in the treatment of glycosidase enzymes implicated in cancer. During the research it was found that many of the enzymes and drugs expressed duality being linked to the influenza virus.

Chapter 5: (Published work)

Has been presented in the format requested by the journal.

The research paper entitled "Understanding the cross-resistance of oseltamivir to H1N1 and H5N1 influenza A neuraminidase mutations using multidimensional computational analyses" addresses the objectives (1.1 and 1.2) outlined in section 1.2. The journal article was submitted to Drug Design, Development and Therapy (IF = 3.0).

Chapter 6: (Manuscript Submitted and under review)

Has been presented in the format requested by the journal.

The manuscript currently under review entitled "Per-residue Free Binding Energy Profiled Pharmacophore Modelling as an Enhanced Approach in Drug Discovery: A Case Study for *in silico* Screening and Validation of Potential Influenza A Neuraminidase Inhibitors" has been submitted to The Journal of Molecular Biosystems (IF = 3.2). The paper addresses the objectives (2.1 – 2.3) outlined in section 1.2.

Chapter 7: (Manuscript Submitted and under review)

Has been presented in the format requested by the journal.

The manuscript is currently under review entitled "A molecular dynamics description of the conformational binding of a potential influenza A endonuclease inhibitor and *in silico* screen of possible drug candidates" has been submitted to the European Journal of Medicinal Chemistry (IF = 3.4). The paper addresses the objectives (3.1 and 3.2) outlined in section 1.2.

Chapter 8: Proposes future work, recommendations, and concluding remarks.

1.5 References

[1] Le, L., Lee, E. H., Hardy, D. J., Truong, T. N., and Schulten, K. (2010) Molecular Dynamics Simulations Suggest that Electrostatic Funnel Directs Binding of Tamiflu to Influenza N1 Neuraminidases, *PLoS Comput Biol 6*, e1000939.

[2] Organization, W. H. (2003) Pandemic preparedness.

[3] Cohen, E. (2009) When a pandemic isn't a pandemic, (CNN.com/health, Ed.), Atlanta.

[4] Doshi, P. (2011) The elusive definition of pandemic influenza, *Bull World Health Organ 89*, 532-538.

[5] Patel, R. B., Mathur, M. B., Gould, M., Uyeki, T. M., Bhattacharya, J., Xiao, Y., and Khazeni, N. (2014) Demographic and clinical predictors of mortality from highly pathogenic avian influenza A (H5N1) virus infection: CART analysis of international cases, *PloS one 9*, e91630.

[6] Prevention, C. f. D. C. a. (2014) How the Flu Virus Can Change: "Drift" and "Shift", In *Seasonal Influenza (Flu)* (Prevention, C. f. D. C. a., Ed.).

[7] Shobugawa, Y., Saito, R., Sato, I., Kawashima, T., Dapat, C., Dapat, I. C., Kondo, H., Suzuki, Y., Saito, K., and Suzuki, H. (2012) Clinical effectiveness of neuraminidase inhibitors--oseltamivir, zanamivir, laninamivir, and peramivir--for treatment of influenza A(H3N2) and A(H1N1)pdm09 infection: an observational study in the 2010-2011 influenza season in Japan, *Journal of infection and chemotherapy : official journal of the Japan Society of Chemotherapy 18*, 858-864.

[8] Han, N., Liu, X., and Mu, Y. (2012) Exploring the mechanism of zanamivir resistance in a neuraminidase mutant: a molecular dynamics study, *PloS one 7*, e44057.

[9] Kim, C. U., Lew, W., Williams, M. A., Liu, H., Zhang, L., Swaminathan, S., Bischofberger, N., Chen, M. S., Mendel, D. B., Tai, C. Y., Laver, W. G., and Stevens, R. C. (1997) Influenza neuraminidase inhibitors possessing a novel hydrophobic interaction in the enzyme active site: design, synthesis, and structural analysis of carbocyclic sialic acid analogues with potent anti-influenza activity, *Journal of the American Chemical Society 119*, 681-690.

[10] Takashita, E., Ejima, M., Itoh, R., Miura, M., Ohnishi, A., Nishimura, H., Odagiri, T., and Tashiro, M. (2014) A community cluster of influenza A(H1N1)pdm09 virus exhibiting cross-resistance to oseltamivir and peramivir in Japan, November to December 2013, *Euro surveillance : bulletin Europeen sur les maladies transmissibles = European communicable disease bulletin 19*.

[11] (2011) Global monitoring of antiviral resistance in currently circulating human influenza viruses, November 2011, *Releve epidemiologique hebdomadaire / Section d'hygiene du Secretariat de la Societe des Nations = Weekly epidemiological record / Health Section of the Secretariat of the League of Nations 86*, 497-501.

[12] Meijer, A., Rebelo-de-Andrade, H., Correia, V., Besselaar, T., Drager-Dayal, R., Fry, A., Gregory, V., Gubareva, L., Kageyama, T., Lackenby, A., Lo, J., Odagiri, T., Pereyaslov, D., Siqueira, M. M., Takashita, E., Tashiro, M., Wang, D., Wong, S., Zhang, W., Daniels, R. S., and Hurt, A. C. (2014) Global update on the susceptibility of human influenza viruses to neuraminidase inhibitors, 2012-2013, *Antiviral research 110*, 31-41.

[13] Garten, R. J., Davis, C. T., Russell, C. A., Shu, B., Lindstrom, S., Balish, A., Sessions, W. M., Xu, X., Skepner, E., Deyde, V., Okomo-Adhiambo, M., Gubareva, L., Barnes, J., Smith, C. B., Emery, S. L., Hillman, M. J., Rivailler, P., Smagala, J., de Graaf, M., Burke, D. F., Fouchier, R. A. M., Pappas, C., Alpuche-Aranda, C. M., López-Gatell, H., Olivera, H., López, I., Myers, C. A., Faix, D., Blair, P. J., Yu, C., Keene, K. M., Dotson, P. D., Boxrud, D., Sambol, A. R., Abid, S. H., St. George, K., Bannerman, T., Moore, A. L., Stringer, D. J., Blevins, P., Demmler-Harrison, G. J., Ginsberg, M., Kriner, P., Waterman, S., Smole, S., Guevara, H. F., Belongia, E. A., Clark, P. A., Beatrice, S. T., Donis, R., Katz, J., Finelli, L., Bridges, C. B., Shaw, M., Jernigan, D. B., Uyeki, T. M., Smith, D. J., Klimov, A. I., and Cox, N. J. (2009) Antigenic and Genetic Characteristics of Swine-Origin 2009 A(H1N1) Influenza Viruses Circulating in Humans, *Science 325*, 197-201.

[14] Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. (2000) The Protein Data Bank, *Nucleic Acids Research 28*, 235-242.

[15] Rose, P. W., Prlic, A., Bi, C., Bluhm, W. F., Christie, C. H., Dutta, S., Green, R. K., Goodsell, D. S., Westbrook, J. D., Woo, J., Young, J., Zardecki, C., Berman, H. M., Bourne, P. E., and Burley, S. K. (2015) The RCSB Protein Data Bank: views of structural biology for basic and applied research and education, *Nucleic Acids Res 43*, D345-356.

[16] Krug, R. M., and Aramini, J. M. (2009) Emerging antiviral targets for influenza A virus, *Trends in Pharmacological Sciences 30*, 269-277.

[17] Das, K., Aramini, J. M., Ma, L.-C., Krug, R. M., and Arnold, E. (2010) Structures of influenza A proteins and insights into antiviral drug targets, *Nature structural & molecular biology 17*, 530-538.

[18] Yuan, P., Bartlam, M., Lou, Z., Chen, S., Zhou, J., He, X., Lv, Z., Ge, R., Li, X., Deng, T., Fodor, E., Rao, Z., and Liu, Y. (2009) Crystal structure of an avian influenza polymerase PAN reveals an endonuclease active site, *Nature 458*, 909-913.

[19] Zhang, Q., Yang, J., Liang, K., Feng, L., Li, S., Wan, J., Xu, X., Yang, G., Liu, D., and Yang, S. (2008) Binding Interaction Analysis of the Active Site and Its Inhibitors for Neuraminidase (N1 Subtype) of Human Influenza Virus by the Integration of Molecular Docking, FMO Calculation and 3D-QSAR CoMFA Modeling, *Journal of chemical information and modeling 48*, 1802-1812.

[20] Duwe, S., and Schweiger, B. (2008) A new and rapid genotypic assay for the detection of neuraminidase inhibitor resistant influenza A viruses of subtype H1N1, H3N2, and H5N1, *Journal of Virological Methods 153*, 134-141.

[21] Moscona, A. (2009) Global Transmission of Oseltamivir-Resistant Influenza, *New England Journal of Medicine 360*, 953-956.

[22] Beyleveld, G., White, K. M., Ayllon, J., and Shaw, M. L. (2013) New-generation screening assays for the detection of anti-influenza compounds targeting viral and host functions, *Antiviral research 100*, 120-132.

[23] Monod, A., Swale, C., Tarus, B., Tissot, A., Delmas, B., Ruigrok, R. W., Crépin, T., and Slama-Schwok, A. (2015) Learning from structure-based drug design and new antivirals targeting the ribonucleoprotein complex for the treatment of influenza, *Expert Opinion on Drug Discovery 10*, 345-371.

[24] Das, K. (2012) Antivirals Targeting Influenza A Virus, *Journal of Medicinal Chemistry 55*, 6263-6277.

[25] Pagano, M., Castagnolo, D., Bernardini, M., Fallacara, A. L., Laurenzana, I., Deodato, D., Kessler, U., Pilger, B., Stergiou, L., Strunze, S., Tintori, C., and Botta, M. (2014) The Fight against the Influenza A Virus H1N1: Synthesis, Molecular Modeling, and Biological Evaluation of Benzofurazan Derivatives as Viral RNA Polymerase Inhibitors, *ChemMedChem 9*, 129-150.

[26] Obayashi, E., Yoshida, H., Kawai, F., Shibayama, N., Kawaguchi, A., Nagata, K., Tame, J. R. H., and Park, S.-Y. (2008) The structural basis for an essential subunit interaction in influenza virus RNA polymerase, *Nature 454*, 1127-1131.

[27] Taylor, N. R., and von Itzstein, M. (1994) Molecular Modeling Studies on Ligand Binding to Sialidase from Influenza Virus and the Mechanism of Catalysis, *Journal of Medicinal Chemistry 37*, 616-624.

[28] Gubareva, L. V., Webster, R. G., and Hayden, F. G. (2001) Comparison of the Activities of Zanamivir, Oseltamivir, and RWJ-270201 against Clinical Isolates of Influenza Virus and Neuraminidase Inhibitor-Resistant Variants, *Antimicrobial agents and chemotherapy 45*, 3403-3408.

[29] von Itzstein, M., Wu, W.-Y., Kok, G. B., Pegg, M. S., Dyason, J. C., Jin, B., Phan, T. V., Smythe, M. L., White, H. F., Oliver, S. W., Colman, P. M., Varghese, J. N., Ryan, D. M., Woods, J. M., Bethell, R. C., Hotham, V. J., Cameron, J. M., and Penn, C. R. (1993) Rational design of potent sialidase-based inhibitors of influenza virus replication, *Nature 363*, 418-423.

[30] von Itzstein, M. (2007) The war against influenza: discovery and development of sialidase inhibitors, *Nature reviews. Drug discovery 6*, 967-974.

[31] Dias, A., Bouvier, D., Crepin, T., McCarthy, A. A., Hart, D. J., Baudin, F., Cusack, S., and Ruigrok, R. W. (2009) The cap-snatching endonuclease of influenza virus polymerase resides in the PA subunit, *Nature 458*, 914-918.

[32] Huarte, M., Sanz-Ezquerro, J. J., Roncal, F., Ortín, J., and Nieto, A. (2001) PA Subunit from Influenza Virus Polymerase Complex Interacts with a Cellular Protein with Homology to a Family of Transcriptional Activators, *Journal of Virology 75*, 8597-8604.

Chapter 2

2. Background on Influenza A virus

2.1 History of Influenza A virus

The influenza virus was first isolated scientifically in 1932. However, the disease has plagued mankind for centuries, although 1173 was the first reported case of a pandemic [1]. There are two criteria that need to be met in order for an outbreak of influenza to be classed as a pandemic. Firstly, the influenza infection should be spread on a worldwide scale. Secondly, a pandemic is caused by a new influenza virus A subtype. The new virus subtype arises from antigenic shifts, resulting in a different glycoprotein, which should be resistant to immunity acquired from past seasonal influenza viruses [2].

There have been several reports of influenza pandemics from the 14th and 15th centuries [3]. The 1729 influenza outbreak lasted three years starting in Russia and claiming an undisclosed number of lives [1, 4-6]. Between 1781 - 1782, a singular strain stemming from China swept through Europe [7]. The peak of the pandemic observed a high incidence of infection in young adults. St. Petersburg recorded an infection rate of 30,000 people per day, the disease infected approximately two-thirds of Rome's population. The greatest "medical holocaust" in history was the influenza pandemic of 1918 – 1920, which claimed approximately 39.3 million lives, impacting North America, India, Africa, Australia and Europe [8]. The most recent outbreak was the 2009-H1N1 influenza pandemic which affected 74 countries. The novel H1N1 virus preferentially infected younger people below the age of 25 years, elderly people and immunocompromised individuals. The estimated number of deaths ranged between 151,700 and 575,400 in its first year of infection [9]. The spread of infection is eminent and from literature the frequency of influenza pandemics has increased. Thus it is important to understand the mechanism for the propagation of disease [10, 11]. In doing so, specific biomolecular targets influencing viral maturation can be identified as potential inhibitory sites. Inhibition of these proteins or enzymes could prevent future influenza virus pandemics [10].

2.2 Life cycle and mechanism of structural components of influenza A virus

The basic life cycle of the virus occurs by attaching itself to the host cell surface receptors. It then enters the cell with the subsequent uncoating of the viral nucleic acid. At this point the host cellular machinery is "hijacked" into replicating the viral genome. The new viral components synthesized are assembled into progeny virions which are released into the host. The sequential cellular functioning of the virus is governed by specialised protein species encapsulated within the virus structure [11].

The influenza A virus is pleomorphic *i.e.* it is spherical or elongated in shape. It contains three surface glycoproteins, embedded in a lipid bilayer, an artefact of the host cell (**Figure 1**). The glycoproteins are: 1) hemagglutinin (HA); 2) neuraminidase (NA); and 3) the M2 proton ion channel. Virus replication is initiated by the binding of HA protein to sialic acid residue receptors contained on the host cell surface (**Figure 2**). The binding of the entities mediates fusion of the viral envelope forming an endosome after receptor-mediated endocytosis. The NA protein plays an integral role in the latter stages of infection. Neuraminidase is responsible for the release of newly assembled viral progeny from the cell surface. The enzyme cleaves the sialic acid residues from sialyloligosaccharides. This process prevents self-aggregation of the virus particles. The HA and NA proteins also serve as a characterisation tool for the subtypes of the influenza virus. There are 17 genetically distinct subtypes for HA (*i.e.* H1 – H17) and nine for NA (*i.e.* N1 – N9), which differentiate influenza viral species (*e.g.* H1N1, H5N1, H3N2 and H7N9).
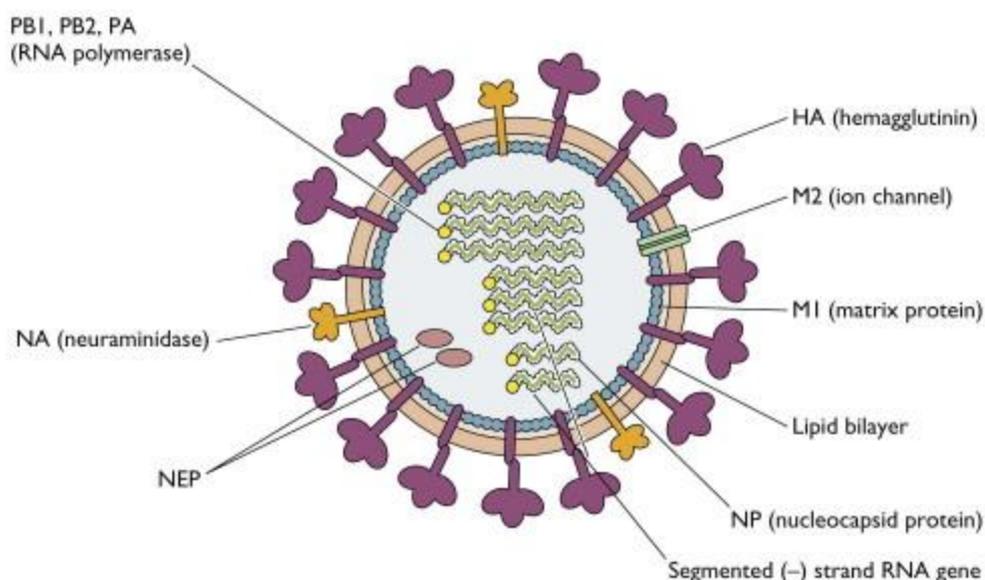
Figure 1: Structure of influenza virus[12]

At the interface of the endosomal wall and the inside of the virus, beneath the bilipid layer lies a critical structural M1 matrix protein. A ribonucleoprotein (RNP) complex lies inside the virus envelope. This is typically characterised by eight segmented, negative-strand RNA genomes; the RNA-dependent RNA polymerase (PB1, PB2 and PA) and the nucleoprotein (NP). The third glycoprotein, the M2 ion channel, promotes the influx of H^+ ions into the virus. This lowers the internal pH of the virus which facilitates binding of the virus to the endosomal wall and the release of viral ribonucleoproteins into the host cell. Once the viral RNPs have infiltrated the cell, they are immediately transported into the nucleus where viral protein synthesis takes place. Because of the lack of self-checking systems during transcription of viral proteins, point mutations are prevalent. Newly synthesized viral proteins are transported from the nucleus to a site in close proximity to the cell membrane for self-assembly. The process of self-assembly has been termed "budding", resulting in the formation of virions. These virions (as previously mentioned), are released into the host and allowed to infect new cells [11, 13-15].
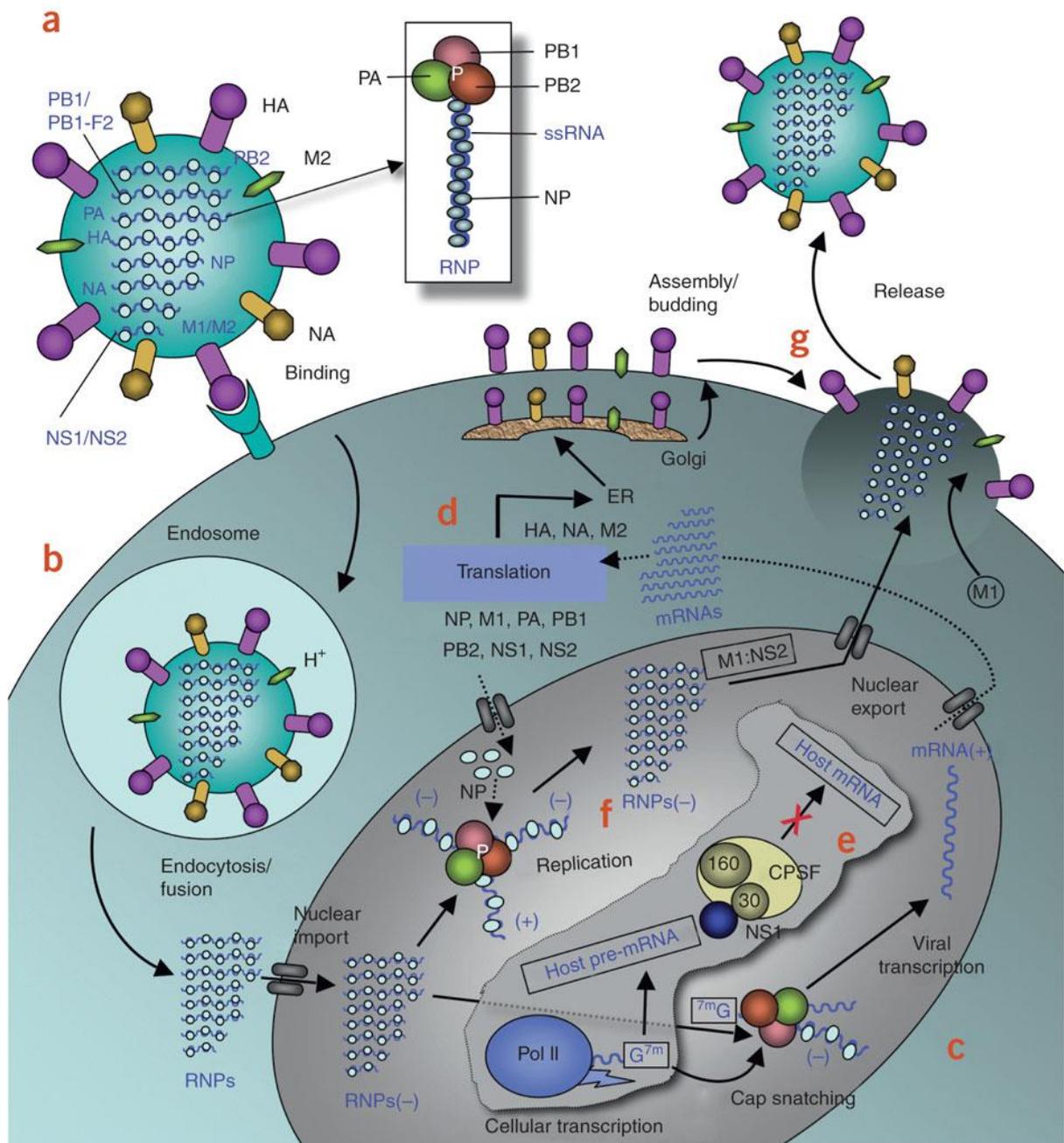
Figure 2: Influenza A life cycle (adapted from the work of Das *et al.* [16])

Figure 2, description (**a**) Influenza A virus has a lipid bilayer envelope, within which are eight RNA genomic segments, each of which is associated with the trimeric viral RNA polymerase (PB1, PB2, PA) and coated with multiple nucleoproteins (NPs) to form the vRNPs. The outer layer of the lipid envelope is spiked with multiple copies of HA, NA and a small number of M2, whereas the M1 molecules keep vRNPs attached to the inner layer. (**b**) The viral surface glycoprotein HA binds to the host cell-surface sialic acid receptors, and the virus is transported into the cell in an endocytic vesicle. The low pH in the endosome triggers a conformational change in the HA protein that leads to fusion of the viral and endosomal membranes. The low pH also triggers the flow of protons into the virus via the M2 ion channel, thereby dissociating the vRNPs from M1 matrix proteins. The

13

vRNPs that are released into the cytoplasm are transported into the nucleus by recognition of the nuclear localization sequences (NLSs) on nucleoproteins[17] only when the M1 molecules are dissociated. (**c**) In the nucleus, the viral polymerase initiates viral mRNA synthesis with 5′-capped RNA fragments cleaved from host pre-mRNAs. The PB2 subunit binds the 5′ cap of host pre-mRNAs[18], and the endonuclease domain in PA subunit cleaves the pre-mRNA 10–13 nucleotides downstream from the cap[19]. Viral mRNA transcription is subsequently initiated from the cleaved 3′ end of the capped RNA segment[19,20]. This 'cap snatching' occurs on nascent pre-mRNAs. (**d**) Viral mRNAs are transported to the cytoplasm for translation into viral proteins. The surface proteins HA, M2 and NA are processed in the endoplasmic reticulum (ER), glycosylated in the Golgi apparatus and transported to the cell membrane. (**e**) The NS1 protein of influenza A virus serves a critical role in suppressing the production of host mRNAs by inhibiting the 3′-end processing of host pre-mRNAs [21, 22], consequently blocking the production of host mRNAs, including interferon-β mRNAs. Unlike host pre-mRNAs, the viral mRNAs do not require 3′-end processing by the host cell machinery. Therefore, the viral mRNAs are transported to the cytoplasm, whereas the host mRNA synthesis is predominantly blocked. (**f**) The viral polymerase is responsible for not only capped RNA-primed mRNA synthesis but also unprimed replication of vRNAs in steps (−) vRNA → (+) cRNA → (−) vRNA. The nucleoprotein molecules are required for these two steps of replication and are deposited on the cRNA and vRNA during RNA synthesis[23]. The resulting vRNPs are subsequently transported to the cytoplasm, mediated by a M1–NS2 complex that is bound to the vRNPs; NS2 interacts with human CRM1 protein that exports the vRNPs from the nucleus[24]. (**g**) The vRNPs reach the cell membrane to be incorporated into new viruses (reviewed in ref.[25]) that are budded out. The HA and NA proteins in new viruses contain terminal sialic acids that would cause the viruses to clump together and adhere to the cell surface. The NA of newly formed viruses cleaves these sialic acid residues, thereby releasing the virus from the host cell. (Adapted from the work of Kaylan *et al.* [16])

2.3 Drug targets, prevention and treatment of infection

Seasonal influenza infections occur as a result of antigenic drifts. An antigenic drift is a small change in the genes of influenza viruses. This occurs continually over time due to viral replication. Thus, as a remedial therapy, vaccinations are used. Unfortunately, this does not prevent infection by pandemic influenza strains and has no effect on those who have contracted a pandemic virus. Influenza pandemics, as mentioned earlier, occur as a result of antigenic shifts. Monitoring of antigenic variation/s in the influenza virus is imperative in preparing for an epidemic or a pandemic. Due to the limitations in preventing potential influenza pandemics, anti-viral chemoprophylaxes offers an alternate method of treatment.

2.3.1 M2 ion channel

The M2 protein is a small homotetrameric transmembrane (TM) protein (**Figure 3**). The protein is comprised of three components:

1) An N-terminal extracellular domain of 24 amino acid residues,
2) A signal anchor transmembrane domain of approximately 19 residues in length, and
3) A 54 residue cytoplasmic tail [26].

The oligo-tetramers are stabilised by disulphide bonds. Its primary function is to channel protons into the virion and host cell, to adjust the pH of intracellular compartments. Little is known about the functional mechanism of the protein. With the aid of computational techniques Sharma *et al.* [27], suggested that a cluster exists in the heart of the TM domain. It comprises residues His to Trp41, and thought to chaperone the protons through the channel. This is achieved by hydrogen bond formation and breaking between the "His-Trp gate". However, it was Wu and Voth [28], who initially suggested an open and closed state of the M2 proton channel. This phenomenon may explain the mechanism of conductance of protons. Thus, in the TM domain lies the channel pore, which is the site for inhibition by anti-influenza virus drugs

FDA-approved anti-viral drugs amantadine (**Figure 3**) and rimantadine, function by blocking the M2 ion channel pore [29]. Co-workers Sansom and Kerr [30], designed a molecular model of the M2 channel. By virtue of energy interaction profiles, they were able to describe the mode of inhibition of amantadine. Two distinct residues S31 and I42, which are positioned at opposite ends of the pore, were implicated in facilitating the inhibition of the protein. Inhibiting the activity of the M2 ion channel, impedes the transport of H^+ ions into the endosome. This affects the proteins ability to adjust the pH, required for the release of RNPs into the host cell which ultimately inhibits viral replication. The activity of the modulator is regulated by changes in pH.

Although these drugs have demonstrated their effectiveness, a major drawback is the severe side effects. Because of these side effects, patients do not complete the prescribed dosages, leading to the emergence of drug resistant influenza viruses with increased virulence. Resistance is acquired by point mutations in the M2 protein. Huang *et al.* [31] and Jing *et al.* [32], made significant contribution to the structural basis of drug-resistance of rimantadine and amantadine, using high resolution NMR studies. They characterised the mechanism of action

of point mutations: D44N, D44A, N44D and R45A; in the TM domain. It was elucidated that mutation R45A specifically interferes with the binding of rimantadine, whilst mutation D44A partially inhibits the binding of amantadine and rimantadine. The presence of the point mutation exerted a conformational stress on the pore preventing either drug from binding inside the channel pore [32].
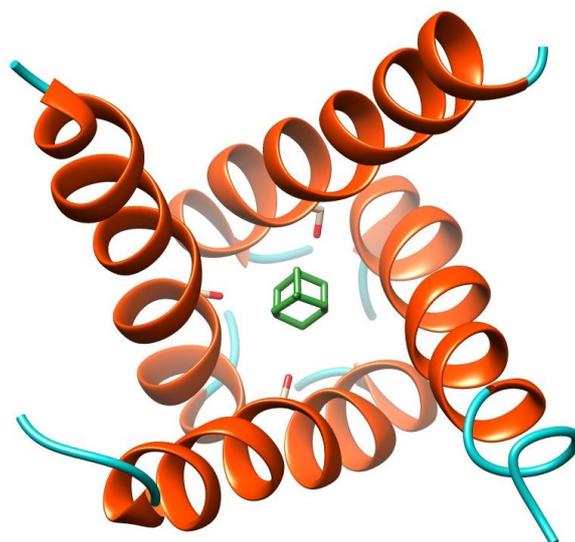
Figure 3: Crystal structure of the transmembrane domain (orange) of the M2 protein complexed with amantadine (green) (PDB code: 3C9J) [33]

2.3.2 Hemagglutinin

Hemagglutinin (HA), is a trimeric, rod shaped glycoprotein which comprises the envelope surrounding the influenza virus [34]. The abundancy of HA on the surface membrane is four times that of neuraminidase, thus making HA an appealing therapeutic target. The glycoprotein is synthesised as a single polypeptide chain (HA0), by post-translational modification the protein is cleaved into two subunits, HA1 and HA2 (**Figure 4**). These subunits are linked covalently by a disulphide bridge. The HA protein facilitates a pH dependent fusion of the endosome to the host cell surface and the release of the viral RNPs into the host cell.

The HA1 protein subunit forms the base of the structure *i.e.* the globular domain. It is responsible for the binding of the virus to the sialic acid receptors. HA1 is the primary antigenic determinant recognised to elicit an immune response. The HA2 protein, is characterised by the fibrous stem. The N-terminus of the HA2 subunit anchors the glycoprotein into the bilipid

layer. This feature is distinguished by 20 consecutive, hydrophobic residues and is essential for membrane fusion.
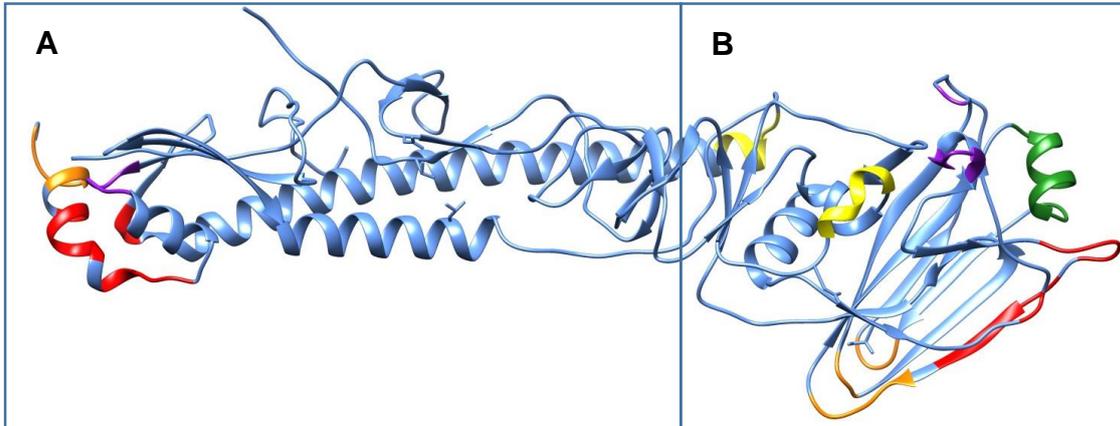
Figure 4: Crystal structure of influenza A hemagglutinin (HA0); A) the HA2 protein (fibrous stem) and B) the HA1 protein (globular head) (PDB code: 2VIU) [34]

(Sa site-green); (Sb site-red); (Ca1 site-orange); (Ca2 site-purple) and (Cb site-yellow)

Viral replication occurs at a high rate, which compromises the accuracy in which proteins are transcribed and translated. The lack of proofreading mechanisms allows antigenic variations of HA to emerge. Therefore unravelling the efforts of vaccinations and antibodies created from immune memory. A variety of HA proteins exists (*i.e.* H1 – H17), each surface antigen is conformationally unique. However, only H1, H2 and H5 subtypes affect the human population.

Analysis of HA sequence variations of isolates causing epidemics have resulted in the identification of five antigenic sites, categorised as being strain specific (S) or common (C). There are two strain specific, the Sa and Sb sites (**Figure 4**) [35]. These antigenic sites are positioned near the tip of HA protein at residues co-ordinates: 128 – 129, 156 – 160, 162 – 167 and 187-198. Three common sites have been isolated: Ca1, Ca2 and Cb. The Ca1 and Ca2 sites are located between adjacent HA monomers mid-way down the globular base. The residue co-ordinates of the Ca1 are: 169 - 173, 206 - 208 and 238 - 240; whilst the Ca2 sites resides within residues 140 - 145 and 224 - 225. The Cb antigenic site resides in the base of the HA1 protein at residue co-ordinate 74 - 79. The vast combinations of antigenic variants imposes great difficulty in the discovery of new drug.

The FDA have approved the administration of the HA protein vaccine FluBlok, in the prevention of infection. However, there are no clinical therapeutic agents that inhibit HA protein activity, in an effort to treat an influenza viral infection. Despite this much effort has

been dedicated to discovery of potential inhibitors. A series of serine protease inhibitors have been identified as potential HA inhibitors. Important inhibitors include:

1. ε-aminocaproic acid [36],
2. Nafamostat [37],
3. pulmonary surfactant (lipoprotein complex) [38], and
4. mucous protease inhibitors [39], have been implicated in inhibiting the cleavage of HA precursors.

Other drug candidates have been identified as promising inhibitors. However, none have yet reached the stage of clinical approval [40-42].

2.3.3 Neuraminidase

Neuraminidase is a tetrameric enzyme (**Figure 5**). Each monomer has a molecular mass of ~60 kDa and comprises three domains:

1) a cytoplasmic head (resides on the virion surface),
2) a transmembrane stem, and
3) a cytoplasmic tail.

Neuraminidases' enzyme activity lies in its cytoplasmic head. The residues implicated in the catalytic site include: Arg118, Asp151, Arg152, Arg224, Glu276, Arg292, Arg371 and Tyr406. The listed residues of various influenza A virus subtypes is largely conserved. This feature only enhances the attractiveness of neuraminidase as target for the treatment of influenza. A calcium binding site is present near the catalytic domain. It serves to stabilise the enzyme structure at low pH. The calcium domain is enclosed within residues 293, 347, 111 - 115 and 139 - 143. This pocket is held together by the interaction of carbonyl oxygens of the αC backbone, specifically residues 297, 345 and 348; as well as the oxygen of Asp324's carboxyl group. Information regarding the three-dimensional structure of the cytoplasmic tail and transmembrane stem is limited. This is further incentive to investigate neuraminidase as a target of influenza [43].
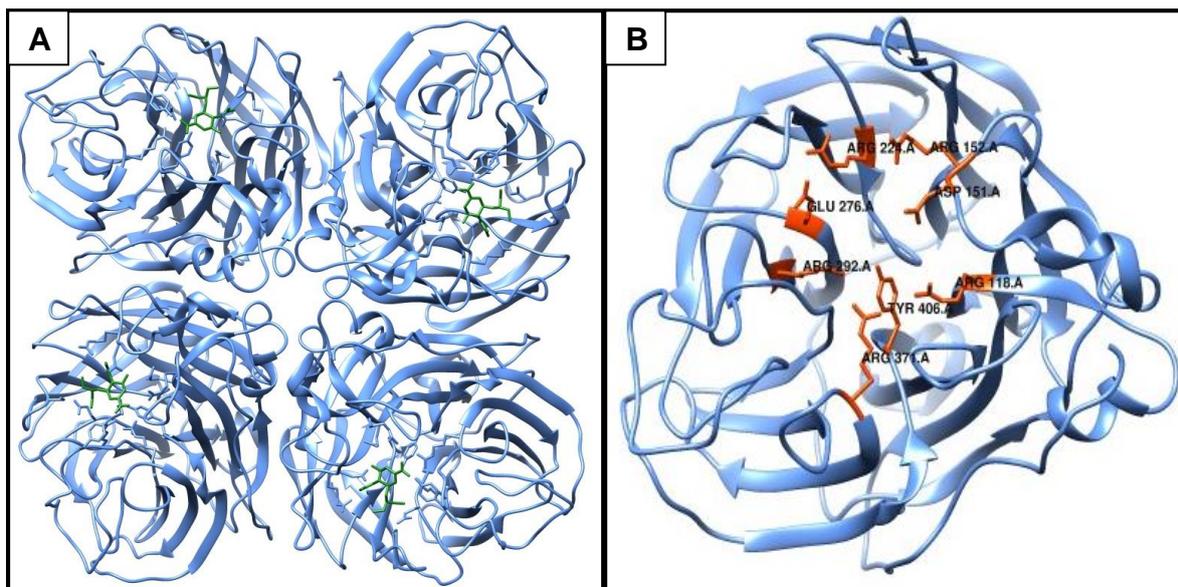
Figure 5: Crystal structure of neuraminidase N1; (A) tetramer with oseltamivir bound, and (B) monomer highlighting active site residues (orange) (PDB code: 2HU4) [44]

As previously mentioned, the catalytic site of influenza A subtypes share a conversed active site profile. As a result, current and potential anti-viral drugs may exert a broader spectrum of inhibition. Inhibition of the NA protein impedes the proliferation of new virions into the host. Prominent researcher Mark von Itzstein, made one the first attempts of using molecular modelling tools in understanding the binding landscape of neuraminidase [45]. With the aid of co-worker, he proposed the rational design of novel anti-influenza drugs [46, 47]. The principles outlining the design of inhibitors are still employed in the discovery of new and improved drug therapies. Von Itzstein et al., described the structural and energetic aspects of the substrate and potential NA inhibitor binding and mechanism of action of influenza neuraminidase. It was found that the two potential NA inhibitors:

1) DANA (2-deoxy-2,3-didehydro-*N*-acetylneuramicnic acid), and
2) an *N*-trifluoroacetyl analogue of FANA,

demonstrated good activity *in vitro* but lacked clinical efficacy [13, 45]. However, these were essential discoveries as they served as leads in the development of FDA-approved drugs: oseltamivir (Tamiflu), and zanamivir (Relenza) [15]. During the 2006 virulent strain of H5N1-avian influenza, the chemotherapeutic agent, oseltamivir, failed to treat the virus. A similar scenario was observed in 2009, with the pandemic strain of H1N1, a point mutation of H274Y was observed to significantly reduce the efficacy of the drug. The emergence of resistance of

viruses of the N1 subtype, placed the human population at risk of future pandemics. Hence, the design of new inhibitors peramivir, which, was made available in the United States by Emergency Use Authorisation (EUA) [48], as well as laninamivir.

There have been accounts of computational investigations performed:

1) to understand the drug resistance mechanisms [49, 50],
2) to determine the point mutations inflicting resistance [50, 51], and
3) investigate the development of enhanced anti-viral agents [52, 53].

However, none have comparatively and comprehensively identified the structural and conformational impact of specific point mutations. With subsequent proposal of compound scaffolds able to bind to the newly evolved active site of neuraminidase.

2.3.4 RNA-dependent RNA polymerase

The viral RNA polymerase enzyme is a heterotrimer of ~250 kDa. It is characterised by eight negative-sense single-stranded RNA segments which encode 10 essential proteins. The enzyme is solely responsible for catalysing RNA replication and viral protein transcription in the infected nuclei of host cells [54]. The three largest RNA segments encode the RNA-dependent RNA polymerase protein subunits:

1) polymerase acidic protein (PA);
2) polymerase basic protein 1 (PB1); and
3) polymerase basic protein 2 (PB2).

The three medium sized RNA segments, when translated reveal the nucleoprotein (NP). The NP coupled with the polymerase subunits and viral RNA segments (vRNA), form the ribonucleoprotein (RNP). RNPs are released into the host cytoplasm once fusion of the viral and endosomal membranes, occurs. The two remain short vRNA segments encode:

1) the matrix protein (M1) that lines the inner membrane surface of the viral bilipid layer,
2) the NS1 protein which participates in evading the innate immune response of the host, and
3) the NS2 or nuclear export protein responsible for chaperoning RNPs from the nucleus into the cytoplasm.

There are no proofreading mechanisms available during RNA polymerase activity. As a result, the antigenic variants of the encoded proteins, occurs at a rate nearly equivalent to the rate of viral replication. A statistical evaluation of this process estimated approximately 10,000 potentially new viral mutants can be produced to infect healthy cells. Viral RNA polymerase is pivotal to viral evolution, inhibition of the enzymes catalytic activity will threaten the survival of the influenza virus. The distinct structure and activity of the enzyme is exclusive to the influenza virus composition and cannot be found in the human genome. The selectivity and promise of the complete extinction of the viral infection emphasizes the RNA polymerase as a crucial target for new anti-viral drugs [54].

The PA subunit has been benchmarked as the primary target of inhibition. This is attributed to the N-terminal domain of PA encompassing the endonuclease activity of RNA polymerase. The PA and PB2 protein subunits compose the cap-binding domain, which is involved in the transcription cap-snatching process. Each subunit projects inwardly toward one another across a solvent channel. The 5' terminal of the polymerase enzyme folds over to form 'hook'. The hook-like structure creates a second highly conserved binding pocket between residues PB1 and PA [55]. The arrangement of the active site and folding of the PA domain resembles the architecture of the PD-(D/E)XK family of nucleases [56].

The endonuclease activity of the PA subunit is responsible for the cleavage of a phosphodiester bond 10 – 13 nucleotides downstream of the cap. The process of cleaving the 'cap' is mediated by divalent metal centers. The metal serves as a thermal stabiliser and shields the enzyme against fluctuations in pH. Structure-based mutagenesis studies have been performed to determine the impact of induced point mutations in the PA active site, on the endonuclease activity. It was established that the endonuclease-independent RNA replication function remained intact, whilst the endonuclease-dependent transcription was inhibited [57]. There have been reports of anti-cancer agents demonstrating dual activity by inhibiting the endonuclease activity of influenza A virus RNA polymerase [58]. A known inhibitor of the influenza endonuclease is 2,4-dioxo-4-phenylbutanoic acid (**Figure 6**) [59, 60]. However, to date there have been no clinically approved chemotherapeutic agents. Thus the need for new drug therapies is imperative [60, 61].
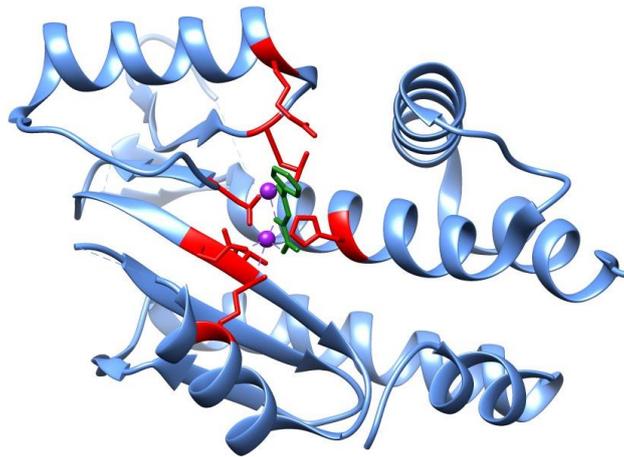
Figure 6: Crystal structure of Influenza A PA endonuclease subunit in complex with 2,4-dioxo-4-phenylbutanoic acid (green) (PDB code: 4AWF) [60]

The active site residues have been highlighted in red and Mn^{2+} metal centers coloured purple.

2.4 Current Limitations

The first 21st century influenza pandemic originated from a genetically divergent form of the H1N1 virus. Since 2009, there have been an estimated $42 - 86$ million cases of infection worldwide. Highly pathogenic avian H5N1 influenza A viruses have become transmissible in humans by overcoming species barriers. There have been reported infections in Asia, Europe and Africa, many with fatal outcomes. Despite the availability of and treatment with current drug therapies, the H1N1 virus (including pandemic 2009 H1N1) and H5N1 maintain virulence and have developed resistance toward active inhibitors. This is of growing concern and emphasises the need for new antiviral drugs.

2.5 References

[1] Hirsch, A. (1883) *Handbook of Geographical and Historical Pathology*, New Sydenham Society, London.

[2] Potter, C. W. (2001) A history of influenza, *J Appl Microbiol 91*, 572-579.

[3] Molineux, D. (1694) Dr. Molineux's Historical Account of the Late General Coughs and Colds; with Some Observations on Other Epidemick Distempers, *Philosophical Transactions 18*, 105-111.

[4] Finkler, D. (1899) *Influenza in twentieth century practice*.

[5] Gold, E. (1987) Pandemic influenza 1700-1900: A study in historical epidemiology, *JAMA 257*, 2656-2656.

[6] Thacker, S. B. (1987) The diffusion of influenza: Patterns and paradigms, *JAMA 258*, 389-389.

[7] Beveridge, W. I. (1991) The chronicle of influenza epidemics, *Hist Philos Life Sci 13*, 223-234.

[8] Johnson, N. P., and Mueller, J. (2002) Updating the accounts: global mortality of the 1918-1920 "Spanish" influenza pandemic, *Bull Hist Med 76*, 105-115.

[9] Dawood, F. S., Iuliano, A. D., Reed, C., Meltzer, M. I., Shay, D. K., Cheng, P.-Y., Bandaranayake, D., Breiman, R. F., Brooks, W. A., Buchy, P., Feikin, D. R., Fowler, K. B., Gordon, A., Hien, N. T., Horby, P., Huang, Q. S., Katz, M. A., Krishnan, A., Lal, R., Montgomery, J. M., Mølbak, K., Pebody, R., Presanis, A. M., Razuri, H., Steens, A., Tinoco, Y. O., Wallinga, J., Yu, H., Vong, S., Bresee, J., and Widdowson, M.-A. Estimated global mortality associated with the first 12 months of 2009 pandemic influenza A H1N1 virus circulation: a modelling study, *The Lancet Infectious Diseases 12*, 687-695.

[10] Palese, P. (2004) Influenza: old and new threats, *Nat Med*.

[11] Watanabe, T., Watanabe, S., and Kawaoka, Y. (2010) Cellular Networks Involved in the Influenza Virus Life Cycle, *Cell Host & Microbe 7*, 427-439.

[12] Racaniello, V. (2009) Structure of influenza virus, In *Virology blog*, Vincent Racaniello.

[13] Palese, P., Shaw, M. L., Knipe, D. M., and Howley, P. M. (2007) *Fields Virology*, Vol. 2.

[14] Skehel, J. J., and Wiley, D. C. (2000) Receptor binding and membrane fusion in virus entry: the influenza hemagglutinin, *Annu. Rev. Biochem. 69*, 531.

[15] Lathrop, S. P. Replication Cycle of the Influenza Virus and Recent Syntheses of Nueraminidase Inhibitor Oseltamivir (TamifluTM).

[16] Das, K., Aramini, J. M., Ma, L.-C., Krug, R. M., and Arnold, E. (2010) Structures of influenza A proteins and insights into antiviral drug targets, *Nature structural & molecular biology 17*, 530-538.

[17] Wu, W. W. H., and Panté, N. (2009) The directionality of the nuclear transport of the influenza A genome is driven by selective exposure of nuclear localization sequences on nucleoprotein, *Virology Journal 6*, 68-68.

[18] Ulmanen, I., Broni, B. A., and Krug, R. M. (1981) Role of two of the influenza virus core P proteins in recognizing cap 1 structures (m(7)GpppNm) on RNAs and in initiating

viral RNA transcription, *Proceedings of the National Academy of Sciences of the United States of America 78*, 7355-7359.

[19] Plotch, S. J., Bouloy, M., Ulmanen, I., and Krug, R. M. (1981) A unique cap(m7GpppXm)-dependent influenza virion endonuclease cleaves capped RNAs to generate the primers that initiate viral RNA transcription, *Cell 23*, 847-858.

[20] Hagen, M., Chung, T. D., Butcher, J. A., and Krystal, M. (1994) Recombinant influenza virus polymerase: requirement of both 5' and 3' viral ends for endonuclease activity, *J Virol 68*, 1509-1515.

[21] Shimizu, K., Iguchi, A., Gomyou, R., and Ono, Y. (1999) Influenza Virus Inhibits Cleavage of the HSP70 Pre-mRNAs at the Polyadenylation Site, *Virology 254*, 213-219.

[22] Nemeroff, M. E., Barabino, S. M. L., Li, Y., Keller, W., and Krug, R. M. Influenza Virus NS1 Protein Interacts with the Cellular 30 kDa Subunit of CPSF and Inhibits 3′ End Formation of Cellular Pre-mRNAs, *Molecular Cell 1*, 991-1000.

[23] Newcomb , L. L., Kuo, R.-L., Ye, Q., Jiang, Y., Tao, Y. J., and Krug, R. M. (2009) Interaction of the Influenza A Virus Nucleocapsid Protein with the Viral RNA Polymerase Potentiates Unprimed Viral RNA Replication, *Journal of Virology 83*, 29-36.

[24] Neumann, G., Hughes, M. T., and Kawaoka, Y. (2000) Influenza A virus NS2 protein mediates vRNP nuclear export through NES-independent interaction with hCRM1, *EMBO J 19*, 6751-6758.

[25] Nayak, D. P., Hui, E. K.-W., and Barman, S. (2004) Assembly and budding of influenza virus, *Virus Research 106*, 147-165.

[26] Holsinger, L. J., Nichani, D., Pinto, L. H., and Lamb, R. A. (1994) Influenza A virus M2 ion channel protein: a structure-function analysis, *J Virol 68*, 1551-1563.

[27] Sharma, M., Yi, M., Dong, H., Qin, H., Peterson, E., Busath, D. D., Zhou, H.-X., and Cross, T. A. (2010) Insight into the Mechanism of the Influenza A Proton Channel from a Structure in a Lipid Bilayer, *Science 330*, 509-512.

[28] Wu, Y., and Voth, G. A. (2005) A Computational Study of the Closed and Open States of the Influenza A M2 Proton Channel, *Biophysical Journal 89*, 2402-2411.

[29] Jefferson, T., Deeks, J. J., Demicheli, V., Rivetti, D., and Rudin, M. (2004) Amantadine and rimantadine for preventing and treating influenza A in adults, *The Cochrane database of systematic reviews*, Cd001169.

[30] Sansom, M. S. P., and Kerr, I. D. (1993) Influenza virus M2 protein: a molecular modelling study of the ion channel, *Protein Engineering 6*, 65-74.

[31] Huang, R.-B., Du, Q.-S., Wang, C.-H., and Chou, K.-C. (2008) An in-depth analysis of the biological functional studies based on the NMR M2 channel structure of influenza A virus, *Biochemical and biophysical research communications 377*, 1243-1247.

[32] Jing, X., Ma, C., Ohigashi, Y., Oliveira, F. A., Jardetzky, T. S., Pinto, L. H., and Lamb, R. A. (2008) Functional studies indicate amantadine binds to the pore of the influenza A virus M2 proton-selective ion channel, *Proceedings of the National Academy of Sciences 105*, 10967-10972.

[33] Stouffer, A. L., Acharya, R., Salom, D., Levine, A. S., Di Costanzo, L., Soto, C. S., Tereshko, V., Nanda, V., Stayrook, S., and DeGrado, W. F. (2008) Structural basis for the function and inhibition of an influenza virus proton channel, *Nature 451*, 596-599.

[34] Fleury, D., Wharton, S. A., Skehel, J. J., Knossow, M., and Bizebard, T. (1998) Antigen distortion allows influenza virus to escape neutralization, *Nature structural biology 5*, 119-123.

[35] Sriwilaijaroen, N., and Suzuki, Y. (2012) Molecular basis of the structure and function of H1 hemagglutinin of influenza virus, *Proceedings of the Japan Academy, Series B 88*, 226-249.

[36] Kido, H., Yokogoshi, Y., Sakai, K., Tashiro, M., Kishino, Y., Fukutomi, A., and Katunuma, N. (1992) Isolation and characterization of a novel trypsin-like protease found in rat bronchiolar epithelial Clara cells. A possible activator of the viral fusion glycoprotein, *J Biol Chem 267*, 13573-13579.

[37] Hosoya, M., Matsuyama, S., Baba, M., Suzuki, H., and Shigeta, S. (1992) Effects of protease inhibitors on replication of various myxoviruses, *Antimicrobial agents and chemotherapy 36*, 1432-1436.

[38] Kido, H., Sakai, K., Kishino, Y., and Tashiro, M. (1993) Pulmonary surfactant is a potential endogenous inhibitor of proteolytic activation of Sendai virus and influenza A virus, *FEBS Letters 322*, 115-119.

[39] Beppu, Y., Imamura, Y., Tashiro, M., Towatari, T., Ariga, H., and Kido, H. (1997) Human Mucus Protease Inhibitor in Airway Fluids Is a Potential Defensive Compound against Infection with Influenza A and Sendai Viruses, *Journal of Biochemistry 121*, 309-316.

[40] Shen, X., Zhang, X., and Liu, S. (2013) Novel hemagglutinin-based influenza virus inhibitors, *J Thorac Dis 5 Suppl 2*, S149-159.

[41] Basu, A., Antanasijevic, A., Wang, M., Li, B., Mills, D. M., Ames, J. A., Nash, P. J., Williams, J. D., Peet, N. P., Moir, D. T., Prichard, M. N., Keith, K. A., Barnard, D. L., Caffrey, M., Rong, L., and Bowlin, T. L. (2014) New small molecule entry inhibitors targeting hemagglutinin-mediated influenza a virus fusion, *J Virol 88*, 1447-1460.

[42] Yang, Z., Yang, G., and Zhou, L. (2013) Mutation effects of neuraminidases and their docking with ligands: a molecular dynamics and free energy calculation study, *Journal of computer-aided molecular design 27*, 935-950.

[43] Shtyrya, Y. A., Mochalova, L. V., and Bovin, N. V. (2009) Influenza virus neuraminidase: structure and function, *Acta Naturae 1*, 26-32.

[44] Russell, R. J., Haire, L. F., Stevens, D. J., Collins, P. J., Lin, Y. P., Blackburn, G. M., Hay, A. J., Gamblin, S. J., and Skehel, J. J. (2006) The structure of H5N1 avian influenza neuraminidase suggests new opportunities for drug design, *Nature 443*, 45-49.

[45] Taylor, N. R., and von Itzstein, M. (1994) Molecular Modeling Studies on Ligand Binding to Sialidase from Influenza Virus and the Mechanism of Catalysis, *Journal of Medicinal Chemistry 37*, 616-624.

[46] von Itzstein, M., Dyason, J. C., Oliver, S. W., White, H. F., Wu, W.-Y., Kok, G. B., and Pegg, M. S. (1996) A study of the active site of influenza virus sialidase: an approach to the rational design of novel anti-influenza drugs, *Journal of medicinal chemistry 39*, 388-391.

[47] von Itzstein, M., Wu, W.-Y., Kok, G. B., Pegg, M. S., Dyason, J. C., Jin, B., Phan, T. V., Smythe, M. L., White, H. F., Oliver, S. W., Colman, P. M., Varghese, J. N., Ryan, D. M., Woods, J. M., Bethell, R. C., Hotham, V. J., Cameron, J. M., and Penn, C. R. (1993) Rational design of potent sialidase-based inhibitors of influenza virus replication, *Nature 363*, 418-423.

[48] Jain, S., and Fry, A. M. (2011) Peramivir: Another Tool for Influenza Treatment?, *Clinical Infectious Diseases 52*, 707-709.

[49] Wang, S.-Q., Du, Q.-S., and Chou, K.-C. (2007) Study of drug resistance of chicken influenza A virus (H5N1) from homology-modeled 3D structures of neuraminidases, *Biochemical and biophysical research communications 354*, 634-640.

[50] Karthick, V., Shanthi, V., Rajasekaran, R., and Ramanathan, K. (2012) Exploring the cause of oseltamivir resistance against mutant H274Y neuraminidase by molecular simulation approach, *Applied biochemistry and biotechnology 167*, 237-249.

[51] Collins, P. J., Haire, L. F., Lin, Y. P., Liu, J., Russell, R. J., Walker, P. A., Skehel, J. J., Martin, S. R., Hay, A. J., and Gamblin, S. J. (2008) Crystal structures of oseltamivir-resistant influenza virus neuraminidase mutants, *Nature 453*, 1258-1261.

[52] Du, Q.-S., Wang, S.-Q., and Chou, K.-C. (2007) Analogue inhibitors by modifying oseltamivir based on the crystal neuraminidase structure for treating drug-resistant H5N1 virus, *Biochemical and biophysical research communications 362*, 525-531.

[53] An, J., Lee, D. C., Law, A. H., Yang, C. L., Poon, L. L., Lau, A. S., and Jones, S. J. (2009) A novel small-molecule inhibitor of the avian influenza H5N1 virus determined through computational screening against the neuraminidase, *Journal of medicinal chemistry 52*, 2667-2672.

[54] Boivin, S., Cusack, S., Ruigrok, R. W. H., and Hart, D. J. (2010) Influenza A Virus Polymerase: Structural Insights into Replication and Host Adaptation Mechanisms, *Journal of Biological Chemistry 285*, 28411-28417.

[55] He, X., Zhou, J., Bartlam, M., Zhang, R., Ma, J., Lou, Z., Li, X., Li, J., Joachimiak, A., Zeng, Z., Ge, R., Rao, Z., and Liu, Y. (2008) Crystal structure of the polymerase PA(C)-PB1(N) complex from an avian influenza H5N1 virus, *Nature 454*, 1123-1126.

[56] Knizewski, L., Kinch, L. N., Grishin, N. V., Rychlewski, L., and Ginalski, K. (2007) Realm of PD-(D/E)XK nuclease superfamily revisited: detection of novel families with modified transitive meta profile searches, *BMC Struct. Biol. 7*, 40.

[57] Yuan, P., Bartlam, M., Lou, Z., Chen, S., Zhou, J., He, X., Lv, Z., Ge, R., Li, X., Deng, T., Fodor, E., Rao, Z., and Liu, Y. (2009) Crystal structure of an avian influenza polymerase PAN reveals an endonuclease active site, *Nature 458*, 909-913.

[58] Kuzuhara, T., Iwai, Y., Takahashi, H., Hatakeyama, D., and Echigo, N. (2009) Green tea catechins inhibit the endonuclease activity of influenza A virus RNA polymerase, *PLoS Currents 1*, RRN1052.

[59] Tomassini, J. E. (1996) A novel antiviral agent which inhibits the endonuclease of influenza viruses, *Antimicrob. Agents Chemother. 40*, 1189-1193.

[60] Kowalinski, E., Zubieta, C., Wolkerstorfer, A., Szolar, O. H. J., Ruigrok, R. W. H., and Cusack, S. (2012) Structural Analysis of Specific Metal Chelating Inhibitor Binding to the Endonuclease Domain of Influenza pH1N1 (2009) Polymerase, *PLoS pathogens 8*, e1002831.

[61] DuBois, R. M., Slavish, P. J., Baughman, B. M., Yun, M.-K., Bao, J., Webby, R. J., Webb, T. R., and White, S. W. (2012) Structural and Biochemical Basis for Development of

Influenza Virus Inhibitors Targeting the PA Endonuclease, *PLoS pathogens 8*, e1002830.

Chapter 3

3. Theoretical and Computational Methods

3.1 Introduction to Computational Chemistry

Computational chemistry (also referred to as molecular modelling), is the amalgamation of science and art. It is a technique that combines computational investigation and design, as well as the fundamentals of chemistry from an atomistic and molecular perspective; to generate realistic molecular models. Molecular modelling has been an essential tool in drug design and understanding the biology of biomolecules [1]. It investigates methodically:

1) the three-dimensional molecular structure;
2) molecular dynamics and flexibility;
3) physical properties;
4) biological activities such as protein folding, stability and conformation.

This is achieved by the using the theoretically principles of physics, mathematics and chemistry; to describe the energetics of a three-dimensional biomolecular system.

There are three basic components of used to describe molecular modelling:

1) Molecular Mechanics (MM);
2) Quantum Mechanics (QM); and
3) Molecular dynamics simulation.

Quantum mechanics is a technique based on first principles used to describe the behaviour of photons, electrons and other particles that compose the universe. The Born-Oppenheimer approximation is used to solve the innate characteristics and mechanistic reactions of biomolecular systems.

Molecular mechanics apply Newtonian mechanics to model molecular systems. The potential energy of the system is estimated by using force fields. Unlike QM, MM ignores the electrons and computes the energy of the system as a function of nuclear positions *i.e.* an implicit method of parameterisation for potential energy approximations. This serves as a useful tool for biochemical and biomedical analyses. Hybrid QM/MM techniques have been employed to reduce computational resources in which biomolecular mechanism models are built.

Molecular dynamics (MD) simulations are trajectory-based methods, designed to be less computationally expensive. MD simulations offer a visual aspect to the motions of atoms using the classical laws of physics.

In this thesis MD simulations were primarily used to:

1) study the conformational changes of proteins,
2) the impact of point mutations in a protein, and
3) propose reasonable explanations for drug resistance.

The three concepts of computational chemistry are further explained, so as to give insight into the relatedness of the energy descriptors.

3.2 Quantum Mechanics

Quantum mechanics is a branch of classical physics. The first version of quantum mechanics was initiated by Werner Heisenberg, Max Born and Pascal Jordan who discovered matrix mechanics. The invention of quantum theory was, however, not single handedly begotten. It was through the great efforts and contributions of principal researchers: Max Planck, Wolfgang Pauli, Heisenberg, Born, Jordan, Erwin Schrödinger, Niels Bohr, Paul Dirac and Albert Einstein [2].

The atomic particles (electrons of an atom) are treated as little packets of energy *i.e.* quantum particles. The energy state of the different atomic particles are associated with unique quantum numbers. The assignment of quantum numbers to the atomic particles are governed by particle characteristics such as; the spin of particle (*i.e.* the intrinsic angular momentum) and molecular electronic states. The change in energy either by release or addition, occurs in the form of a photon which is computed into a wave.

Analysis of a biomolecule by QM involves the arrangement of nuclei and corresponding electrons, in a three-dimensional space. The electrons are mapped using the continuous electronic density method, and the energetics measured are solved by the Schrödinger equation. For larger molecular systems the Born-Oppenheimer approximation coupled with the Hartree-Fock self-consistent field which maps the electron density of the system.

3.2.1 The Schrödinger equation

In 1925 Erwin Schrödinger described the evolution of wave functions by a differential equation[3]. There are two derivative of the Schrödinger equation they are:

1) The time-dependent Schrödinger equation gives a description of a system that evolves with time,

$$i\hbar \frac{\partial}{\partial t} \psi(r,t) = \hat{H}\psi(r,t) \qquad eqn. 1$$

Where i is the imaginary unit, $\hbar$ is Planck's constant divided 2π, $\frac{\partial}{\partial t}$ represents the partial derivative with respect to time, Ψ (psi) represents the wave function of the quantum system, and $\hat{H}$ is the Hamiltonian operator which characterises the total energy and is dependent on the wave function and form of the system.

2) The time-independent Schrödinger equation which predicts standing waves which translate to atomic and molecular orbitals,

$$E\Psi = \hat{H}\Psi \qquad eqn. 2$$

The term E, represents the energy of state. The Hamiltonian operator is governed by a specific wave function.

When using either equation the solution remains ambiguous, thus certain assumptions are required to obtain an approximate answer.

3.2.2 Born-Oppenheimer approximation

The Born-Oppenheimer approximation is one of the most important approximations in molecular QM. It was developed in 1927 by Max Born and J. Robert Oppenheimer. The basis of the approximation is built on the assumption that the motion of electrons and nuclei can be separated [4].

The theory is derived from the consideration of the mass of atomic particles and the influence they have on one another. There exists an interaction between the electron and nucleus, this interaction constitutes the fundamental character of an atom. Supposing one component either the electron or nucleus, was by mass heavier than the other. The motion of the lighter particle would thus be dominated and steered by the heavier particle [5]. Translating this phenomenon

into a wave function, the heavier particle would exert and interaction force on the lighter particle causing it to oscillate more rapidly around the heavier particle, in order to stay with the sphere of interaction. This can be a considered a compensatory effect of the lighter particle. Electrons naturally have a higher diffusion constant which make them lighter than the nucleus. Thus in the Born-Oppenheimer approximation the nuclear positions are fixed. The approximation is as follows [4]:

$$\Psi^{(mol)}(x, R) = \Psi^{(elec)}(x, R)\Psi^{(nuc)}(R) \qquad eqn.\, 3$$

R represents the nuclear position, and $\Psi^{(elec)}(x, R)$ the electronic wave function solved by the nuclear position.

The total classical energy of the molecule is calculated as follows:

$$K_e + K_n + V_{ee}(r) + V_{en}(r, R) + V_{nn}(R) = E \qquad eqn.\, 4$$

K_e and K_n measure the electronic and nuclear kinetic energies. The symbol V is representative of the potential energy. It must be noted that the Born-Oppenheimer approximation is dependent on the nuclear wave function and not the velocity.

3.3 Molecular Mechanics

Molecular mechanics (also known as force field calculations), is designed to compute large datasets. It has been the preferred method for the estimation of energies of biomolecules. As a result it used to elucidate structural details of the MD ensembles. The mathematical model analyses the nuclear position of atoms whilst ignoring the electronic contribution. Based on the previous discussion regarding the Born-Oppenheimer approximation, the nuclear position is regarded as fixed due to mass differentials. In order for the electron approximations to be ignored, the model has to recognise the electronic component as one unit. The unit sum of the electrons can be compensated for in the potential energy function [6].

Molecular mechanics is a classical approach to molecular modelling. The atoms in this approach are represented as balls and the bonds elastic sticks. Structures acquired from x-ray crystal data or NMR studies, requires refinement by energy minimisation. The atoms with a molecule are not static, they are in constant motion due to the bond type. As a result biomolecules may adopt a many conformations of differing energies. In some instances, the protein may have an average of its conformations at a local or global energy minimum, which

are separated by activation energy barriers. There are different forces which act upon the system when being prepared. Force fields are used to solve the energy equation [7, 8].

3.3.1 MM Potential energy function

The potential energy (PE) function is a mathematical characterisation of the molecules' geometry by minimisation, and the nuclear probability distribution by solving the nuclear Schrodinger equation. It is capable of determining nuclear dynamics, with the aid of solving the time-dependent Schrodinger equation. A solid understanding of the distribution of electrons in molecules is essential. An additional approximation derived from the PE function, involves the treatment of nuclei as classical particles with a potential energy. Using the description of Newton's Second Law, the nuclei would move as a classic point particle [9].

The potential energy function is thus described as follows [10]:

$$U(\vec{R}) = \underbrace{\sum_{bonds} k_i^{bond} (r_i - r_0)^2}_{U_{bond}} + \underbrace{\sum_{angles} k_i^{angles} (\theta_i - \theta_0)^2}_{U_{angle}}$$

$$+ \underbrace{\sum_{dihedrals} k_i^{dihedrals} [1 + \cos(n_i\phi_i + \delta_i]}_{U_{dihedral}}$$

$$+ \underbrace{\sum_i \sum_{j \neq i} 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^6 \right] + \sum_i \sum_{j \neq i} \frac{q_i q_j}{\epsilon r_{ij}}}_{U_{nonbond}}$$

Where, U_{bond} measures the oscillations about the equilibrium, U_{angle} defines the oscillations of three atoms about an equilibrium bond angle, $U_{dihedral}$ expresses the torsional rotation of four atoms about a central bond and $U_{nonbond}$ is the descriptor of all non-bonding energy terms (*i.e.* electrostatics and Lenard-Jones interactions).

3.3.2 Force Fields

There are a variety of biomolecular system force fields available. A few examples include:

1) AMBER [11],
2) CHARMM [12],
3) GROMOS [13],
4) OPLS-AA [14],

In each of the aforementioned force fields, the biomolecule is characterised as a collective mass interacting with one another through harmonic forces. The atoms in the molecule are depicted as balls of differing sizes joined together by springs i.e. bonds, of variable strength, length and angular distances. In the studies conducted, the AMBER force field was implemented to characterise the enzymes, whilst GAFF was used to interpret the ligands.

This AMBER force field provides a favourable balance in energy between the helical and extended regions of the peptide and protein backbones with improved dihedral torsions [15]. Further to this the parameter set achieves better agreement of the dihedral contributions of the conformational model and experimental NMR data.

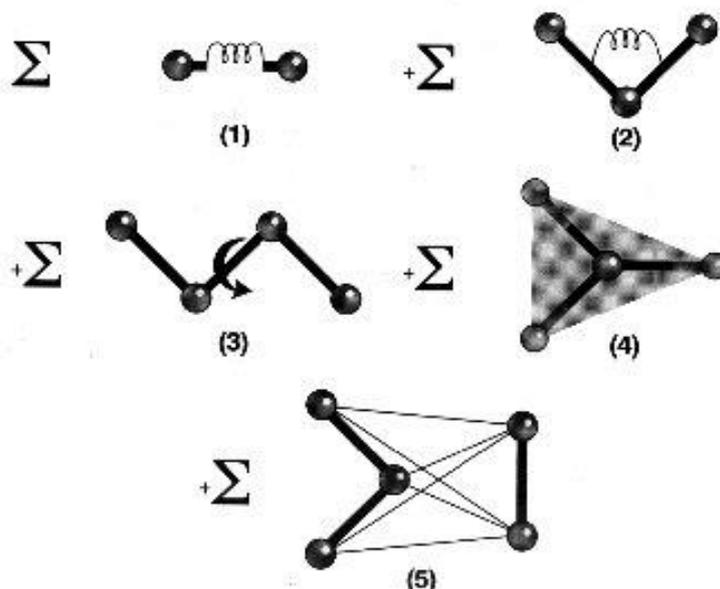**Figure 7:** A diagrammatic representation expressing the summation of total energy in a molecule [6]

Bonding terms: (1) bond stretching, (2) bond bending, (3) bond dihedrals. Non-bonding terms: (4) electrostatic, and (5) van der Waals interactions

The total energy (E_{tot}), of a molecule is computed from the deviation in energy contributions of different bond types (eqn. 1), *i.e.* bonding and non-bonding interactions (**Figure 7**) [16, 17]. The harmonic bond type terms include:

1) bond stretching energy (E_{str})- describes the energy change as bonds stretch and contract
2) angle-bending energy (E_{bend})- describes the energy change of 'spring-like' or 'scissor-like' bond harmonic
3) torsional energy (E_{tors})- is a measure of the dihedral potential energy
4) van der Waals energy (E_{vdw})- is the energetic contributions of van der Waals interactions
5) electrostatic energy (E_{elec})- is the energetic characterisation of the Coulombic charges

$$E_{tot} = E_{str} + E_{bend} + E_{tors} + E_{vdw} + E_{elec} \dots \qquad eqn.\, 5$$

Total energy can be described as "a measure of intramolecular strain" of a three-dimensional biomolecular model in equilibrium, which is useful when comparing different conformations.

3.4 Molecular dynamics (MD) simulation

MD simulations estimate the correlated motions of a protein during a trajectory. A single trajectory is capable of generating three ensembles of which the conformations extracted are closer to the x-ray crystal structure. Molecular dynamics simulations were predominantly used in this study due to their excellent approximation of biomolecular systems [7, 18, 19].

Classical molecular dynamics (MD), is an integrated computed algorithm system which incorporates Newton's equations of motion. Molecular dynamic simulation can provide high probability real-time view of many chemical reactions and/or biological systems. Based on highly-evolved mathematical and physical criteria, the technique proposes possible conformations or mechanisms of action of systems on an atomic and molecular scale.

Unlike other 2D analyses such as NMR and X-ray, MD is not a static technique and the resulting data is a snapshot of its motion at that moment. However, MD simulation has proven an accurate and incomparable tool in estimating phenomena or unique occurrences. The importance of understanding the three-dimensional nature of a bound and unbound protein structures would prove invaluable.

MD simulation allows us to study a complex environment to predict the time evolution of a system of interacting particles (atoms, molecules, granules, etc.) with the aid of manipulating parameters such as temperature, velocity, flexibility of molecule and runtime. The essence of the technique is to solve classical equations in an attempt to understand the energies and molecular mechanics of a particular network. We are able to specify the initial velocity, conformation or positioning of our system, as well as we can direct the interaction potential. Computing Newton's equations of motion, MD trajectories of complex biomolecular are generated [20]:

$$F_i = m_i \frac{d^2 r_i(t)}{dt^2} \qquad eqn.\, 6$$

Where $r_i(t)$ represents the co-ordinate of the particle position vector (i), t = time, m_i is the mass of the particle and F_i depicts the force acting on (i) particle at a specific t and m_i.

There are advantages of running MD simulations, as all the inputs of a particular model are derived from crystal structures or well-constructed homology models which provide a description of the interatomic and intermolecular interactions. Further to this, the results derived from the simulation are predominantly unbiased and made to mimic a real-time system with no assumptions. Critical information can be extracted from the MD ensembles generated, with regards to the atomic and molecular interactions [21].

3.4.1 Molecular dynamics solvent parameters

Solvated biomolecular systems are best parameterised using the Particle-Mesh Ewald method (PME). When using this method the protein is embedded in an explicit solvent which leads to stable trajectories. The method has proven to be an accurate and efficient computational tool, used to calculate the three-dimensional Coulomb interactions with two-dimensional periodicity [22-24]. The PME method separates the potential energy into Ewald's standard direct (U^r) and the reciprocal sums (U^m). Conventional Gaussian charge distributions are used to describe the electrostatic potential. The direct evaluation of the sum of point charges *i.e.* U^r is measured explicitly using cut-offs, whilst U^m term is calculated using the Fast Fourier Transform (FFT), which solves the three-dimensional Coulombic charges into a two-dimensional grid. This process reduces the memory and time required for PME estimation, whereby the differentiation in energy can be computed analytically rather than interpolated.

3.4.2 Post-dynamics Analysis

The post-dynamics techniques and calculations, discussed in this section have been used to describe the protein features in the studies of this thesis. Post-dynamics analysis of trajectories obtained from the MD simulation is imperative in determining the:

1) relative binding free energies,
2) three-dimensional conformations and configurations, and
3) to determine the state of thermodynamic equilibrium of the biomolecular system [25].

3.4.2.1 Convergence

Convergence is an empirical description of protein dynamics. It is based on the fluctuations of bond type and bond angle vibrations, such that time magnitude and time range (from angstroms to a femto second to seconds) during the unfolding of a protein [26]. The point of convergence is often regarded as the stage at which the system enters a state of equilibrium such that in a graphical representation the curve is reported to plateau. It is the point at which optimum distance from the reference structure is reached exhibiting the most energetically stable conformations of our protein-ligand system.

The root mean square deviation (RMSD), provides invaluable information about the structural equilibrium of a simulated system. Essential information derived from RMSD calculations, related to how long the system took to equilibrate can determine the number of conformations that exist for the biomolecule [27]. It may highlight biomolecules with 'open' and 'closed' conformations. RMSD values may also shed light on the type of sampling of a structure whether it be extensive or non-convergent. It can also aid in monitoring the atoms in the backbone and identifying the evolution of trajectories. RMSD of a trajectory is calculated by continual measurement of the spatial difference between two static structures, *i.e.* a reference structure and the structure at that point in time of simulation [28].

The evaluation of a systems potential energy (PE), during an MD simulation, confers information about the stability and functional motions of the structure. Due to potential energy being quantised by velocity and time, a biomolecule takes to reach a state of equilibrium, different convergence behaviour may be monitored by calculating the ergodicity [29].

3.4.2.2 Conformational and structural analysis

The manner in which a protein folds influences the conformational freedom of that protein. Thus it is important to evaluate the secondary structure of protein. These determinations can be achieved by the following approximations:

1) RMSF

The root mean square fluctuation (RMSF) captures, for each atom, the fluctuation about its average position. Providing insight into the flexibility of regions of the protein that corresponds to the crystallographic β-factors (temperature factors) [27].

2) PCA

Principal component analysis (PCA), is a technique which is based on the reduction of size or dimensionality of a data set, which originally consists of a large number of interrelated variables whilst retaining as much of the variation as possible. Reduction in the sample size is achieved by transforming a new set of variables, *i.e.* the principal components, which are marked as uncorrelated [30, 31]. However, they are ordered in such a way that the first few components retain most of the variation present in all original variables. Thus for the purposes of our study a restricted analysis to the first two modes was instituted. A useful tool in acquiring information with regards to the conformational freedom of proteins is covariance. Conformational freedom implicates the potential energy in the stability and functional motions of proteins. Equations 7 and 8 below, offer a classical means for the numerical diagnosis of flexibility *i.e.* covariance, within a system.

$$cov(X,Y) = \frac{1}{n-1}\left(\sum_{i}^{n}(X_i - \bar{X})(Y_i - \bar{Y})\right) \qquad (eqn.\,7)$$

$$cov(X,Y) = \frac{1}{n-1}\left(\sum_{i}^{n}X_iY_i - n\bar{X}\bar{Y}\right) \qquad (eqn.\,8)$$

Covariance is denoted as, $cov(X,Y)$; and defined by the size of the data set (n), the mean of X and Y values $(\bar{X}\bar{Y})$, and the sum product of data values X and Y $(\sum_{i}^{n}X_iY_i)$. An additional feature which allows the identification and fair comparison of protein during a trajectory is correlation. Correlation (eqn. 9), concatenates the strength of linear association between two variables, which are quantified as being independent of one another. Correlation (r_{XY}), is not governed by any measure or unit inflicted by the system. This is accomplished by relating the covariance $(Cov(X,Y))$ and variance of X and Y $(Var(X)Var(Y))$ [32, 33].

$$r_{XY} = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}} \qquad\qquad (eqn.\,9)$$

3) Solvent accessible surface area

The solvent accessible surface area (SASA), is a derivative of the area over which absolute contact between the proteins' Van der Waals surface and the solvent, arises. This feature imparts information relative to the compactness of the structure as well as the extent of hydrophobicity in the interior of the folded protein [34].

4) Radius of gyration

The radius of gyration (Rg) is defined as the root mean square distance from each atom of the protein to their centroid. It is classed as a qualitative assurance technique of MD simulations providing an understanding of the compactness of a biomolecule. The compactness has been defined as the ratio of the accessible surface area of a protein to the surface area of an ideal sphere of the same volume [35]. The Rg is independent of the protein size. The Rg and folding rate of a protein, share an inversely proportional relationship, e.g. A high radius of gyration value indicates a loosely-packed structure.

3.4.2.3 Binding free energy calculations

The binding free energy of MD ensembles can be calculated using different algorithmic approaches. Some examples include:

1) The free energy perturbation estimation,
2) The Thermodynamic integration,
3) The Linear interaction energy calculation,
4) The Molecular Mechanics/Generalised Born Surface Area (MM/GBSA) calculation,
5) The Molecular Mechanics/Poisson-Boltzmann Surface Area (MM/PBSA) calculation, and
6) The Molecular docking calculation.

The MM/GBSA and MM/PBSA free energy calculations infer information based on protein structure determinations. The techniques employ the principles of molecular mechanics, by using an implicit solvent model to solve the binding free energy

computation. Evaluation of ligand-protein complexes using MM/GBSA or MM/PBSA, provides a quantitative analysis of the binding affinity of the ligand to the protein. The use of MM in these approaches, allows for averaged energy estimations to be made over certain frames from the MD trajectory to derive an estimate of the absolute binding free energy. The total binding free energy (ΔG_{bind}), in equations 10 and 11, is defined by the collective contributions of:

7) the gas-phase energy (E_{gas}) term,
8) the solvation free energy (G_{sol}) term, and
9) multiple of the temperature (T) in Kelvins and entropy (*S*) approximation

$$\Delta G_{bind} = G_{complex} - G_{receptor} - G_{ligand} \qquad (eqn.\,10)$$

$$\Delta G_{bind} = E_{gas} + G_{sol} - TS \qquad (eqn.\,11)$$

$$E_{gas} = E_{int} + E_{vdw} + E_{ele} \qquad (eqn.\,12)$$

$$G_{sol} = G_{GB/PB} + G_{SA} \qquad (eqn.\,13)$$

$$G_{SA} = \gamma SASA \qquad (eqn.\,14)$$

G_{SA}, represents the non-polar solvation energy term and G_{GB} describes the polar solvation energy. Internal potential terms such as bond types, bond angles, torsional stresses and dihedral angles, influence the potential energy stability of the protein complex, which ultimately affect the binding free energy value.

From the two methods MM/GBSA has better accuracy and determination capability of the ligand-protein binding free energy [36]. Thus it has been the preferred calculation used in the studies.

3.5 References

[1] Ooms, F. (2000) Molecular modeling and computer aided drug design. Examples of their applications in medicinal chemistry, *Current medicinal chemistry 7*, 141-158.
[2] Kleppner, D., and Jackiw, R. (2000) One hundred years of quantum physics, *Science 289*, 893-898.

[3] Weinsstein, E. W. (2007) Schrödinger equation In *Quantum Mechanics* (Weisstein, E., Ed.).

[4] Born, M. (1927) Born-Oppenheimer Approximation, *Quantum 2*, 4.

[5] Niklasson, A. M. (2008) Extended born-oppenheimer molecular dynamics, *Physical review letters 100*, 123004.

[6] Molecular Modeling - An Introduction to Molecular Mechanics

[7] Weinhold, F. (2001) Chemistry: A new twist on molecular shape, *Nature 411*, 539-541.

[8] MacKerell Jr, A. D., and Nilsson, L. (2008) Molecular dynamics simulations of nucleic acid–protein complexes, *Current Opinion in Structural Biology 18*, 194-199.

[9] Pophristic, V., and Goodman, L. (2001) Hyperconjugation not steric repulsion leads to the staggered structure of ethane, *Nature 411*, 565-568.

[10] (2004) Force Fields for MD simulations, In *Parameters for Classical Force Fields*

[11] D.A. Case, T. A. D., T.E. Cheatham, III, C.L. Simmerling, J. Wang, R.E. Duke, R. Luo, R.C. Walker, W. Zhang, K.M. Merz, B. Roberts, S. Hayik, A. Roitberg, G. Seabra, J. Swails, A.W. Goetz, I. Kolossváry, K.F. Wong, F. Paesani, J. Vanicek, R.M. Wolf, J. Liu, X. Wu, S.R. Brozell, T. Steinbrecher, H. Gohlke, Q. Cai, X. Ye, J. Wang, M.-J. Hsieh, G. Cui, D.R. Roe, D.H. Mathews, M.G. Seetin, R. Salomon-Ferrer, C. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko, P.A. Kollman. (2012) AMBER 12, University of California.

[12] Brooks, B. R., Brooks, C. L., 3rd, Mackerell, A. D., Jr., Nilsson, L., Petrella, R. J., Roux, B., Won, Y., Archontis, G., Bartels, C., Boresch, S., Caflisch, A., Caves, L., Cui, Q., Dinner, A. R., Feig, M., Fischer, S., Gao, J., Hodoscek, M., Im, W., Kuczera, K., Lazaridis, T., Ma, J., Ovchinnikov, V., Paci, E., Pastor, R. W., Post, C. B., Pu, J. Z., Schaefer, M., Tidor, B., Venable, R. M., Woodcock, H. L., Wu, X., Yang, W., York, D. M., and Karplus, M. (2009) CHARMM: the biomolecular simulation program, *J Comput Chem 30*, 1545-1614.

[13] Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R. D., Kale, L., and Schulten, K. (2005) Scalable molecular dynamics with NAMD, *J Comput Chem 26*, 1781-1802.

[14] Christen, M., Hunenberger, P. H., Bakowies, D., Baron, R., Burgi, R., Geerke, D. P., Heinz, T. N., Kastenholz, M. A., Krautler, V., Oostenbrink, C., Peter, C., Trzesniak, D., and van Gunsteren, W. F. (2005) The GROMOS software for biomolecular simulation: GROMOS05, *J Comput Chem 26*, 1719-1751.

[15] Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A., and Simmerling, C. (2006) Comparison of multiple Amber force fields and development of improved protein backbone parameters, *Proteins: Structure, Function, and Bioinformatics 65*, 712-725.

[16] Salomon-Ferrer, R., Case, D. A., and Walker, R. C. (2013) An overview of the Amber biomolecular simulation package, *Wiley Interdisciplinary Reviews: Computational Molecular Science 3*, 198-210.

[17] Ponder, J. W., and Case, D. A. (2003) Force Fields for Protein Simulations, In *Advances in Protein Chemistry*, pp 27-85, Academic Press.

[18] Clementi, E. (2005) Chapter 6 - Computational chemistry: Attempting to simulate large molecular systems A2 - Scuseria, Clifford E. DykstraGernot FrenkingKwang S. KimGustavo E, In *Theory and Applications of Computational Chemistry*, pp 89-114, Elsevier, Amsterdam.

[19] Kapral, R., and Ciccotti, G. (2005) Chapter 16 - Molecular dynamics: An account of its evolution A2 - Scuseria, Clifford E. DykstraGernot FrenkingKwang S. KimGustavo E, In *Theory and Applications of Computational Chemistry*, pp 425-441, Elsevier, Amsterdam.

[20] Mirjalili, V., Noyes, K., and Feig, M. (2013) Physics-based protein structure refinement through multiple molecular dynamics trajectories and structure averaging, *Proteins: Structure, Function and Bioinformatics 82*.

[21] Zhigilei, L. Molecular dynamics, In *Introduction to Atomistic Simulations*, University of Virginia.

[22] Petersen, H. G. (1995) Accuracy and efficiency of the particle mesh Ewald method, *The Journal of chemical physics 103*, 3668-3679.

[23] Cheatham, T. E., Miller, J. L., Fox, T., Darden, T., and Kollman, P. A. (1995) Molecular dynamics simulations on solvated biomolecular systems: the particle mesh Ewald method leads to stable trajectories of DNA, RNA, and proteins, *Journal of American Chemical Society 117*, 4193-4194.

[24] Kawata, M., and Nagashima, U. (2001) Particle mesh Ewald method for three-dimensional systems with two-dimensional periodicity, *Chemical Physics Letters 340*, 165-172.

[25] Karplus, M., and McCammon, J. A. (2002) Molecular dynamics simulations of biomolecules, *Nature 9*, 646.

[26] Galindo-Murillo, R., Roe, D. R., and Cheatham, T. E., 3rd. (2015) Convergence and reproducibility in molecular dynamics simulations of the DNA duplex d(GCACGAACGAACGAACGC), *Biochimica et biophysica acta 1850*, 1041-1058.

[27] Knapp, B., Frantal, S., Cibena, M., Schreiner, W., and Bauer, P. (2011) Is an Intuitive Convergence Definition of Molecular Dynamics Simulations Solely Based on the Root Mean Square Deviation Possible?, *Journal of Computational Biology 18*, 997-1005.

[28] Schreiner, W., Karch, R., Knapp, B., and Ilieva, N. (2012) Relaxation Estimation of RMSD in Molecular Dynamics Immunosimulations, *Computational and Mathematical Methods in Medicine 2012*, 9.

[29] Moyano, G. E., and Collins, M. A. (2004) Molecular potential energy surfaces by interpolation: strategies for faster convergence, *The Journal of chemical physics 121*, 9769-9775.

[30] Cocco, S., Monasson, R., and Weigt, M. (2013) From principal component to direct coupling analysis of coevolution in proteins: low-eigenvalue modes are needed for structure prediction, *PLoS Comput Biol 9*, e1003176.

[31] Altis, A., Nguyen, P. H., Hegger, R., and Stock, G. (2007) Dihedral angle principal component analysis of molecular dynamics simulations, *The Journal of chemical physics 126*, 244111.

[32] Shlens, J. (2005) A tutorial on Principal Component Analysis, pp 1-13, University of California, San Diego.

[33] Jolliffe, I. T. (2002) *Principal Component Analysis*, Springer New York.

[34] Richmond, T. J. (1984) Solvent accessible surface area and excluded volume in proteins: Analytical equations for overlapping spheres and implications for the hydrophobic effect, *Journal of Molecular Biology 178*, 63-89.

[35] Lobanov, M. Y., Bogatyreva, N. S., and Galzitskaya, O. V. (2008) Radius of gyration as an indicator of protein structure compactness, *Molecular Biology 42*, 623-628.

[36] Hou, T., Wang, J., Li, Y., and Wang, W. (2011) Assessing the performance of the MM/PBSA and MM/GBSA methods. 1. The accuracy of binding free energy calculations based on molecular dynamics simulations, *Journal of chemical information and modeling 51*, 69-82.

Chapter 4

Anti-cancer Glycosidase Inhibitors from Natural Products: A Computational and Molecular Modelling Perspective

Ashona Singh[a], Ndumiso Mhlongo[a] and Mahmoud E.S. Soliman[a,b,*]

[a]*School of Health Sciences, University of KwaZulu-Natal, Westville, Durban 4001, South Africa;*

[b]*Department of Pharmaceutical Organic Chemistry, Faculty of Pharmacy, Zagazig University, Zagazig 44519, Egypt*

Abstract: The implementation of computational tools in pharmaceutics has proven an effectual strategy in creating harmony between the physical and chemical aspects of proteins and potential inhibitors. This is achieved by bringing to life the three dimensional retrospect of biological systems, which takes into consideration computational approaches such as quantum mechanics and molecular dynamics to facilitate drug design and discovery. In this work, we aim to provide a summary of the computational aspects of naturally derived anti-cancer inhibitors targeting the enzyme family of glucosidase's. Our study offers insight into the evolution of drug discovery, molecular modelling and molecular binding modes of natural product inhibitors associated with glycosidase enzymes.

Keywords: Anti-cancer inhibitors, glycosidase, molecular modelling, natural product.

INTRODUCTION

Natural Products in Drug Discovery

Since the beginning of time man has exploited the use of plant and animal based products for their medicinal application. Ancient Eastern traditional healers have documented the healing properties of roots, bulbs, flowers and other herbs like ginseng, for at least 4,500 years [1]. European healers dating back to the 10th century used foxglove (*Digitalis purpurea* L.), which was found to contain the active ingredient digitoxin which is effective in the management of congestive heart failure, along with its many structure based analogues [2]. The pharmacological use of natural products was, however, only established in Western culture by the 19th century the most famous being the synthesis of the anti-inflammatory, aspirin, derived from salicin, isolated from the bark of the willow tree (*Salix alba* L.). By 1803 the alkaloid morphine was isolated from opium poppy (*Papaver somniferum* L.) which had been used by

the Sumerians and Ancient Greeks, and dubbed by the Arabs as being addictive. Later during the 1870's crude morphine treated with acetic anhydride yielded heroin, which was found to readily convert to codeine, a painkiller [2, 3]. The track record of nature lending itself as a curative, preventative or symptomatic treatment has been our only saving grace in eradicating infection and alleviating disease and other syndromes. The current status of pharmaceutics worldwide indicates that approximately 40% of all dispensable medicines are either natural or semi-synthetic analogues of natural derivatives [4]. We are only at the precipice in drug discovery in terms of natural products and have yet to uncover and explore hidden secrets in different species genome using genome sequencing or single molecule real-time methods. Such discovery would allow new introspection in our management of disease, how some organisms remain immune to certain infections, as well as it would provide a new mechanism by which alternate biocatalysts and natural products may be identified. The bulk of natural product derivatives currently in clinical study are dedicated to the treatment of cancer [5]. There are a multitude of natural products sanctioned as leads in anticancer, there exists four main plant based drug classes, these include: 1) vinca alkaloids, 2) epipodophyllotoxins, 3) taxanes, and 4) camptothecines. Each class targets different stages or pathways of the cell cycle of cancerous cells. The alkaloids vinblastine and vincristine (Fig. **1**), block mitosis with metaphase arrest by binding to tubulin resulting in its depolymerisation. In combination with other cancer chemotherapeutic drugs the alkaloids can be used in the treatment of a variety of cancers including leukaemia's, lymphomas, advanced testicular cancer, breast and lung cancer, and Kaposi's sarcoma. Podophyllotoxin derivatives etoposide and teniposide (Fig. **2**), bind to tubulin, leading to DNA strand degradation irreversibly inhibiting topoisomerase II. These two derivatives are used in the treatment of lymphomas, as well as bronchial and testicular cancers. The class of taxane, which include paclitaxel (taxol®) and other like derivatives act by disrupting the assembly of tubulin, they show specific activity in patients diagnosed with breast, ovarian and non-small cell lung cancer. Camptothecin analogues selectively inhibit topoisomerase I. The more effective and safer analogues being topotecan, an ovarian and small cell lung cancer chemotherapy and irinotecan, used in patients with colorectal cancer (Fig. **3**). There exists many other natural products implicated in the treatment of different cancers, which only amplifies the cause for continued research efforts toward investigating natural products and their derivatives for their medicinal activity [6-8].

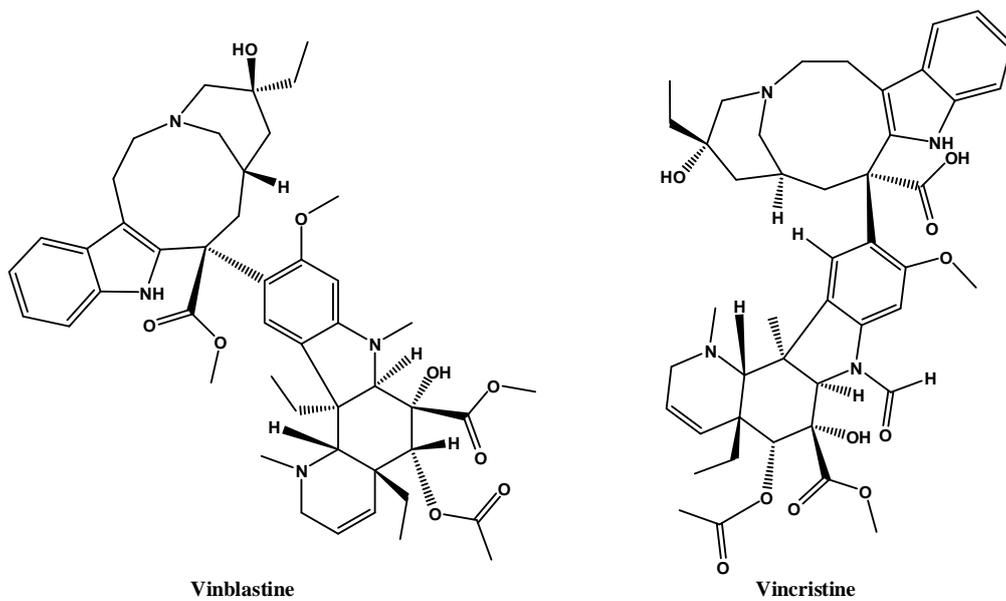**Vinblastine**

Vincristine

Figure 1: Alkaloid anti-cancer derivatives vinblastine and vincristine [6]
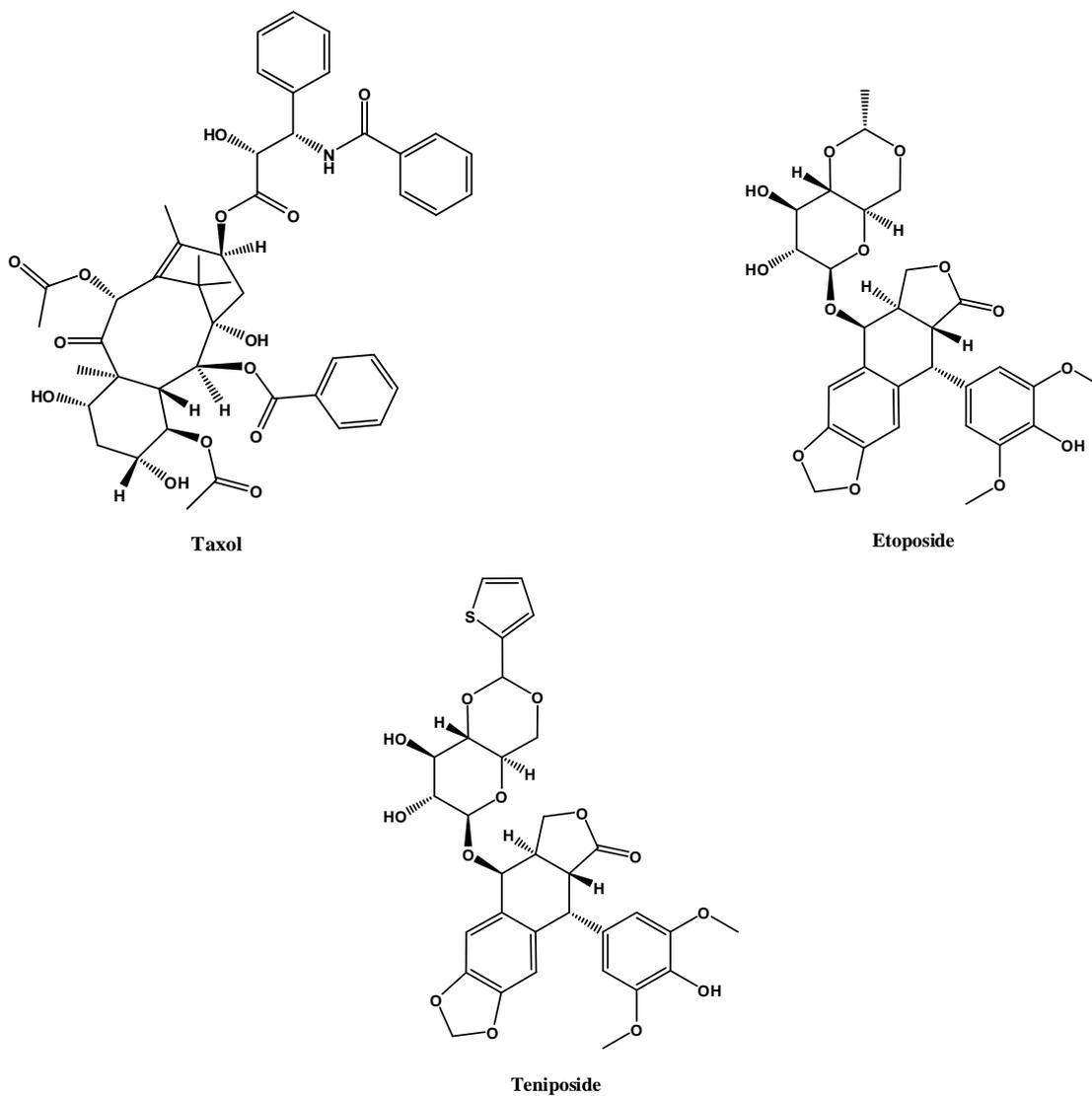

Taxol

Etoposide

Teniposide

Figure 2: Podophylltoxins (etoposide and teniposide) and taxane (taxol®) anti-cancer agents [6]
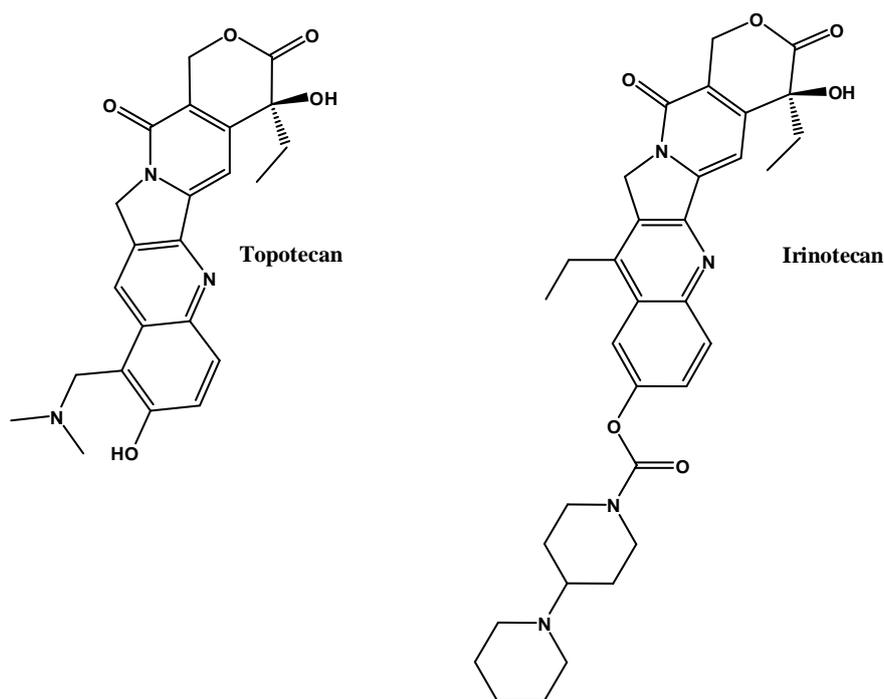
Figure 3: Camptothecins; topotecan and irinotecan cancer inhibitors [7]

Cancer and the Role of Glycosylation

Cancer has been deemed one of the world's major causes of mortality. Characterised by the out-of-growth" cell growth, prohibiting normal bodily functions depending on the type of cell infected. Cancer is a disease instigated by the function or dysfunction of catalytic pathways or cellular proteins. Being a heterogeneous disease there are a number of biological therapies available to patients, which target specific genetic markers or enzymes relevant to the development and growth of the condition. Through high through-put screening/dock it has been established that carbohydrates play an important role in cancer; and circulating or cell surface tumour-associated carbohydrate antigens serve as diagnostic markers [9-11]. Cell surface glycosylation is universal to all living cells reflecting their physiological state, and are perfectly positioned to mediate adhesion and motility. A shift from the normal glycosylation pathway leads to altered glycan expression due to one or more of the following changes: (1) under- or overexpression of glycosyltransferases deregulated at the level of epigenetics [12, 13], transcription [14-16], post-transcription [17], and/or chaperone [18]; (2) altered glycosidase activity [19-21]; (3) altered expression of glycoconjugate acceptor together with availability and abundance of the sugar nucleotide donor [22]; (4) modified sugar nucleotide transporter activity [23]; and (5) malfunction of the Golgi structure, the warehouse of

47

glycosyltransferases [9, 24, 25]. Abundant literature has suggested aberrant glycosylation contributes to various aspects of cancer development and progression, including proliferation, invasion, angiogenesis, metastasis and immunity [9, 26, and 27]. Fig. **4**, depicts the *N*-glycosylation pathway on the surface of the Golgi apparatus. Overall the biosynthesis of *N*-glycan's have been deemed an expensive process in terms of the number of actively participating enzymes in the synthesis and trimming of *N*-glycan's. Structural diversity of *N*-glycan's in mature proteins in the cell surface are introduced by glycosyltransferases in the Golgi complex as a terminal step. The core glycan's in the endoplasmic reticulum are universal from yeast to mammal and are considered intermediates [28].
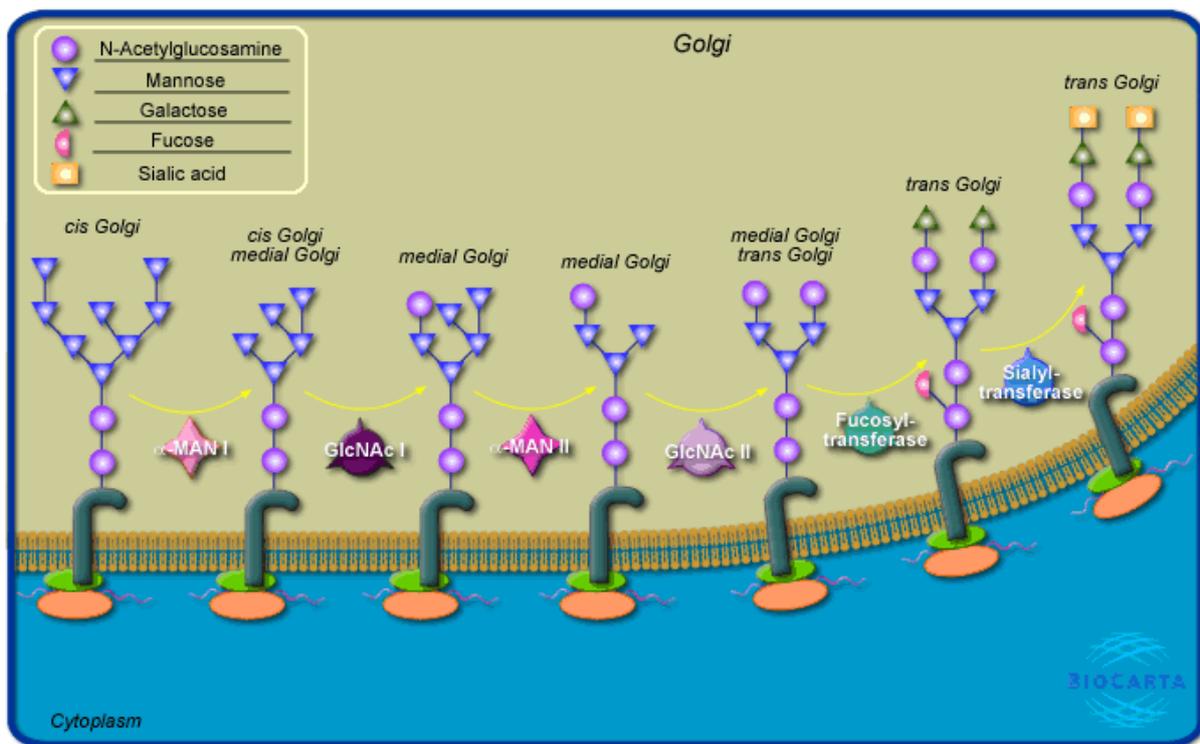

Figure 4: Diagram of N-glycosylation pathway in human Golgi complex [29]

Glycosidases are a vast class of enzymes designated to cleave glycosidic bonds of carbohydrates or polysaccharides to essentially assist in biochemical processes such as protein folding in living cells [27, 29]. Vocadlo and Davies provided insight into the mechanistic behaviour of glycosidase enzymes by understanding the enzyme reaction co-ordinates, through collective works in computational and structural studies [27]. The fundamental function of these enzymes is to hydrolyse glucose residues from glucosides. Their specificity originates at a structural level, in terms of their *α*- and/or *β*- configurations as well as the configuration of its particular substrate. The inability of a glycosidase to function correctly or expressions of the enzyme have been implicated as a cause of cancer. Related cancers include: breast cancer

[30], brain tumours [31], prostate [32], lung, gastric, bone and ovarian [33]. The glycosylation pathway is an imperative target in the prevention of development of cancer or further spread of the disease [34]. There are an estimated 16 enzymes steam rolling the glycosylation pathway within the cellular framework. These sites include the endoplasmic reticulum, golgi apparatus, cytosol and nucleus. Each of the glycosidase enzymes, based on tissue loci and organism, exists in either an α- and/or β-configuration and are governed by the structural feature of the bond on which they exert themselves [35]. A prime example of the uniqueness of this class of enzyme is glucosidase. The α-glucosidase depicted in Fig. **5**, is derived from sugar beet and classified as an endoplasmic reticulum inner membrane bound protein, which is responsible for the cleavage of α-(1,2)- linked glucose units from amylose to extended carbohydrate chains [36].
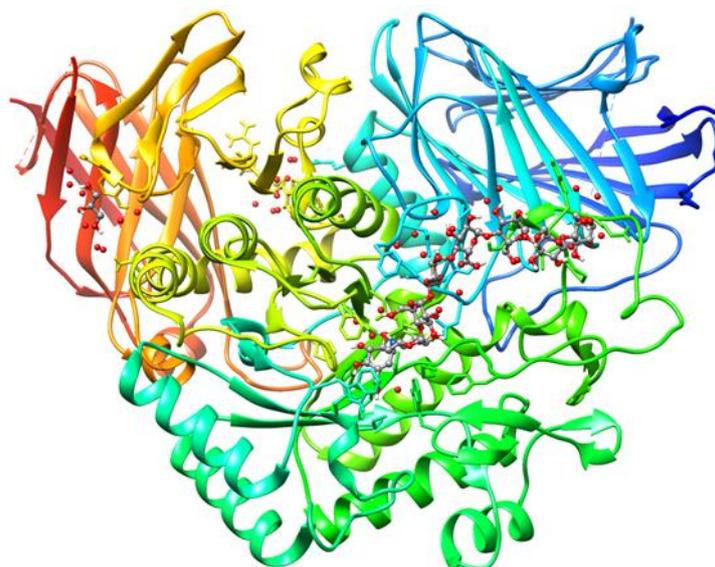
Figure 5: Crystal structure of sugar beet α-glucosidase (PDB code: 3WEO) [37]

Fig. **6**, highlights the covalent intermediate of human cytosolic β-glucosidase. This enzyme has been associated with essential organs of the human metabolism such as the small intestine, kidneys and liver. It is thought to play a role in the detoxification of xenobiotics (foreign chemical compounds which pose a toxic threat to the healthy cells), by hydrolysing β-glucoside moiety providing a site for conjugation which would lead to rapid excretion in the bile and urine [38, 39].

Figure 6: Crystal structure of the covalent intermediate of human cytosolic β-glucosidase (PDB code: 2ZOX) [40]

A vast array of natural product glycosidase inhibitors have been explored due to their extensive ability in the treatment of cancer. These include but are not limited to: castanospermine and deoxynojirimycin (glucosidase I & II and lysosomal α-glucosidase inhib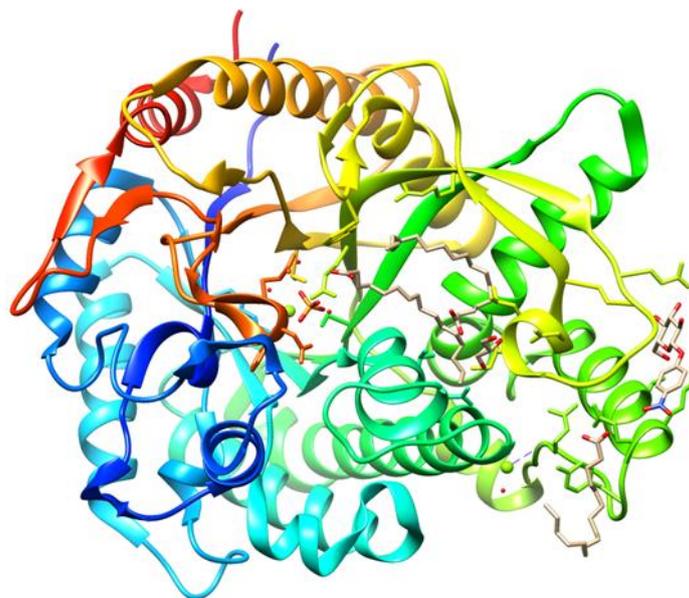itors), swainsonine (lysosomal and aryl-mannosidases inhibitor, and a potent mannosidase II inhibitor), kifunensine (potent mannosidase I inhibitor), allosamidin, salacinol, nojirimycin, mannostatin A & B, isogaomine, kojibiose, sophorose, nigerose, trehalose and xanthone derivatives.

Evolution of Computational Techniques to Present Day

By the 1950's much improvement had been made in the way of instrumental techniques developed to analyse chemical reactions and molecular structures. However, the lack of understanding in terms of kinetic models, visualisation of dynamic interactions as well as how and what the conformations, transition states and reaction intermediates look like and existence in their natural systems, left many an unanswered question. It was not until 1957 when researchers Alder and Wainwright, reported the first successful molecular dynamic (MD) simulation. Even in its infancy the pair brought to life phase transition. Their investigation was based on a solid-fluid system evolution comprised of rigid spheres colliding instantaneously. With limited computing, a 500-particle system was designed, emulating collisions between particles with a duration of one hour on an IBM 704 computer [41]. This sparked a ripple effect where in 1960, Gibson and colleagues implemented the first continuous repulsive Born-Mayer interaction potential in MD simulation. This attempt may have been the

first recorded MD method in materials science. Their study was based on the radiation damage in a copper (Cu) target. This was performed by applying a constant force towards each atom on the boundary of the crystallite to account for the attractive part of the interatomic interaction [42]. In 1964 Aneesur Rahman described the attractive and repulsive forces in an 864 argon atom system using the Lennard-Jones potential [43]. The computational methods applied in his study, such as pair correlation function, velocity autocorrelation function and mean square displacement calculated for liquid Argon, are still relevant in current studies. Current studies have observed the evolution of MD simulation and the vast approaches and information that can be derived from such analysis. MD studies have offered a remarkable platform in the field of pharmaceutics; with regards to design and discovery of new and effective drug therapies for a multitude of diseases [44]. Nobel prize winner Richard Feynman described the motion of atoms being governed by probability functions, where chemical bonds are not formed mechanically but rather by the shifting of electron clouds which act as both waves and particles [45]. Thus the standard of the 'lock and key' theory has been set aside to accommodate the new age of binding models which accounts for conformational changes as well as the random motions of receptors and ligands. This seemingly minor difference allows molecular dynamic simulations to stand apart from other static two dimensional models, thus playing an important role in drug discovery. With the ability to analyse a three dimensional (3D) crystal structure of protein or target, we may gain perspective into the key features that enable such structures to function and possibly enhance or inhibit the expression of the protein species. Such techniques which exploit these attributes are molecular dynamics and quantum mechanics, homology modelling and virtual screening. Each of which offers different introspection into the specific aim of study [46].

COMPUTATIONAL APPROACHES ON NATURAL AGENTS AGAINST CANCER
Homology Modelling

Homology modelling is a technique applied in the prediction of protein structure based on the act that proteins of similar sequences bear similar structures. The technique has promoted the identification of protein function and mechanism, drug binding to specific sites and rationalisation of select amino acids in the discovery of pertinent biological function ascribable to mutagenesis. Due to computational methods being economically viable high throughput docking is an ideal method to select promissory compounds of appropriate chemical nature from extensive virtual chemical libraries, since they incur minor errors. Lead optimisation

is a monotonous process of modifying the chemical structure of a known hit by modifying its physico-chemical and pharmacological properties to improve bioavailability, minimize unwanted toxicity and obtain the admissible drug profile appropriate for animal model studies and clinical trials [47]. Protein-ligand interaction comprehension is crucial in homology-based lead optimization, since the character of the structure obtained can steer ligand optimization towards enhanced pharmacological profiles. Homology modelling has a significant contribution to structure-based drug discovery through prediction and generation of rational 3D models for drug targets. It generates plausible protein structures by presenting a 3D model from a protein template sequence subject to previously reported homologous protein structures. When building a homology model for proteins, the following typical procedure should be adhered to: 1) 3D protein structures with ~30% primary sequence similarity over the protein of unknown structure should be used [46, 48]; 2) the alignment of template and target protein sequences; 3) identification of variable and structurally conserved regions; 4) generation of structurally conserved residues of the unknown structure from template structure(s); 5) generation of loop conformations for the unknown structure; 6) generation of side chain conformations for the modelled protein; 7) refinement and evaluation of the generated unknown structure. The RMSD (root of mean square deviation distance between Cα atoms) of the homology model is compared to that of the experimental structure, which should be ~ 1-2 Å, to verify the accuracy of the homology model [49]. The accuracy of a homology model is defined by the template of choice, alignment accuracy and application of efficient refinement methods [47]. Homology models generated with over 50% sequence identity have satisfactory accuracy for drug discovery applications; whereas models of sequence identities of 25% to 50% can be applied in target druggability assessment and design mutagenesis. Homology models have proven invaluable as a rationalising tool of SAR data and the prediction of binding modes of experimental compounds [50, 51]. Heparanase is an endo-*β*-D-glucuronidase, which has been reported as a target for antimetastatic agents. Its unique function in degrading heparin sulphate glycosaminoglycans in mammalian tissue has highlighted the enzyme. Over expression would result in invasive normal and malignant cells such as; immune cells, lymphoma, melanoma, and carcinoma cells as well as human head and neck tumours. Ishida *et al.*, had proposed the design of selective inhibitors of heparanase by using a homology model of the enzyme. The homology modelled enzyme was based on the sequence alignment of human heparanase and 1,4-*β*-xylanase from *Penicillium simplicissimum*. Based on this model it was identified that the interaction between inhibitor and heparanase enzyme was stabilised by arylalkylation [52]. Glucosidases are last stage carbohydrate digestive enzymes. They are

responsible for hydrolysing the glycosidic bond of oligosachharides. Park *et al.*, wished to design new inhibitors derived from structure based virtual screening. In order to accomplish this they were required to obtain a high quality amino acid sequence of the enzyme. This was achieved by the alignment of sequence from baker's yeast α-glucosidase and oligo-1,6-glucosidase from *Bacillus cereus*. Such that these sequences shared an amino acid identity and similarity of 38.5% and 58.4% respectively. The homology model was sufficient to perform docking studies. It was observed that the target and the template displayed highly similar folding structures, whereby the catalytic residues were conserved in terms of position in the active site when compared to the x-ray crystal structure of oligo-1,6-glucosidase (Figs. **7** & **8**) [53].
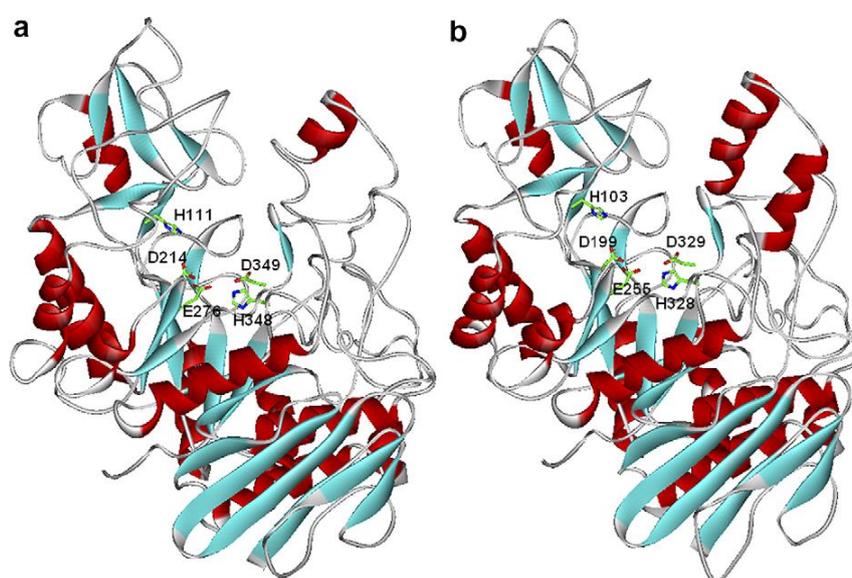

Figure 7: Comparative view of (a) homology modelled structure of α-glucosidase and (b) the x-ray crystal structure of oligo-1,6-glucosidase [53]
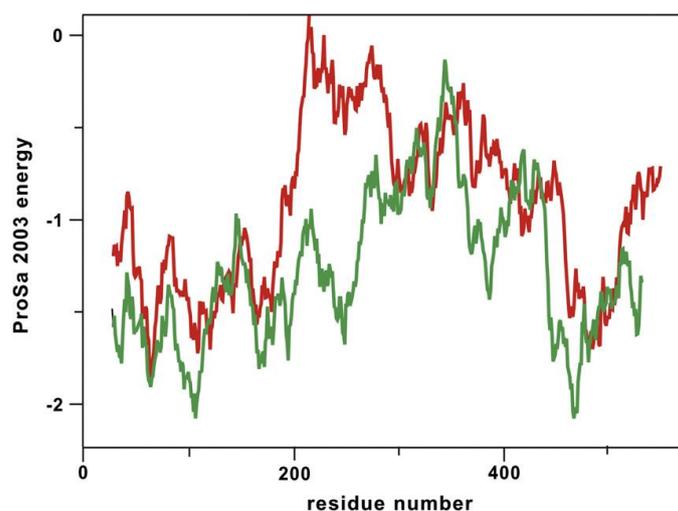

Figure 8: Comparison of the ProSa energy profiles of the homology modelled structure of α-glucosidase (red) and the x-ray crystal structure of oligo-1,6-glucosidase (green) [53]

Researchers Moorthy, Ramos and Fernandes conducted homology modelling of an *α*-glucosidase enzyme which was later used in a quantitative structure activity relationship study of xanthone derivatives. They had deduced that the active site of the model retained residues aspartic acid, histidine and glutamic acid, which contributed to the specific structural features of potential inhibitors [54]. Human glycan's which include *N*-linked glycan's, ABO blood group and Lewis antigens have prominent structural moieties of *α*-L-fucose a residue by-product of reaction catalysed by *α*-L-fucosidase [55, 56]. Many cancers are associated with the high levels of fucosylation which increases with expression of the enzyme [57, 58]. Bueren *et al.*, analysed the reaction co-ordinate of *α*-L-fucosidases by combining structural and quantum mechanical approaches. They were able to interpret inhibitor binding by deriving a homology model of enzyme from *B. thetaiotaomicron* as the sequence identity was in high correlation with that of the human enzyme [59].

Qualitative Structure Activity Relationship (QSAR)

QSAR is a technique that has infiltrated the scene of drug design and discovery as an invaluable tool. It has been implemented to identify ligands with high affinities for specific macromolecular targets, by screening drugs and providing potential outcomes of synthesis and testing. It has also been applied in extended research in predicting adsorption, distribution, metabolism, elimination, toxicity properties as well as bioavailability of compounds. Such studies can be undertaken on different dimensions ranging from 1D-QSAR, where only a single property of the ligand is exploited correlating to its activity; to 6D-QSAR which involves the ligand being represented as an ensemble of configurations with the explicit representation of different induced fit models and the representation of different solvation scenarios [60]. Glycosidases have evolved finely tuned active site configurations for the specific hydrolysis of glycosidic bonds. The inhibition of glucosidases has offered a noteworthy effect on the glycon structure ultimately affecting the maturation, transport, secretion and function of glycoproteins. Thus this could potentially have an altered recognition of cell-cell processes. Moorthy *et al.*, attempted QSAR studies on associated xanthone derivatives based on the data set published by Yan *et al.*, as inhibitors of *α*-glucosidase. Xanthones are natural derivatives extracted from different medicinal plant material (Fig. **9**) [61, 62]. Due to the limited information presented from current literature, they decided to delve into the link of biological activity data and physiochemical descriptors of molecules. By exploring structural features which have a determining effect on the binding affinity, mechanism of inhibition, topology and hydrophobicity. It was established that heteroatoms such as; oxygen bonded to carbon, are

favourable features for enzyme inhibition. Using an E-state count descriptor, the most favourable framework for inhibition of enzyme activity is a carbon atom linked with three aromatic bonds and hydrogen or other atoms [54, 63, and 64]. Moorthy and colleagues performed additional studies on the *α*-glucosidase enzyme of *B. Stearothermophilus* and *S. cerevisiae* and their interaction with chlorogenic acid inhibitors [65].
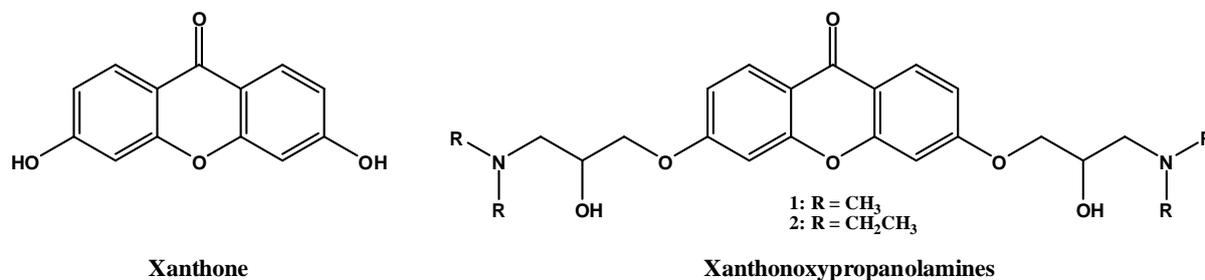

Figure 9: Xanthone and subsequent xanthone derivatives

Virtual Screening

Virtual screening has offered a cost effective method by which we may search for lead structures for further development as therapeutic agents in different biological systems and against a multitude of targets. An advantage of this technique is the resulting computational prediction of binding affinity, which contributes to bottlenecking and reduces the number of compounds required for testing. In terms of drug design, a range of computational methods are employed at distinctive stages of the procedure. As previously mentioned, high through-put screening of large compound libraries focuses on decreasing the number of potential ligands. Post lead optimization, reduction of experimental costs and time is emphasized. Molecular dynamic simulations have improved docking procedures in terms of Molecular Mechanics Poison-Boltzmann/ Generalised Born Surface Area (MM/PBSA or MM/GBSA), Free Energy Perturbation (FEP), Thermodynamic Integration (TI) and Linear Interaction Energy (LIE) approaches [66-69]. They add invaluable insight into the dynamic behaviour of proteins at various time-frames, from speedy internal motions to steady conformational changes or even protein folding processes. Studying of the explicit solvent molecule effect [70] on protein structure and stability to characterise a biomolecular system is plausible. Such system characteristics include density, conductivity, dipolar moment and thermodynamic parameters inclusive of interaction energies as well as entropies [71, 72]. Several MD-based in silico methods for binding energy predictions have been extensively applied in drug design as they provide statistically meaningful conformational ensembles for thermodynamic calculations at a reasonable computational cost. MM/PBSA and MM/GBSA are considered

more computationally efficient calculations as opposed to FEP/TI. This is attributed to the former free binding energy calculations having fewer constraint rules to obey, as well as their capability to dissect total binding free energy into different interaction terms [73]. MM-PBSA/GBSA are, however, theoretically more rigorous as such calculations take into consideration the conformation of free ligand, receptor and ligand-receptor complex. The FEP theory was initially introduced by Zwanzig in 1954, where he correlated the free energy distinction between reference and target state of a system to its average function of energy estimated by sampling initial states [74]. Since then much improvement has been made by way of calculating free energy differences using this technique. These include coupling the mathematical algorithm with advanced molecular dynamics and Monte Carlo sampling to elucidate respective solvation free energies, pKa values, medium-effects on conformational equilibria, host-guest binding affinities, organic and biochemical reactions free energy surfaces [75]. Alternatively, binding energies can be calculated by scoring functions as they compute results faster than the MM-GB/PBSA models. However, scoring functions lack precision, with an average error of ~2.5 kcal/mol [69]. In 2008 Park and colleagues suggested discovering novel inhibitors of *α*-glucosidase through virtual screening encouraged by a homology-modelled protein structure [53]. The technique allowed the screening of a library estimating 85,000 compounds, resulting in the reduction of sample size to 200 of the highest scoring compounds. It was later established that only 13 of these suggested agents presented a 50% inhibition at a concentration range between 0 μM and 50 μM [53]. Figs. **10**, **11**, are representative inhibitors, their scaffolds now form the templates for further development and investigation of structure based *de novo* synthetic inhibitors.
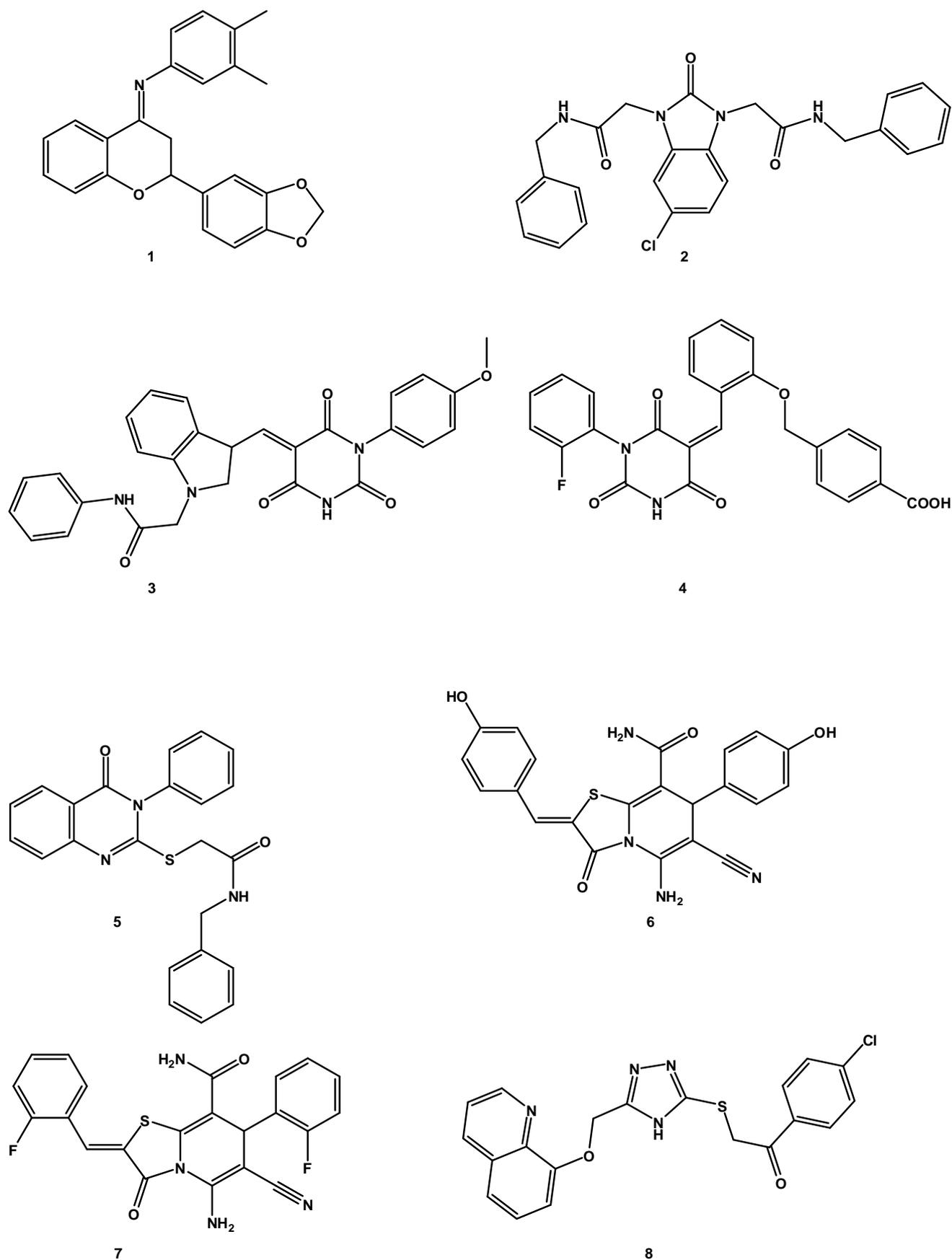
Figure 10: Compounds 1 - 8 of newly identified α-glucosidase inhibitors [53]
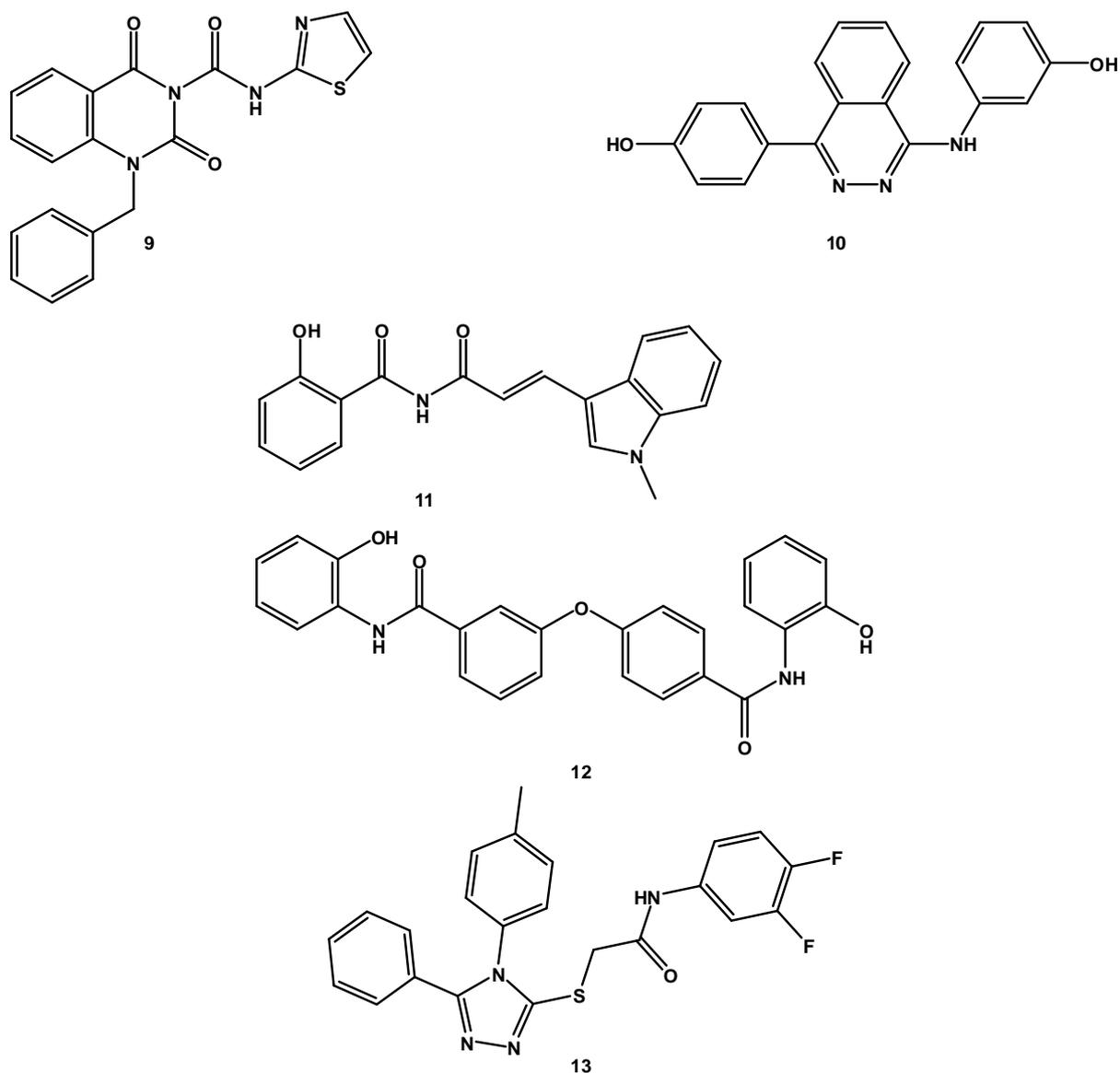
Figure 11: Chemical structures 9 - 13 of newly identified α-glucosidase inhibitors [53]

Molecular Modelling and Binding Modes

Molecular modelling is a fusion of structure and function of molecules in the form of proteins, ligands and cellular entities. The manipulation of chemistry and laws of protein folding allows scientists to seek understanding of physiology of disease. The essence of the technology utilises molecular biology, x-ray crystallography and quantum mechanics [71, 76, and 77]. The allure of molecular modelling techniques lies in the atomistic level description of molecular systems. There are three common stages involved in molecular modelling studies. The first stage is a selection process whereby a model is chosen to best describe the intra- and inter-molecular interactions in a system. The two most frequently used models are quantum mechanics and molecular mechanics, both follow the fundamentals of the energy associated

with the arrangement of atoms and molecules within the system. The second stage encompasses the actual mathematical calculation of the system, this includes energy minimisation, molecular dynamic or Monte Carlo simulations or a conformational search. This leads directly into the final stage of post-analysis of the calculations carried out in stage two. It is no small wonder that from these stages computational chemistry or molecular modelling is regarded as a classical example of scientific art [78]. From the post-analysis of calculations performed a multitude of information may be generated with respect to molecular geometry, energies, electronic properties, spectroscopic properties and bulk properties [78, 79]. Allosamidin has been classified as a pseudo-trisaccharide (Fig. **12**), isolated from the mycelia of *Streptomyces* sp. [80].
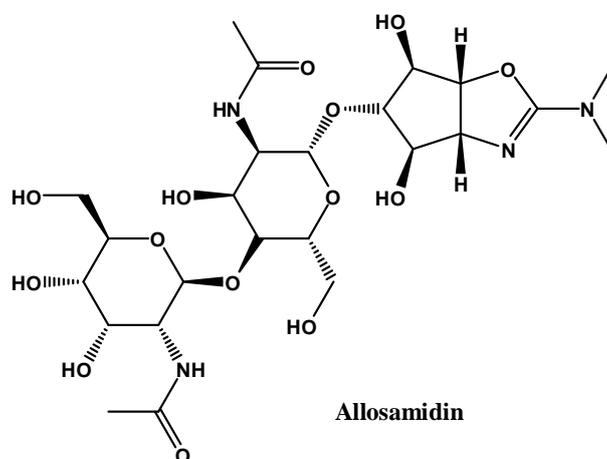
Figure 12: 2D structure of allosamidin

Germer and colleagues performed molecular modelling studies on the allosamidin and six additional analogues based on the allosamidin framework in an effort to validate conformational information extracted from 2D NMR studies. Of which it was concluded that the results of each method were aligned. Proving once again that computational approaches to scientific investigations can act both as a tool of diagnosis and validation [81]. Kara *et al.*, proposed a mechanism of action of allosamidin in chitinase, which is regarded as a type of glycohydrolase enzyme. With modelling, Kara and colleagues suggested that mechanism of action of the enzyme involves the bending of acetalamido group to oxygen of the substrate ring which subsequently neutralises the charge. Allosamidin mimics the transition state of the catalytic reaction and thus perpetuates competitive and selective inhibition of the enzyme [82]. *α*-Mannosidase is a class II enzyme partaking in the *N*-glycosylation reaction, with a function to link oligosaccharides to distinct asparagine amino acids in incipient proteins. Thus subsequent inhibitors show radical antitumour and antimetastatic activity. Such inhibitors include kifunensine an alkaloid derived from actinomycete *Kitasatosporia kifunense,* salacinol

extracted from *Salacia reticulata* and swainsonine [83-85]. Shah *et al.*, compared the binding of kifunensine and 1-deoxymannojirimycin to α-mannosidase (Fig. **13**).

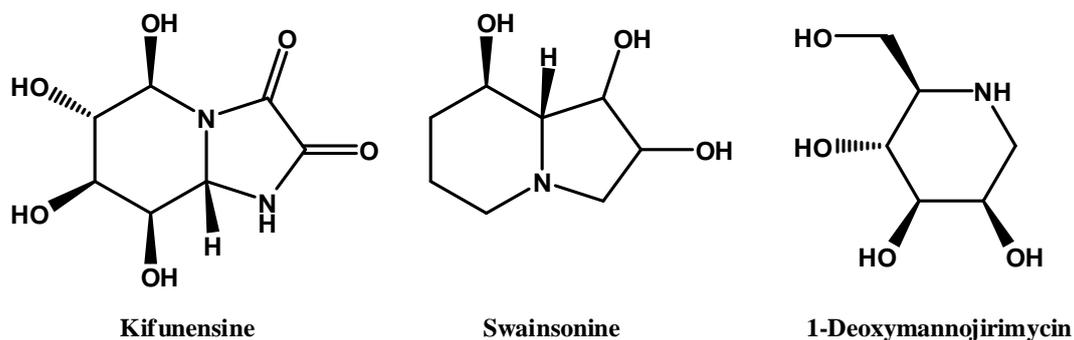

Figure 13: 2D structure of α-mannosidase inhibitors kifunensine, swainsonine and 1-deoxymannojirimycin [88]

1-Deoxymannojirimycin, once thought only a derivative of the natural inhibitor nojirimycin a piperidine alkaloid isolated from *Streptomyces* sp., was found naturally occurring from the extract of mulberry leaves, *Bacillus* and *Streptomyces* sp. as well as Micronesian marine sponge [5, 86, 87]. However, both entities proved viable inhibitors of glycosidases and thus exerted an effective activity as antitumor and anticancer agents. Shah and colleagues performed energy directed studies which proposed kifunensine inferring moderate inhibition to class II mannosidase as compared to being more potent against class I mannosidase. With kifunensine maintaining a 1*C*4 conformation in each of the different active sites. 1-Deoxymannojirimycin being a smaller inhibitor observed a 4*C*1 conformation in the class I mannosidase active site but a 1*C*4 conformation in the class II mannosidase active site [88]. Wen *et al.*, investigated the binding mode off the previously mentioned inhibitors docked in the glycosidase enzyme. The study revealed that the five membered ring of kifunensine is planar. There exists three distinct binding modes differing by the distance between zinc and hydroxyl polar groups. In mode I the distance was estimated to 2.3 Å, mode II 3.31 Å, whereas with mode III the ligand appeared to turn completely within the active site and thus the distance was of polar group to zinc was too far. With respect to salacinol the inflexibility of its five membered ring offered pseudo rotation (Fig. **14**). The three distinct binding modes of this system reasoned that the relative conformations are independent of the interaction between the polar species of the ligand and active site interactive moieties. In conclusion these findings satisfied those observed from x-ray crystallography [89].
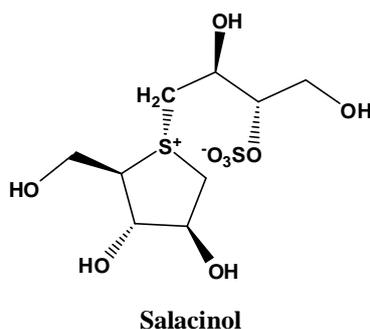
Salacinol

Figure 14: 2D structure of salacinol

Mannostatin A and B (Fig. **15**), were discovered by Aoyagi and collaborators in 1989 from isolates of *Streptoverticilliu verticillus* variation *quantum*. Classed as novel aminocylcopentitol structures they were found to be active inhibitors of α-mannosidase and α-glucosidase [90]. Kawatkar *et al.*, explored the ligand-protein interactions of mannostatin A and α-mannosidase II. Their research team had elucidated that mannostatin A mimicked covalently linked mannosyl intermediate which adopts a 1S5 skew boat conformation. It was established that the thiomethyl group is required for high affinity binding, which was reported to have good overlay with the C-6 hydroxyl of the covalently linked intermediate [91]. Kuntz *et al.*, had performed binding mode studies on a similar system. Their findings were *in cognito* with corresponding literature, that inhibition sees crucial contribution of zinc interactions, interactions with Asp341 and Asp472, as well as hydrophobic interactions with Phe206 of the enzyme [92].

Mannostatin A, K_i = 36nM   **Mannostatin B, K_i = 150nM**
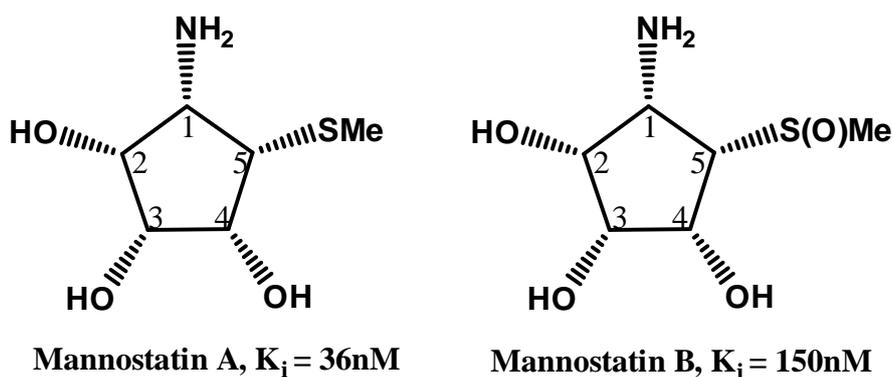
Figure 15: 2D structure of mannostatin A and B with corresponding inhibition values [92]

Cardona *et al.*, conducted molecular dynamic simulations on glucoamylase II from *Aspergillus awamori* with natural inhibitors 1-deoxynojirimycin and lentiginosine (Fig. **16**) in an attempt to understand binding for further investigation in the design of new inhibitors in human systems as antitumour agents. It was then determined that lentiginosine observed optimal conformation

in the enzyme cavity when hydrogen bonding to residues Arg54 and Arg55. As well as inhibition was dominated by the interaction of hydroxyl groups of the substrate with key enzyme residues [93]. Zhou *et al.*, performed similar simulations, however, the system involved comprised of 1-deoxynojirimycin and isogaomine. They showed that 1-deoxynojirimycin bound to glucosidase in a protonated chair conformation with a binding energy value of -46.76 kJ/mol.

Major thermodynamic contributions implying favourable enthalpy of binding were attributed to strong hydrogen bonds and electrostatic interactions between enzyme and ligand [94].
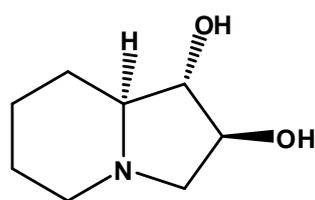
Lentiginosine

Figure 16: 2D structure if lentiginosine

Pereira *et al.,* performed a conformational and dynamical study on disaccharides in water using explicit-solvent molecular dynamic simulations [95, 96]. They analysed eight reducing disaccharides of *β*-anomeric configuration isolated from honey: kojibiose, sophorose and nigerose (Fig **17**), as well as laminarabiose, maltose, cellobiose, isomaltose and gentiobiose. In current literature most of the disaccharides have been characterised and theoretically investigated using MD simulations, molecular mechanics calculations or quantum mechanics calculations. In the study a continuous 50 ns run time was performed, during which time it was concluded that the preference of dihedral angles which dictates that polar substituents are to be orientated away from the ring [97].
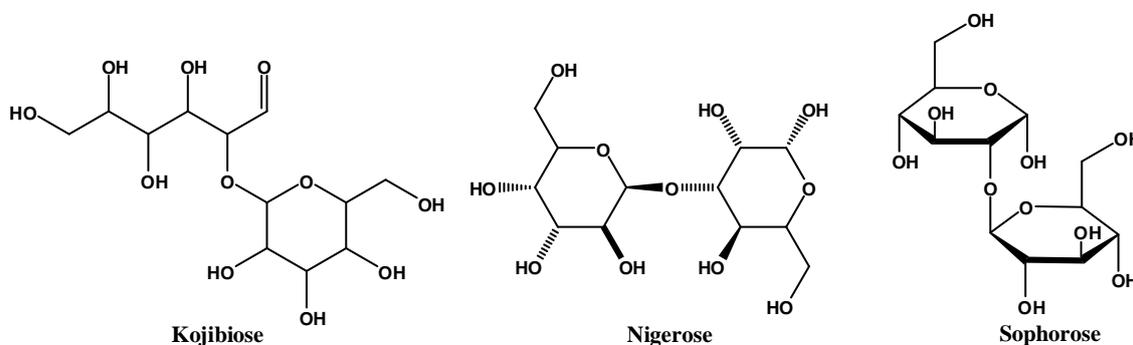

Figure 17: 2D structure of kojibiose, nigerose and sophorose disaccharides

Kräutler *et al*., explored conformation, dynamic, solvation and respective stabilities of specific *β*-hexopyranoses in water using a molecular dynamic platform. Four *β*-D-aldohexopyranoses monosaccharides were studied, namely: *β*-D-glucose, *β*-D-mannose, *β*-D-galactose and *β*-D-talose; which were simulated in a series of 200 ns timescale. During the simulation the latter three substrates maintained a $^4C_1$ chair conformation, unlike glucose which has evolving boat and twisted configurations. Kräutler *et al.*, described the intramolecular hydrogen-bonding pattern with each substrate, and were deemed opportunistic contributors to the relative conformation and stability of the structure. They had ranked the estimated epimerisation energies in order of decreasing stability, with talose exhibiting the least stability superseded by galactose and mannose in turn, with glucose registered as most stable. This confirmed results obtained from intramolecular effects and hydrophilicity investigations [98]. Trehalose (Fig. **18**), a sugar extract from yeast, exists as a naturally occurring disaccharide in mushrooms and other fungi. This unique combination of *α*-D-glucopyranosyl-*α*-D-glucopyranoside has been implicated as prominent inhibitor of the glucosidase enzyme [99]. The dysfunction of the enzyme has proven a target in the prevention of growth and development of metastatic cancer. Trehalose has been shown to form direct hydrogen bonds to proteins which endure the ligand requiring occupation of specific active site in a specific orientation [100]. There have been a number of reported simulations investigating the structural dynamics of binary sugar/water solution in an attempt to elucidate physical properties [101-111]. Sum *et al*. [112] and Pereira *et al*. [113, 114] had conducted a simulation of a lipid membrane in the presence of sugars. They discovered that specific hydrogen bonding governed the interaction between sugar and lipid molecules. Lerbret *et al*. had decided to simulate a lysozyme in aqueous solution of trehalose, sucrose and maltose. It was suggested that trehalose was the most hydrated of all the three sugars as it maintained a greater number of hydrogen bonds with water [115].
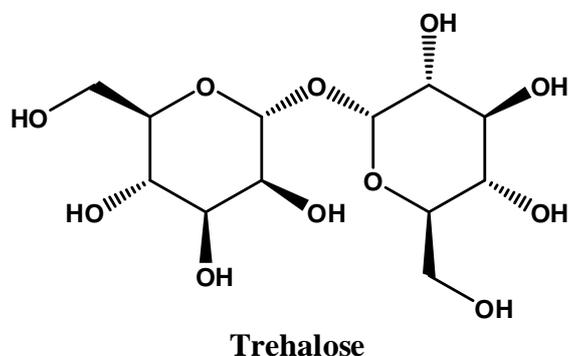

Trehalose

Figure 18: 2D structure of trehalose

Park and collaborators, took their study of novel α-glucosidase inhibitors a step further by exploiting the binding modes of the 13, then, newly identified molecules in the active site of the enzyme during which they noticed that the phenolic oxygen of the select inhibitor was a key feature in the migration of a hydrogen bond from His348 to Asp349, essentially translating to the phenol moiety playing a pivotal role for binding to the active site. It was also established that the stabilisation of docked ligands can be attained by inducing hydrophobic interactions at across from the active site [53]. Pistarà *et al.*, had provided *in silico* characterisation of cyclitols, inosose the ketone structure of inositols found in mammalian tissues and *Streptomyces griseus*. The inosose molecules were docked in human maltase-glucoamylase which has been crystallised with miglitol (Fig. **19**). It was determined that nitrogen that engages in a hydrogen bond with Asp443 in the catalytic residue is an essential entity for inhibitory activity. Additional structural components such as benzyl have the ability of filling the active site [116]. The binding free energy of compound 1 was calculated to be -6.0 kcal/mol and was found to possess the greatest inhibitory effect with an inhibition constant of 36.9 µM.

Miglitol   **Compound 1**
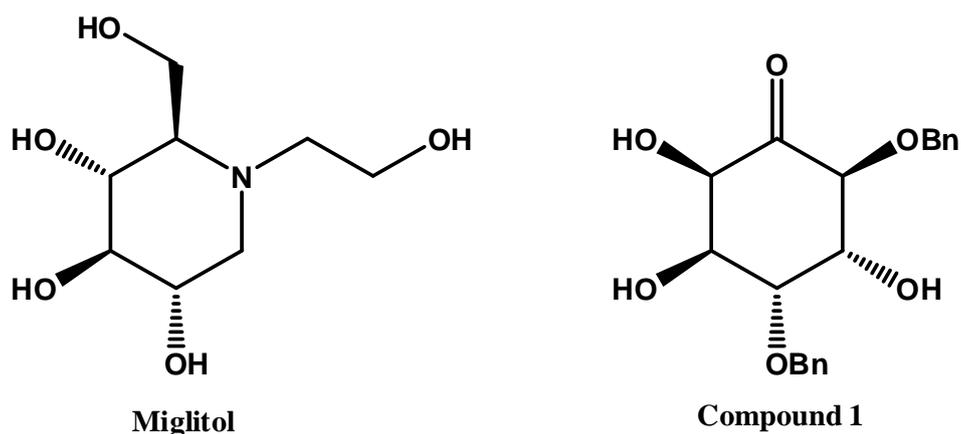
Figure 19: (left) 2D structure of miglitol and (right) 2D structure of compound 1

CONCLUSION

From the topics described above, it is evident that there exist considerable gaps in the way of computational studies performed on the natural element inhibitors of glycosidase enzymes. These include the understanding of kinetic mechanisms of inhibitor enzyme interaction, conformational support within active site, effects of mutagenesis, selective and competitive inhibition of enzyme species; and many more. Thus, there is great opportunity for continued research toward answering many of these essential components by way of a computational approach. Additional research toward understanding the reaction kinetics and mechanism of glycosidase enzymes in their natural biological system, would be imperative in our quest to discover and design innovative anticancer inhibitors. Despite the large array of high resolution

crystal structures available within the protein databank, much effort is required in acquiring crystal structures of human glycosidase. Such that accurate study may be performed to investigate possible inhibitors. Intensive molecular dynamic study encompassing molecular biology techniques would aid research in determining critical enzyme reactions that perpetuate cancerous species. Thus by isolating specific drug targets, it would allow us to build into the inhibitor, selective drug delivery systems exerting a specialised response. An attractive feature of obtaining refined crystal structures of human glycosidase enzymes, offers insight into the active site residues. Thus accurate binding modes may be established as well as chemical structuring of potential inhibitors may be designed according to supported conformational and steric properties. Despite limitations of the crystal library with the aid of computational tools we may continue studies in this field by accurately constructing best-fit homology models of associated enzymes. Extensive studies based on QSAR and QM/MM of biological systems will dramatically improve our understanding of the influence of glycosidase enzymes in cancer development. As well as it could bridge the gap between theoretical and experimental studies, by economising on time, cost of study and prioritising higher possibility enzyme structure and mechanistic predictions. The above mentioned attributes contribute to the design and development of new drugs as anticancer agents as well as provides pivotal information with regards to the specific enzyme pathways implicated in cancer. Molecular dynamic studies present an opportunistic methodology whereby we may unlock and unveil the mystery underlying the disease thus unleashing great strides in our fight against cancer. Computational techniques boast an arsenal to unravel and offer a broad introspection into the molecular dynamics of glycosidase enzymes. Thus the study of this family is not exclusive or restricted to its effect on cancer but rather a multi-dimensional study on other congenital disorders of glycosylation may be investigated. Armed with computational tools the study of glycosidase enzymes remains limitless, and with the limited literature available within this topic much research is yet to be done in this field.

CONFLICT OF INTEREST

The authors confirm that this article content has no conflict of interest.

REFERENCES

[1] Wong, K.C., Wang, C.; Wu, L. History of Chinese medicine: Being a chronicle of medical happenings in china from ancient times to the present period 1936: Tientsin Press.

[2] Dias, D.A.; Urban, S.; Roessner, U. A historical overview of natural products in drug discovery. *Metabolites,* **2012**, *2*(2), 303-336.

[3] Balunas, M.J. Kinghorn, A.D. Drug discovery from medicinal plants. *Life Sci.,* 2005. **78**(5), 431-441.

[4] Li, J.W.; Vederas, J.C. Drug discovery and natural products: End of an era or an endless frontier? *Science,* **2009**, *325*(5937), 161-165.

[5] Cragg, G.M.; Newman, D.J. Plants as a source of anti-cancer agents. *J. Ethnopharmacol.,* **2005**, *100*(1-2), 72-79.

[6] Srivastava, V. Plant-based anticancer molecules: A chemical and biological profile of some important leads. *Bioorg. Med. Chem.,* **2005**, *13*(21), 5892-5908.

[7] Newman, D.J.; Giddings, L.A. Natural products as leads to antitumor drugs. *Phytochem. Rev.,* **2013**, *13*(1), 123-137.

[8] Gordaliza, M. Natural products as leads to anticancer drugs. *Clin.Trans. Oncol.,* **2008**, *9*(12), 767-776.

[9] Padler-Karavani, V. Aiming at the sweet side of cancer: Aberrant glycosylation as possible target for personalized-medicine. *Cancer Lett.*, **2013.**

[10] Dube, D.H.; Bertozzi, C.R. Glycans in cancer and inflammation [mdash] potential for therapeutics and diagnostics. *Nat. Rev. Drug Discov.,* **2005**, *4*(6), 477-488.

[11] Ludwig, J.A.; Weinstein, J.N. Biomarkers in cancer staging, prognosis and treatment selection. *Nat. Rev. Cancer,* **2005**, *5*(11),845-856.

[12] Kannagi, R.;Yin, J.; Miyazaki, K.; Izawa, M. Current relevance of incomplete synthesis and neo-synthesis for cancer-associated alteration of carbohydrate determinants—Hakomori's concepts revisited. *Biochimica. Biophysica. Acta,* **2008**, *1780*(3), 525-531.

[13] Miyazaki, K.; Ohmori, K.; Izawa, M.; Koike, T.; Kumamoto, K.; Furukawa, K.; Ando, T.; Kiso, M.; Yamaji, T.; Hashimoto, Y.; Suzuki, A.; Yoshida, A.; Takeuchi, M.; Kannagi, R. Loss of disialyl lewisa, the ligand for lymphocyte inhibitory receptor sialic acid-binding immunoglobulin-like lectin-7 (siglec-7) associated with increased sialyl lewisa expression on human colon cancers. *Cancer Res.,* **2004**, *64*(13), 4498-4505.

[14] Buckhaults, P.; Chen, L.; Fregien, N.; Pierce, M.Transcriptional regulation of N-acetylglucosaminyltransferase by the srconcogene. *J. Biol. Chem.,* **1997**, *272*(31), 19575-19581.

[15] Le Marer, N.; Laudet, V.; Svensson, E.C.; Cazlaris, H.; Van Hille, B.; Lagrou, C.; Stehelin, D.; Montreuil, J.; Verbert, A.; Delannoy, P. The c-Ha-ras oncogene induces increased expression of β-galactoside α-2,6-sialyltransferase in rat fibroblast (FR3T3) cells. *Glycobiology,* **1992**, *2*(1), 49-56.

[16] Hatano, K.; Miyamoto, Y.; Mori, M.; Nimura, K.; Nakai, Y.; Nonomura, N.; Kaneda, Y. Androgen-regulated transcriptional control of sialyltransferases in prostate cancer cells. *PLoS One,* **2012**, *7*(2), e31234.

[17] Kim, J.K.; Noh, J.H.; Jung, K.H.; Eun, J.W.; Bae, H.J.; Kim, M.G.; Chang, Y.G.; Shen, Q.; Park, W.S.; Lee, J.Y.; Borlak, J.; Nam, S.W. Sirtuin7 Oncogenic Potential in Human Hepatocellular Carcinoma and its regulation by the tumour suppressors MiR-125a-5p nd MiR-125b. *Hepatology,* **2013**, *57*(3), 1055-1067.

[18] Schietinger, A.; Philip, M.; Yoshida, B.A.; Azadi, P.; Liu, H.; Meredith, S.C.; Schreiber, H. A mutant chaperone converts a wildtype protein into a tumor-specific antigen. *Science,* **2006**, *314*(5797), 304-308.

[19] Miyagi, T. Aberrant expression of sialidase and cancer progression. *Proc. JPN. Acad. Ser. B. Phys. Biol. Sci.,* **2008**, *84*(10), 407-418.

[20] Miyagi, T.; Wada, T.; Yamaguchi, K.; Hata, K. Sialidase and malignancy: A minireview. *Glycoconj. J.,* **2003**, *20*(3), 189-198.

[21] Kakugawa, Y.; Wada, T.; Yamaguchi, K.; Yamanami, H.; Ouchi, K.; Sato, I.; Miyagi, T. Up-regulation of plasma membrane associated ganglioside sialidase (Neu3) in human colon cancer and its involvement in apoptosis suppression. *Proc. Nat. Acad. Sci.,* **2002**, *99*(16), 10718-10723.

[22] Dennis, J.W.; Laferté, S.; Waghorne, C.; Breitman, M.L.; Kerbel, R.S. Beta 1-6 branching of Asn-linked oligosaccharides is directly associated with metastasis. *Science,* **1987**, *236(*4801), 582-585.

[23] Kumamoto, K.; Goto, Y.; Sekikawa, K.; Takenoshita, S.; Ishida, N.; Kawakita, M.; Kannagi, R. Increased expression of udp galactose transporter messenger rna in human colon cancer tissues and its implication in synthesis of thomsen-friedenreich antigen and sialyl lewis a/x determinants. *Cancer Res.,* **2001**, *61*(11), 4620-4627.

[24] Kellokumpu, S.; Sormunen, R.; Kellokumpu, I. Abnormal glycosylation and altered Golgi structure in colorectal cancer: Dependence on intra-Golgi pH. *FEBS Lett.*, **2002**, *516*(1-3), 217-224.

[25] Moremen, K.W.; Tiemeyer, M.; Nairn, A.V. Vertebrate protein glycosylation: Diversity, synthesis and function. *Nat. Rev. Mol. Cell Biol.,* **2012**, *13*(7), 448-462.

[26] Meany, D.L.; Chan, D.W. Aberrant glycosylation associated with enzymes as cancer biomarkers. *Clin. Proteomics,* **2011**, *8*(1), 7.

[27] Vocadlo, D.J.; Davies, G.J. Mechanistic insights into glycosidase chemistry. *Curr. Opin. Chem. Biol.,* **2008**, *12*(5), 539-555.

[28] Gramatikoff, K. Steps in the glycosylation of mammalian n-linked oligosaccharides. Pathways [cited 2014 02 September]; Available at: http://www.biocarta.com/pathfiles/h_aMANPathway.asp#contributors.

[29] Divakar, S.; Glycosidases, *Enzymatic Transforation,* **2013**, 5-21.

[30] Kushen Ramessur, P.G.A.M.D. Glycosidases and breast cancer cell migration through the extracellular matrix. *The Biomed. Scient.,* **2010**.

[31] Prabha, M.N.R.S.; Ravi, V. Specific activity of glycosidases in brain tumors and their expression in primary explants culture. *J. Biochem. Technol.,* **2013**, 5(1), 654-665.

[32] Bellezza, I.; Aisa, M.C.; Palazzo, R.; Costanzi, E.; Mearini, E.; Minelli, A. Extracellular marix degrading enzymes at the prostasome surface. *Prostate Cancer Prost. Dis,* **2005**, 8, 344-348.

[33] Niedbala, M.J.; Madiyalakan, R.; Matta, K.; Crickard, K.; Sharma, M.; Bernacki, R.J. Role of glycosidases in human ovarian carcinoma cell mediated degradation of subendothelial extracellular matrix. *Cancer Res.,* **1987**, *47*, 4634-4641.

[34] Yarema, K.J., Bertozzi, C.R. Characterizing glycosylation pathways. *Genome Biol,* **2001**, 2(5).

[35] Gerber-Lemaire, S.; Juillerat-Jeanneret, L. Glycosylation pathways as drug targets for cancer: Glycosidase Inhibitors. *Mini-Rev. Med. Chem.,* **2006**, *6*(9), 1043-1052.

[36] Yamasaki, Y.; Konno, H. Purification and properties of alphaglucosidase from suspension-cultures sugar-beet cells. *Phytochemistry,* **1991**, *30*(9). 2861-2863.

[37] Tagami, T.; Yamashita, K.; Okuyama, M.; Mori, H.; Yao, M.; Kimura, A. Sugar be*et alpha*-glucosidase with acarviosylmaltohexaose. [cited 2014 02 September]; Available at: http://www.rcsb.org/pdb/explore/explore.do?structureId=3weo.

[38] Day, A.J.; DuPont, M.S.; Ridley, S.; Rhodes, M.; Rhodes, M.J.; Morgan, M.R.; Williamson, G. Deglycosylation of flavonoid and isoflavonoid glycosides by human small intestine and liver β-glucosidase activity. *FEBS Lett.,* 1998. *436*(1), 71-75.

[39] Pelkonen, O.; Jorma, T.A.; Jukka, H.; Miia, T. Consideration of metabolism in *in vitro* cellular systems, in In Vitro toxicology Systems, *Meth. Pharmacol. Toxicol.,* **2014**, 501-519.

[40] Noguchi, J.; Hayashi, Y.; Baba, Y.; Okino, N.; Kimura, M.; Ito, M.; Kakuta, Y. Crystal structure of the covalent intermediate of human cytosolic beta-glucosidase. *Biochem. Biophys. Res. Commun.,* **2008**, *374*(3), 549-552.

[41] Alder, B.J.; Wainwright, T.E. Studies in molecular dynamics. I. general method. *J. Chem. Phys.,* **1959**, *31*(2), 459-466.

[42] Gibson, J.B.; Dynamics of radiation damage. *Phys. Rev.,* **1960**, *120*(4), 1229-1253.

[43] Rahman, A. Normalization of diffraction data from liquids. *J. Chem. Phys*., **1965**, *42*(10), 3540-3542.

[44] Karplus, M.; McCammon, J.A. Molecular dynamics simulations of biomolecules. *Nat. Struct. Biol.,* **2002**, *9*(9), 646-652.

[45] Durrant, J.; McCammon, J.A. Molecular dynamics simulations and drug discovery. *BMC Biol.,* **2011**, *9*(1), 71.

[46] Honarparvar, B.; Govender, T.; Maguire, G.E.; Soliman, M.E.; Kruger, H.G. Integrated approach to structure-based enzymatic drug design: Molecular modeling, spectroscopy, and experimental bioactivity. *Chem. Rev.,* **2013**, *114*(1), 493-537.

[47] Cavasotto, C.N.; Phatak, S.S. Homology modeling in drug discovery: Current trends and applications. *Drug Disc. Today,* **2009**, *14*(13), 676-683.

[48] Power, T.D.; Ivanciuc, O.; Schein, C.H.; Braun, W.Assessment of 3D models for allergen research. *Proteins,* **2013**, *81*(4), 545-554.

[49] Rodrigues, J.P.; Melquiond, A.S.; Karaca, E.; Trellet, M.; Van Dijk M.; Van Zundert G.C.; Schmitz, C.; De Vries S.J.; Bordogna, A.; Bonati, L.; Kastritis, P.L.; Bonvin, A.M. Defining the limits of homology modeling in information-driven protein docking. *Proteins*, **2013**, *81*(12), 2119-2128.

[50] Schmidt, T.; Bergner, A.; Schwede, T. Modelling threedimensional protein structures for applications in drug design. *Drug Discov. Today*, **2014**, *19*(7), 890-897.

[51] Nakamura, S. Homology modeling of human alpha-glucosidase catalytic domains and sar study of salacinol derivatives. *Open J. Med. Chem.*; **2012**, *2*(03), 50-60.

[52] Ishida, K.; Hirai, G.; Murakami, K.; Teruya, T.; Simizu, S.; Sodeoka, M.; Osada, H. Structure-based design of a selective heparanase inhibitor as an antimetastatic agent. *Mol. Cancer Therapeut.,* **2004**, 3(9).

[53] Park, H.; Hwang, K.Y.; Oh, K.H.; Kim, Y.H.; Lee, J.Y.; Kim, K. Discovery of novel α-glucosidase inhibitors based on the virtual screening with the homology-modeled protein structure. *Bioorg. Med. Chem.,* **2008**, *16*(1) 284-292.

[54] Moorthy, N.S.; Ramos, M.J.; Fernandes, P.A. Topological, hydrophobicity, and other descriptors on alpha-glucosidase inhibition: A QSAR study on xanthone derivatives. *J. Enzyme Inhib. Med. Chem.,* **2011**, 26(6), 755-766.

[55] Liu, T.W.; Ho, C.W.; Huang, H.H.; Chang, S.M.; Popat, S.D.; Wang, Y.T.; Wu, M.S, Chen, Y.J.; Lin, C.H. Role for alpha-Lfucosidase in the control of Helicobacter pylori-infected gastric cancer cells. *Proc. Natl. Acad. Sci. USA,* **2009**, *106*(34), 14581-14586.

[56] Liu, T.W.; Ho, C.W.; Huang, H.H.; Chang, S.M.; Popat, S.D.; Wang, Y.T.; Wu, M.S.; Chen, Y.J.; Lin, C.H. Identification of essential residues of human α-l-fucosidase and tests of its mechanism. *Biochemistry*, **2008**, *48*(1), 110-120.

[57] Patrick, J.; Willems, H.C.S.; Paul, C. Rossana tonlorenzi and john s o'brien, spectrum of mutations in fucosidosis. *Eur. J. Human Genet.,* **1999**, *7*(1), 60-67.

[58] Yuan, K.; Kucik, D.; Singh, R.K.; Listinsky, C.M.; Listinsky, J.J.; Siegal, G.P. Alterations in human breast cancer adhesion-motility in response to changes in cell surface glycoproteins displaying alpha-L-fucose moieties. *Int. J. Oncol,* **2008**, *32*(4), 797-807.

[59] Lammerts van Bueren, A.; Ardèvol, A.; Fayers-Kerr, J.; Luo, B.; Zhang, Y.; Sollogoub, M.; Blériot, Y.; Rovira, C.; Davies, G.J. Analysis of the reaction coordinate of α-l-fucosidases: A combined structural and quantum mechanical approach. *J. Am. Chem. Soc.,* **2010**, *132*(6), 1804-1806.

[60] Lill, M.A. Multi-dimensional QSAR in drug discovery. *Drug Discov. Today*, **2007**, *12*(23-24), 1013-1017.

[61] Pinto, M.M.; Sousa, M.E.; Nascimento, M.S. Xanthone Derivatives: New insights in biological activities. *Curr. Med. Chem.,* **2005**, 12, 2517-2538.

[62] Negi, J.S.; Singh, V.K.B.P.; Rawat, M.S.M.; Joshi, G.P. Naturally occurring xanthones: Chemistry and biology. *J. App. Chem.,* **2013**.

[63] Liu, Y.; Ke, Z.; Cui, J.; Chen, W.H.; Ma, L.; Wang, B. Synthesis, inhibitory activities, and QSAR study of xanthone derivatives as α- glucosidase inhibitors. *Bioorg. Med. Chem.,* **2008**, *16*(15) 7185-7192.

[64] Liu, Y.; Zou, L.; Ma, L.; Chen, W.H.; Wang, B.; Xu, Z.L. Synthesis and pharmacological activities of xanthone derivatives as α-glucosidase inhibitors. *Bioorg. Med. Chem.,* **2006**, *14*(16), 5683-5690.

[65] Narayana Moorthy N.S.; Ramos, M.J.; Fernandes, P.A. Comparative structural analysis of alpha-glucosidase inhibitors on difference species: A computational study. *Arch. Pharm. Chem. Life Sci.,* **2012**, *345*, 265-274.

[66] Yang, Y.; Qin, J.; Liu, H.; Yao, X. Molecular dynamics simulation, free energy calculation and structure-based 3d-qsar studies of b-raf kinase inhibitors. *J. Chem. Inform. Mod.,* **2011**, *51*(3), 680-692.

[67] Xue, Q.; Zhang, J.L.; Zheng, Q.C.; Cui, Y.L.; Chen, L.; Chu, W.T.; Zhang, H.X. Exploring the molecular basis of dsRNA recognition by Mss116p using molecular dynamics simulations and free-energy calculations. *Langmuir,* **2013**, 29(35), 11135-11144.

[68] Cui, Y.L.; Zheng, Q.C.; Zhang, J.L.; Xue, Q.; Wang, Y.; Zhang H.X. Molecular dynamic investigations of the mutational effects on structural characteristics and tunnel geometry in CYP17A1. *J. Chem. Inform. Mod,* **2013**, *53*(12), 3308-3317.

[69] Xu, L.; Sun, H.; Li, Y.; Wang, J.; Hou, T. Assessing the performance of MM/PBSA and MM/GBSA methods. 3. The impact of force fields and ligand charge models. *J. Phys. Chem. B,* **2013**, *117*(28), 8408-8421.

[70] Pang, X.; Zhou, H.X. Poisson-Boltzmann Calculations: Van der waals or molecular surface? *Commun. Comput. Phys.,* **2013**, *13*(1), 1-12.

[71] Alonso, H.; Bliznyuk, A.A.; Gready, J.E. Combining docking and molecular dynamic simulations in drug design. *Med. Res. Rev.,* **2006**, **26**(5), 531-568.

[72] McInnes, C. Virtual screening strategies in drug discovery. *Curr. Opin. Chem. Biol.,* **2007**, *11*(5), 494-502.

[73] Sun, H.; Li, Y.; Tian, S.; Xu, L.; Hou, T. Assessing the performance of MM/PBSA and MM/GBSA methods. 4. Accuracies of MM/PBSA and MM/GBSA methodologies evaluated by various simulation protocols using PDBbind data set. *Phys. Chem. Chem. Phys.,* **2014**, *16*(31) 16719-16729.

[74] Jorgensen, W.L.; Thomas, L.L. Perspective on free-energy perturbation calculations for chemical equilibria. *J. Chem. Theory Comput.,* **2008**, *4*(6), 869-876.

[75] Chipot, C. Frontiers in free-energy calculations of biological systems. Wiley Interdisciplinary Reviews: *Comput. Mol. Sci.,* **2014**, *4*(1), 71-89.

[76] Meyer, E.F.; Swanson, S.M.; Williams, J.A. Molecular modelling and drug design. Pharmacology & Therapeutics, **2000**, *85*(3), 113-121.

[77] Caflisch, A. Focal Point: Medicinal Chemistry. *Chimia,* 1999. *50,* 35-36.

[78] Redhu, S.; Jindal, A. Molecular modelling: A new scaffolding for drug design. *Int. J. Pharm. Pharmaceut. Sci.,* **2013**, 5(1).

[79] Friedman, R.; Boye, K.; Flatmark, K. Molecular modelling and simulations in cancer research. *Biochim. Biophys. Acta,* **2013**, 1836(1), 1-14.

[80] Berecibar, A.; Grandjean, C.; Siriwardena, A. Synthesis and biological activity of natural aminocyclopentitol glycosidase inhibitors: Mannostatins, trehazolin, allosamidins, and their analogues. *Chem. Rev.,* 1999. *99*, 779-844.

[81] Germer, A.; Peter, M.G.; Kleinpeter, E. Solution-State conformational study of the hevamine inhibitor allosamidin and six potential inhibitor analogues by NMR spectroscopy and molecular modeling. *J. Org. Chem.,* **2002**, *67*(18), 6328-6338.

[82] Bortone, K.; Monzingo, A.F.; Ernst, S.; Robertus, J.D. The Structure of an allosamidin complex with the coccidioides immitis chitinase defines a role for a second acid residue in substrate assisted mechanism. *J. Mol. Biol.,* **2002**, *320*(2), 293-302.

[83] Kawatkar, S.P. Molecular modeling insights into the inhibition and catalytic mechanism of golgi alpha-mannosidase II, **2006**, University of Georgia Athens, Georgia.

[84] Elbein, A.D.; Tropea, J.E.; Mitchell, M.; Kaushal, G.P. Kifunensine, a potent inhibitor of the glycoprotein processing mannosidase I. *J. Biol. Chem.*, **1990**, *265*(26), 15599-15605.

[85] Michael, J.P. Indolizidine and quinolizidine alkaloids. *Nat. Prod. Rep.,* **2004**, *21*(5), 625-649.

[86] Borges de Melo, E.; Da Silveira Gomes, A.; Carvalho, I. α- and β- Glucosidase inhibitors: Chemical structure and biological activity. *Tetrahedron,* **2006**, *62*(44), 10277-10302.

[87] Da Rocha A.B.; Lopes, R.M.; Schwartsmann, G. Natural products in anticancer therapy. *Curr. Opin. Pharmacol.,* **2001**, *1*, 364-369.

[88] Shah, N.; Kuntz, D.A.; Rose, D.R. Comparison of kifunensine and 1-deoxymannojirimycin binding to class I and II alphamannosidases demonstrates different saccharide distortions in inverting and retaining catalytic mechanisms. *Biochemistry*, **2003**, *42*(47), 13812-13816.

[89] Wen, X.; Yuan, Y.; Kuntz, D.A.; Rose, D.R.; Pinto, B.M. A combined STD-NMR/molecular modeling protocol for predicting the binding modes of the glycosidase inhibitors kifunensine and salacinol to Golgi alpha-mannosidase II. *Biochemistry*, **2005**, *44*(18), 6729-6737.

[90] Nobili, S.; Lippi, D.; Witort, E.; Donnini, M.; Bausi, L.; Mini, E.; Capaccioli, S. Natural compounds for cancer treatment and prevention. *Pharmacol. Res*., **2009**, 59(6), 365-378.

[91] Kawatkar, S.P.; Kuntz, D.A.; Woods, R.J.; Rose, D.R.; Boons GJ. Structural basis of the inhibiton of golgi alpha-mannosidase II by Mannostatin A and the role of the thiomethyl moiety in ligandprotein interactions. **2006**.

[92] Kuntz, D.A.; Zhong, W.; Guo, J.; Rose, D.R.; Boons, G.J. The molecular basis of inhibition of Golgi alpha-mannosidase II by mannostatin A. *Chem.Bio.Chem.,* **2009**, *10*(2) 268-277.

[93] Francesca, C.; Andrea, G.; Alberto, B.; Maria, S.; Neri, N.; Stefano, M. Molecular dynamics simulations on the complexes of Glucoamylase II (471) from Aspergillus awamori var. X100 with 1-Deoxynojirimycin and lentiginosine. *Mol. Mod. Ann.,* **1997**, *3*(7), 249-260.

[94] Jin-Ming, Z.; Jun-Hong, Z.; Yi, M.; Min-Bo, C. Molecular dynamics simulation of iminosugar inhibitor−glycosidase complex: insight into the binding mechanism of 1-deoxynojirimycin and isofagomine toward β-Glucosidase. *J. Chem. Theor. Comput.,* **2005**, *2*(1), 157-165.

[95] Karplus, M.; Kushick, J.N. Method for estimating the configurational entropy of macromolecules. *Macromolecules,* **1981**, *14*(2) 325-332.

[96] Ilario, G.T.; René, S.; Paul, E.S.; Wilfred, F.; Van, G. A generalized reaction field method for molecular dynamics simulations. *J. Chemi. Phys.,* **1995**, *102*(13), 5451-5459.

[97] Pereira, C.S.; Kony, D.; Baron, R.; Müller, M.; Van Gunsteren, W.F.; Hünenberger, PH.Conformational and dynamical properties of disaccharides in water: A molecular dynamics study. *Biophys. J.,* **2006**, *90*(12), 4337-4344.

[98] Kräutler, V.M.; Müller, M.; Hünenberger, P.H. Conformation, dynamics, solvation and relative stabilities of selected β-hexopyranoses in water: a molecular dynamics study with the gromos 45A4 forcefield. *Carbohydrate Res.,* **2007**, *342*(14), 2097-2124.

[99] Lerbret, A.; Bordat, P.; Affouard, F.; Descamps, M.; Migliardo, F. How homogeneous are the trehalose, maltose, and sucrose water solutions? An insight from molecular dynamics simulations. *J. Phys. Chem. B,* **2005**, *109*(21), 11046-11057.

[100] Sastry, G.M.; Agmon, N. Trehalose prevents myoglobin collapse and preserves its internal mobility. *Biochemistry*, 1997. *36*(23), 7097-7108.

[101] Brady, J.W.; Schmidt, R.K. The role of hydrogen bonding in carbohydrates: Molecular dynamics simulations of maltose in aqueous solution. *J. Phys. Chem.,* **1993**, *97*(4), 958-966.

[102] Qiang, L.; Schmidt, R.K.; Teo, B.; Karplus, P.A.; Brady, J.W. Molecular dynamics studies of the hydration of α,α-Trehalose. *J. Am. Chem. Soc.,* **1997**, *119*(33), 7851-7862.

[103] Minoru, S.; Masashi, M.; Yoshio, I.; Akihiro, H.; Syouichi, K. Molecular-dynamics study of aqueous solution of trehalose and maltose: Implication for the biological function of trehalose. *Bull. Chem. Society Japan,* **1997**, *70*(4), 847-858.

[104] Roberts, C.J.; Debenedetti, P.G. Structure and dynamics in concentrated, amorphous carbohydrate−water systems by molecular dynamics simulation. *J. Physical Chem. B,* **1999**, *103*(34), 7308-7318.

[105] Conrad, P.B.; De Pablo, J.J. Computer simulation of the cryoprotectant disaccharide α,α-Trehalose in aqueous solution. *J. Physical Chem.* A, **1999**, *103*(20), 4049-4055.

[106] Bonanno, G.; Noto, R.; Fornili, S.L. Water interaction with [small alpha],[small alpha]-trehalose: Molecular dynamics simulation. *J. Chem. Soc, Faraday Trans.,* **1998**, *94*(18), 2755-2762.

[107] Naidoo, K.J.; Kuttel, M. Water structure about the dimer and hexamer repeat units of amylose from molecular dynamics computer simulations. *J. Comput. Chem.,* **2001**, *22*(4), 445-456.

[108] Engelsen, S.B.; Monteiro, C.; Hervé de Penhoat, C.; Pérez, S. The diluted aqueous solvation of carbohydrates as inferred from molecular dynamics simulations and NMR spectroscopy. *Biophys. Chem.,* **2001**, 93(2-3),103-127.

[109] Ekdawi-Sever, N. Diffusion of Sucrose and α,α-Trehalose in Aqueous Solutions. *J. Phy. Chem. A,* **2003**, *107*(6), 936-943.

[110] Molinero, V.; Çağın, T.; Goddard Iii, W.A. Sugar, water and free volume networks in concentrated sucrose solutions. *Chem. Phy. Lett.*, **2003**, *377*(3-4), 469-474.

[111] Bordat, P.; Lerbret, A.; Demaret, J.P.; Affouard, F.; Descamps, M. Comparative study of trehalose, sucrose and maltose in water solutions by molecular modelling. *Euro.Phys. Lett.,* **2004**, *65*(47).

[112] Sum, A.K.; Faller, R.; Pablo, J.J.D. Molecular simulation study of phospholipid bilayers and insights of the interactions with disaccharides. *Biophys. J.*; **2003**, *85*(5), 2830-2844.

[113] Pereira, C.S.; Lins, R.D.; Chandrasekhar, I.; Freitas, L.C Hünenberger, P.H. Interaction of the disaccharide trehalose with a phospholipid bilayer: A molecular dynamics study. *Biophys. J.,* 2004. *86*(4), 2273-2285.

[114] Pereira, C.S.; Hünenberger, P.H. Interaction of the Sugars trehalose, Maltose and glucose with a phospholipid bilayer: A comparative molecular dynamics study. *J. Phys. Chem. B,* **2006**, *110*(31), 15572-15581.

[115] Lerbret, A. Molecular dynamics simulations of lysozyme in water/sugar solutions. *Chem. Phys.,* **2008**, *345*(2-3), 267-274.

[116] Venerando, P.; Giuseppe, M.L.; Antonio, R.; Alessia, B.; Felicia D', A.; Francesco, P. Experimental and in silico characterization of a biologically active inosose. *Struct. Chem.*, **2013**, *24*(3) 955-965.

Chapter 5

Understanding the cross-resistance of oseltamivir to H1N1 and H5N1 influenza A neuraminidase mutations using multidimensional computational analyses

Ashona Singh, Mahmoud E Soliman

School of Health Sciences, University of KwaZulu-Natal, Westville, Durban, South Africa

Correspondence: Mahmoud E Soliman

School of Health Sciences, University of KwaZulu-Natal, Westville, Campus, Private Bag X54001, Durban 4000, South Africa

Tel +27 31 260 7413, Fax +27 31 260 7413, Email soliman@ukzn.ac.za

Abstract: This study embarks on a comprehensive description of the conformational contributions to resistance of neuraminidase (N1) in H1N1 and H5N1 to oseltamivir, using comparative multiple molecular dynamic simulations. The available data with regard to elucidation of the mechanism of resistance as a result of mutations in H1N1 and H5N1 neuraminidases is not well established. Enhanced post-dynamic analysis, such as principal component analysis, solvent accessible surface area, free binding energy calculations, and radius of gyration were performed to gain a precise insight into the binding mode and origin of resistance of oseltamivir in H1N1 and H5N1 mutants. Three significant features reflecting resistance in the presence of mutations H274Y and I222K, of the protein complexed with the inhibitor are: reduced flexibility of the α-carbon backbone; an improved ΔE_{ele} of ~15 (kcal/mol) for H1N1 coupled with an increase in ΔG_{sol}(~13 kcal/mol) from wild-type to mutation; a low binding affinity in comparison with the wild-type of ~2 (kcal/mol) and ~7 (kcal/mol) with respect to each mutation for the H5N1 systems; and reduced hydrophobicity of the overall surface structure due to an impaired hydrogen bonding network. We believe the results of this study will ultimately provide a useful insight into the structural landscape of neuraminidase-associated binding of oseltamivir. Furthermore, the results can be used in the design and development of potent inhibitors of neuraminidases.

Keywords: neuraminidase, molecular dynamics, resistance, mutation, binding free energy

Introduction

The rapid evolution of highly pathogenic and resistant variants of influenza A viruses, H5N1 and H1N1, takes us to the precipice of a possible new-age influenza pandemic. The first human influenza pandemic documented in the 20th century was the 1918–1919 "Spanish flu", which claimed the lives of approximately 50 million people worldwide.[1–4] Since 1997, the highly pathogenic avian influenza A (H5N1) has had a detrimental effect on the socioeconomic development in countries across Africa, Asia, and the Middle East. Pharmacological advances in viable influenza chemotherapeutics and vaccines could not have predicted the influenza A (H1N1) virus outbreak in 2009–2010, which claimed an estimated 284,500 lives. The viruses H5N1 and H1N1 display similar pathogenesis of respiratory tract infection in humans. However, statistically H5N1 has a higher mortality rate.[5]

Influenza virus A is highly communicable. H1N1 is easily transmissible between humans, while avian influenza has not yet adapted sufficiently to facilitate human-human spread; the spread of H5N1 is largely dependent on close avian-human contact. Based on the research, the H1N1 and H5N1 viruses can both trace their lineage back to pigs as potential intermediate hosts. Both viruses can undergo reassortment in the intermediate host prior to infection in the new host.[6,7] This raises the concern that more virulent and highly transmissible influenza virus strains may infect humans.[8]

Both the H1N1 and H5N1 viruses express two main surface antigens, H5/H1 (hemagglutinin type 5/1, respectively) and N1 (neuraminidase type 1).[9–11] Neuraminidase has been the main target for potential pharmacological interventions.[12] It plays a vital role in the spread of infection through cleavage of the viral receptor, sialic acid, from both viral and host proteins. This results in the release of newly replicated virus from an infected cell.[12,13] The active site residues of the neuraminidase enzyme are highly conserved within all virus types and subtypes, ie, designed neuraminidase inhibitors may be effective against types A and B of the influenza virus.[14–17]

There are four effective neuraminidase inhibitors currently available: zanamivir,[18] laninamivir,[19] peramivir,[20] and oseltamivir.[21] Zanamivir is commercially known as Relenza® and is administered as an aerosol into the nasal cavity. Zanamivir has been approved in many countries since 1999/2000. Laninamivir is administered in the form of an inhaler and was made available to the Japanese market in 2010 (under the trade name Inavir®). Peramivir, commercially known as PeramiFlu®, is administered by intravenous drip and is approved in Japan, the Republic of Korea, and the People's Republic of China. Oseltamivir commercially known as Tamiflu® (Figure 1), has long been the "drug of choice" due to ease of oral

administration, improved bioavailability, and easy accessibility. There has been an increase in the number of reports describing oseltamivir resistance in H1N1 and H5N1, and the rapid development of resistance toward the drug poses a threat in the event of an outbreak.[22,23]
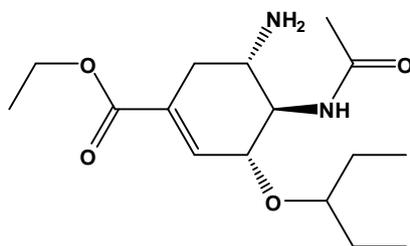

Figure 1: Structure of drug oseltamivir

Research directed toward gaining an insight into the mechanism of resistance of influenza viruses is imperative in the future design of drugs as well as in the development of pre-emptive measures against possible mutations. The comparative multiple molecular dynamic (MD) simulations conducted in this study offer an atomistic perspective into the complex nature of protein motions as a function of time.[24,25] Due to H5N1 virus sharing a 91.47% sequence identity and a conserved binding site with the 2009 H1N1, the results of molecular simulation may apply to both viruses.[26]

Several computational approaches have been used in an attempt to understand the impact of mutations on oseltamivir resistance to neuraminidase. Karthick *et al* initiated one of the first MD studies, which was based solely on the natural mutation H274Y of H1N1 isolated during the 2007–2008 Eurasia influenza season.[27,28] Malaisree *et al* refined the source of oseltamivir resistance in avian influenza H5N1 virus with the H274Y mutation using MD simulation.[29,30] Wang and Zheng performed a similar study that incorporated the mutations N294S and E119G.[31] In the above studies, it was found that although residue 274 is a framework residue in the active site,[21] the mutation of histidine (His) to tyrosine (Tyr) significantly impaired the binding of oseltamivir because the binding pocket could not accommodate the steric bulk of the drug. Further to this, the presence of additional mutations was found to amplify resistance of the neuraminidase enzyme to oseltamivir. A large database of crystallized glycoprotein of influenza virus A and B exists, each with a different mutation in the neuraminidase amino acid sequence.[26,27,32–34] The MD studies for each mutation revealed enhanced surface expression of oseltamivir-resistant influenza neuraminidase.[35] In many instances, these included zanamivir and peramivir resistance.[36]

Nguyen *et al*[37] and Huang *et al*[38] identified a mutation at position 222 from isoleucine to lysine (Lys) in strains of wild-type (WT) and H274Y mutants of H1N1 and H5N1. The point mutation from isoleucine to Lys exhibited increased resistance toward oseltamivir. Structural

residue I222 of H1N1 and H5N1 neuraminidase offered stability to oseltamivir binding.[39,40] Structure stability was reinforced by the presence of the hydrophobic pocket created by E276 and R224, which in turn supported a seven hydrogen bond interaction between the active site residues and the drug.[30,36]

Insufficient information about the origin of resistance of oseltamivir against neuraminidase H1N1 and H5N1 mutations, I222K and H274Y (Figure 2) prompted us to perform a comprehensive multidimensional analysis using a multiple MD approach. Multiple MD simulations have demonstrated better sampling efficiency.[41] Multiple trajectories with different initial conditions improved the conformational sampling of MD simulations in proteins.[42,43] Several post-dynamic analyses, such as root mean square fluctuation (RMSF), root mean square deviation (RMSD), free binding energy calculations, radius of gyration, solvent accessible surface area (SASA), and principal component analysis (PCA) were performed in order to gain a comprehensive understanding of the impact of mutations on binding and the conformational landscape of the oseltamivir–protein complex. Our research should contribute to understanding the effect of resistance and provide insight into the future development of innovative chemotherapeutics.
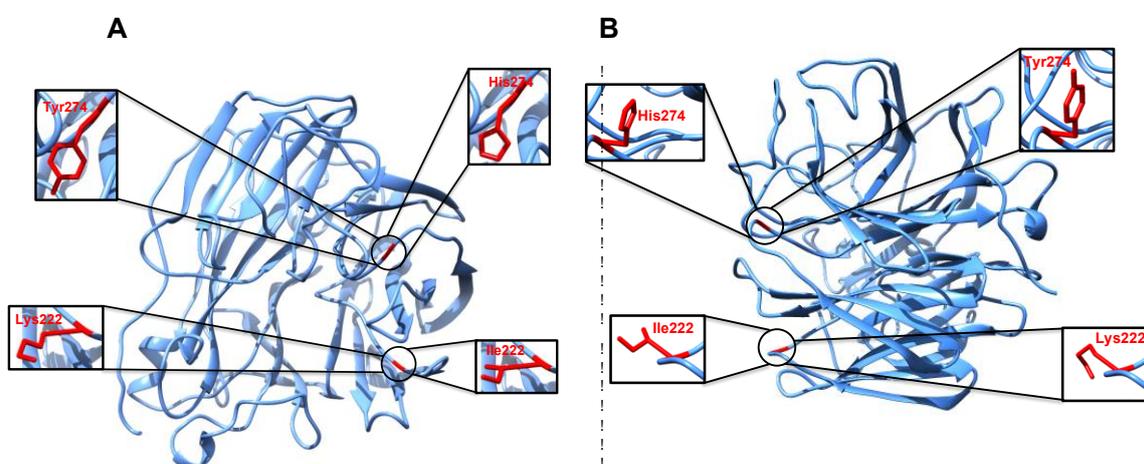

Figure 2: Three dimensional structure of H5N1 and H1N1 neuraminidase, A and B, respectively, showing the positions of the studied mutations.

Computational methods

System preparation

The X-ray crystal structures of H1N1 and H5N1 WT and mutant neuraminidase were extracted from Protein Data Bank (PDB) codes 4B7R (WT-H1N1),[44] 2HU4 (WT-H5N1),[15] and 3CL0 (H274Y mutation-H5N1)[45] complexed with oseltamivir, obtained from the Research Collaboratory for Structural Bioinformatics. Neuraminidase is a tetrameric enzyme, using the Chimera software package[46] a single subunit was selected with the inclusion of an active site drug complex to reduce computational cost. Influenza virus, H1N1 PDB code 4B7R, formed the WT sequence with a His274 residue. Using Chimera, point mutations were introduced at positions 222 (from I to K) and 274 (from H to Y). Table 1 lists the three enzymes with relevant mutations resulting in eight simulations.

Table 1: The crystal structure of the simulated systems, PDB codes and abbreviations

Simulated system	PDB code	Abbreviation*
H1N1 wild-type	4B7R	WT_{H1N1}
H1N1 I222K	4B7R	$I222K_{H1N1}$
H1N1 H274Y	4B7R	$H274Y_{H1N1}$
H1N1 H274Y-I222K	4B7R	$H274Y\text{-}I222K_{H1N1}$
H5N1 wild-type	2HU4	WT_{H5N1}
H5N1 I222K	2HU4	$I222K_{H5N1}$
H5N1 H274Y	3CL0	$H274Y_{H5N1}$
H5N1 H274Y-I222K	3CL0	$H274Y\text{-}I222K_{H5N1}$

* These abbreviations are used throughout the manuscript

MD simulation

Multiple MD simulations were performed to establish the impact of mutation I222K on the binding of oseltamivir to the WT enzyme and in the presence of mutation H274Y of H1N1 and H5N1. A long continuous MD trajectory may incur greater statistical errors as the protein denatures and evolves from one conformation to another during the time of the simulation.[47] Thus, a multiple MD approach was undertaken. This method in turn reduced the force field artefacts, statistical bias, and computational time. A total simulation time of 100 ns partitioned in five distinct 20 ns MD runs was executed, with each trajectory having a different initial velocity. The multiple MD protocol applied in this study is described in Figure 3.
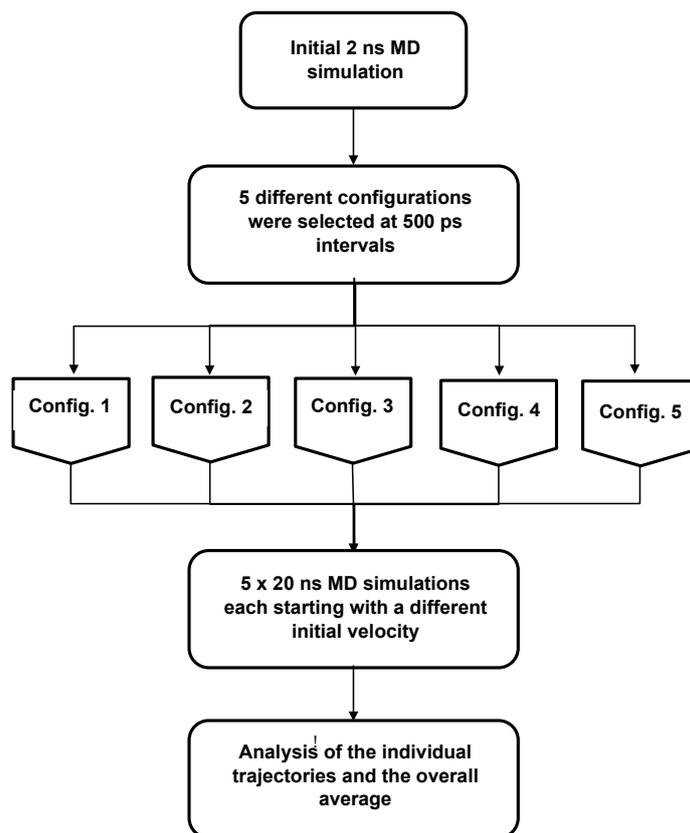
Figure 3: Multiple MD trajectory approach adopted in this report

MD simulation set-up and parameters

The MD simulation was performed using the graphics processing unit version of the PMEMD engine provided with the Amber 12 and 14 package.[48,49] The FF99SB variant of the Amber force field[50] was used to describe the protein. The LEAP module of Amber 12/14 allowed for the addition of hydrogen atoms to the protein and three Na+ counter ions for neutralization. The system was suspended within a TIP3P[51] water box such that all protein atoms were within 8 Å of any box edge. After energy minimization, a 1,000 step restraint gradient minimization was performed. Selective boundary conditions were enforced concurrently with the particle-mesh Ewald method,[52] a component of Amber 12/14, with set parameters of direct space and a van der Waals cut-off of 12 Å. This would be conducive to the treatment of long-range electrostatic interactions. The system passed through an initial energy minimization with 2,500 steps of steepest descent and a restraint harmonic potential of 500 kcal/mol Å2 being applied to the solute. After this, an additional 1,000 step unrestrained conjugate gradient energy minimization was carried out on the complete system, i.e., protein, ligand, solvent molecules, and added ions. Gradual heating of the MD simulation from 0 to 300 K was

executed for 50 ps, such that the system maintained a fixed number of atoms and a fixed volume, i.e., a canonical (NVT) ensemble. The solutes within the system were imposed with a potential harmonic restraint of 10 kcal/mol $Å^2$ and a collision frequency of 1.0 ps^{-1}. Superseding the heating input, a final equilibration estimating 500 ps of the systems was required. The operating temperature was kept constant at 300 K, accompanied by the number of atoms and pressure, mimicking an isobaric-isothermal (NPT) ensemble. The systems pressure was maintained at 1 bar using the Berendsen barostat.[53] The total time for each MD simulation was 20 ns, and five trajectories for each of the eight systems were generated. In each simulation, the SHAKE algorithm[54] was employed to constrict the bonds of the hydrogen atoms. The time step of each simulation was 2 fs and a single-precision floating-point precision model[55] was used. The simulations coincided with randomized seeding of the NPT ensemble. It was accompanied by constant pressure of 1 bar maintained by the Berendsen barostat, a pressure coupling constant of 2 ps, with a temperature of 300 K and Langevin thermostat with a collision frequency of 1.0 ps^{-2}. Co-ordinates were saved every 1 ps and the trajectories were analysed every 1 ps using the PTRAJ module implemented in Amber 12.0 and CPPTRAJ module in Amber 14.0.

Post-dynamic analysis

Thermodynamic calculations

An implicit solvent model was employed to describe the thermodynamic free binding energies of oseltamivir bound to WT and mutant H1N1 and H5N1 neuraminidase in this study. The molecular mechanics/generalized Born surface area method was used to evaluate the ligand-protein complex binding affinities.[56–59] To calculate the free binding energy contributions, 1,000 snapshots were extracted from each of the 20 ns trajectories. The following set of equations describes the calculation of the binding free energy:

$$\Delta G_{bind} = G_{complex} - G_{receptor} - G_{ligand} \qquad (1)$$

$$\Delta G_{bind} = E_{gas} + G_{sol} - TS \qquad (2)$$

$$E_{gas} = E_{int} + E_{vdw} + E_{ele} \qquad (3)$$

$$G_{sol} = G_{GB} + G_{SA} \qquad (4)$$

$$G_{SA} = \gamma SASA \qquad (5)$$

Where E_{gas} represents the gas-phase energy, E_{int} is the internal energy, E_{ele} is the Coulomb energy, and E_{vdW} is the van der Waals energy. The term E_{gas} is directly measured from the FF99SB force field terms. The solvation energy (G_{sol}) is the summation of contributions from polar and nonpolar states. The polar solvation energy contribution, G_{GB}, is derived from solving the GB equation. The term G_{SA} corresponds to the non-polar solvation energy contribution, which is estimated from the SASA determined using a water probe radius of 1.4 Å. The temperature and total solute entropy are represented by T and S, respectively.[60]

Principal component analysis

PCA reveals the structure of atomic fluctuations, and describes the motion of the system in terms of eigenvectors (planar of motion) and eigenvalues (magnitude of motion).[61] The individual MD trajectories were stripped of solvent and ions using the PTRAJ and CPPTRAJ modules in Amber 12.0/14.0. The resulting trajectories were aligned against a fully minimized structure. PCA was performed on a Cα backbone with 1,000 snapshots taken every 20 ps. The first two eigenvectors (PC1 and PC2) corresponding to the first two modes of PCA covariance matrices were generated using in-house scripts. An average of the PC1 and PC2 for the 5×20 ns trajectories of the H1N1 and H5N1 WT and mutant systems was generated. The corresponding PCA scatters were plotted using Origin software (http://www.originlab.com/) and structural postscript diagrams were created using visual MDs.[62] Porcupine plots of the first and second modes developed by the normal mode wizard using the ProDy interface of visual MDs were sketched for each of the systems.[63]

Results and discussion

MD simulations and system stability

RMSD and potential energy plots in the Supplementary materials (Figures S1–12 of H5N1 and H1N1 neuraminidase, respectively) graphically monitor the convergence of the studied systems.[64] All the systems of H5N1 and H1N1 influenza viruses converge at approximately 5,000 ps by both RMSD and potential energy calculations.

Post-dynamic analysis

RMSF and radius of gyration analysis was used to relate conformational changes and plasticity of the Cα backbone to mutation of each system.

Root mean square fluctuation

The RMSF captured the fluctuation of each residue, providing insight into the flexible regions of the protein that correspond to crystallographic β (temperature) factors.[65] Figure 4 illustrates the fluctuations of systems WT_{H5N1} and $H274Y_{H5N1}$. The crystal structure of the WT neuraminidase consists of three α-helices and 24 β-strands. Evidence from the graph demonstrates an unfolding of the protein helix at residue 28 in $H274Y_{H5N1}$ in comparison with the WT, shortening the cardinal α-helix and β-strand by one and three residues, respectively. Closer inspection revealed a distinct change in fluctuation corresponding to region 108–144, possibly due to the presence of a hydrogen bond between Lys125 and Thr131. A similar effect was observed through β-strands 156–162 and 170–177 which share a hydrogen bond between amino acids Thr161 and Tyr171. The residue region 260–320 containing the compensatory mutation did not demonstrate large variations in fluctuations.

As depicted in Figure S13, by interacting with residues in the active site, the role of residue 222 appeared to shift from a structural one to a functional one. The Lys222 mutation contains a second amino group that acts as an electrophile at physiological pH; this prompts an interaction with the active site residues. Mutation $I222K_{H5N1}$ caused a change in orientation of Gly332, promoting a disulphide link between parallel cysteine residues 339 and 364 and causing a shift in the β-strand orientation.[66]

Analysis of Figure S14 revealed a migration from a system of high flexibility ($H274Y_{H5N1}$) to one of low flexibility ($H274Y$-$I222K_{H5N1}$). The change from a hydrophobic residue, His, to the aromatic Tyr274, which contains an ionisable phenolic group, introduced instability in a previously predominant hydrophobic region. Therefore, it is plausible to assume that the presence of mutation I222K in system $H274Y$-$I222K_{H5N1}$ restricts interaction with the external solvent, thus minimizing flexibility throughout the protein.

In the $H274Y_{H1N1}$ system (Figure 5), an increase in fluctuation at positions 58–65, 246–260, 284, and 301–305 occurred in comparison with WT_{H1N1}. The H274Y mutation introduced a new hydrogen bond, and the aromatic group imposed a steric bulk disrupting previous hydrophobic interactions within that region. A considerable fluctuation was observed in the latter regions of the protein. There is only a single mutation in the $I222K_{H1N1}$ system (Figure S15), which induced an increase in protein flexibility because Lys enhances solvent interaction. An overlay of the RMSF plots for systems $H274Y$-$I222K_{H1N1}$ and $H274Y_{H1N1}$ (Figure S16) revealed no distinct difference between the two systems, indicating that no potential compensatory effect occurred to accommodate the double mutant species of protein $H274Y$-$I222K_{H1N1}$. Thus, the conformation of the protein remains unchanged.

Figure 4: RMSF comparison of WT_{H5N1} and $H274Y_{H5N1}$: T1, T2, T3, T4, T5 and T_{avg} presenting the 5 individual 20 ns MD trajectories and overall average, respectively.
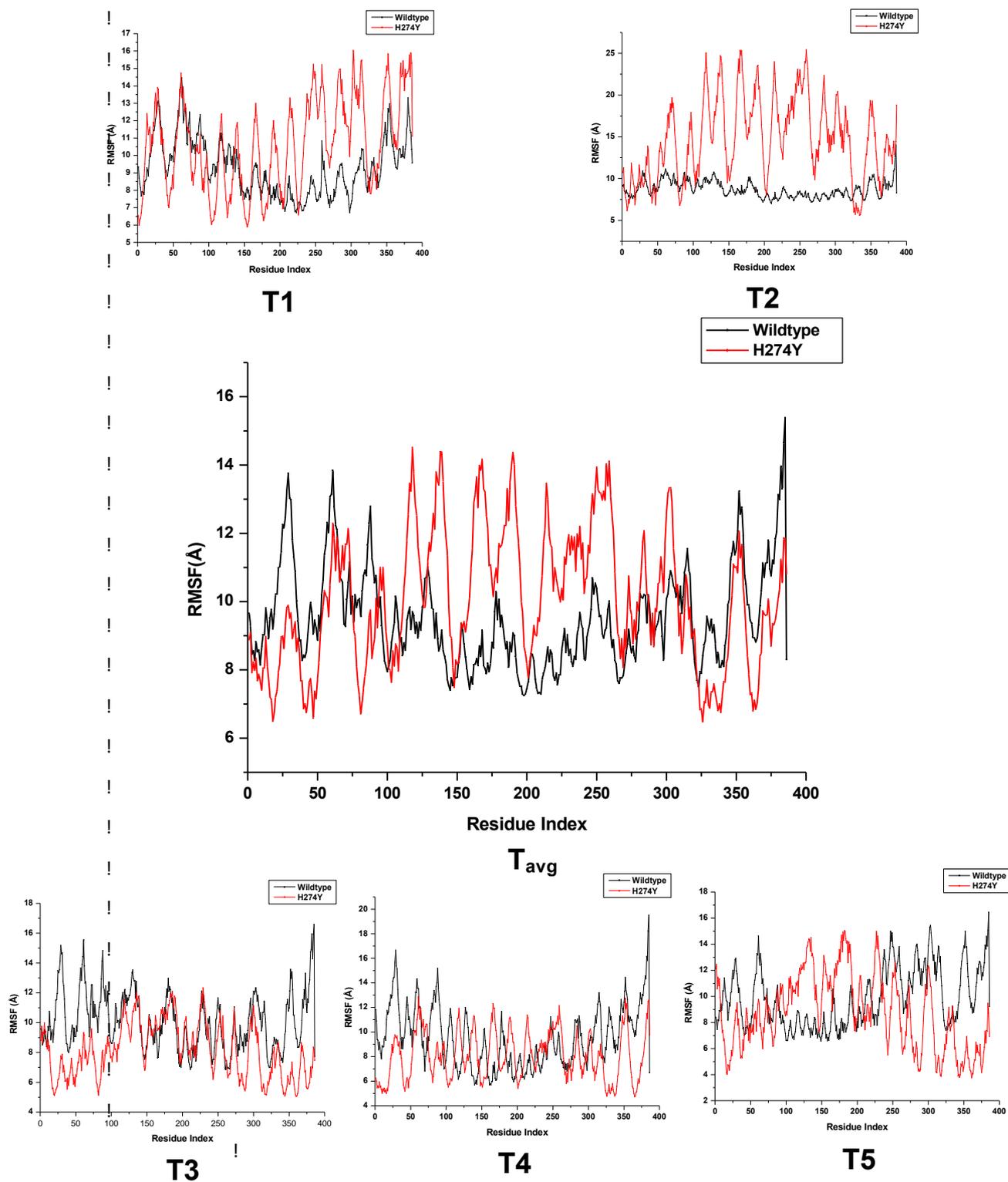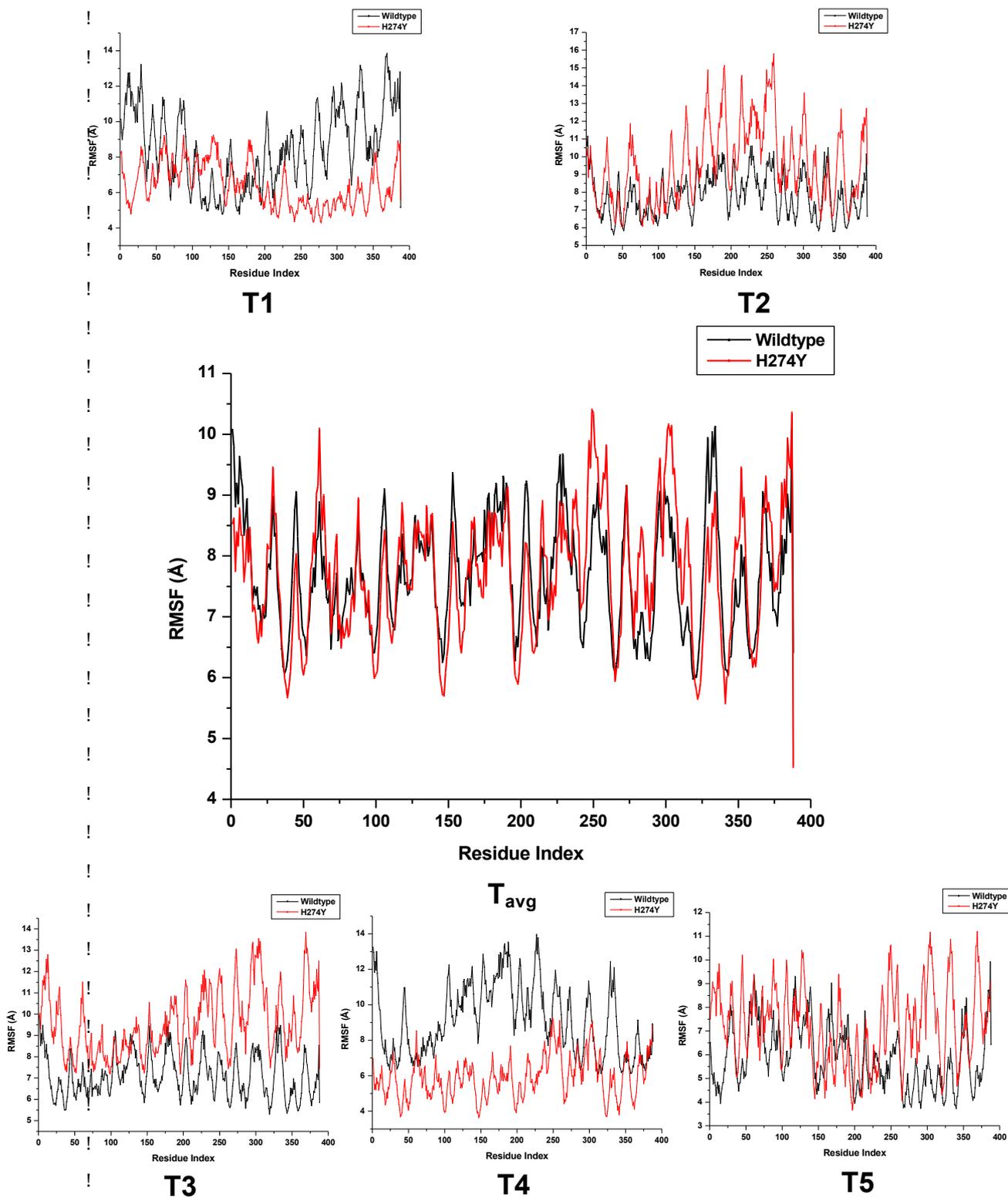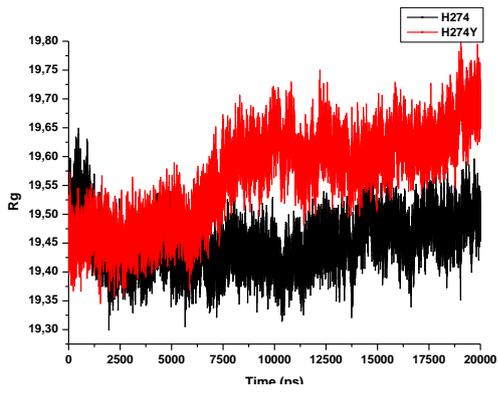
Figure 5: RMSF comparison of WT_{H1N1} and $H274Y_{H1N1}$: T1, T2, T3, T4, T5 and T_{avg} presenting the 5 individual 20 ns MD trajectories and overall average, respectively.
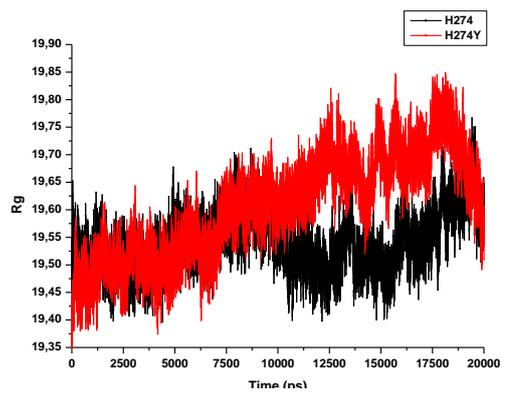
Radius of gyration

Radius of gyration demonstrates the compactness of protein structures, providing insight into complex changes in the molecular shape.[67,68] It is evident that systems WT_{H5N1} and $H274Y_{H5N1}$ (Figure 6) share a similar arrangement of amino acids in secondary and tertiary structures with an almost identical initial compactness. Although both systems share a close resemblance, the $H274Y_{H5N1}$ system showed a distinguishable increase in its radius of gyration over time, transmutating to a less compact structure.

System comparison of WT_{H5N1} and $I222K_{H5N1}$ (Figure S17) unveiled the I222K mutation as the more compact protein complex. Condensation of the structure created a semipermeable complex that prevented unwarranted solvent exposure of interior amino acid residues stabilizing the mutation $I222K_{H5N1}$. Comparative analysis of systems $H274Y_{H5N1}$ and $H274Y\text{-}I222K_{H5N1}$ (Figure S18) proposed an increase in flexibility of system $H274Y\text{-}I222K_{H5N1}$. The compactness of system $H274Y\text{-}I222K_{H5N1}$ attempts to preserve the enzymes bioactivity whilst inferring resistance against oseltamivir. Both systems $H274Y_{H5N1}$ and $H274Y\text{-}I222K_{H5N1}$ share a similar radius of gyration profile, which suggests that both have a similar complex compactness.

System $H274Y_{H1N1}$ showed a larger radius of gyration than the parent protein WT_{H1N1}, indicating that $H274Y_{H1N1}$ is less tightly packed (Figure 7). Solvent interaction within the amino acid sequence perpetuated partial protein unfolding, facilitating a less compacted structure. The single mutation $I222K_{H1N1}$ (Figure S19) shared a similar trend by way of possessing a large radius of gyration value as compared with the WT_{H1N1} system. However, isoleucine expels solvent, and when Lys was introduced, an ionisable species was generated, which, much like the Tyr residue at position 274, was capable of interacting with the surrounding solvent, relieving steric strain. The $H274Y\text{-}I222K_{H1N1}$ system has a folding profile similar to that of $H274Y_{H1N1}$, which is indicative of an undisturbed compaction of structure. The $H274Y\text{-}I222K_{H1N1}$ system (Figure S20) appeared unaffected by the mutation at position 222. This could be due to the presence of the solvent-interacting amino acid species Tyr having previously rendered the inner folding's of the protein susceptible to solvation. The introduction of an additional ionisable moiety better complemented the structure in this conformation.
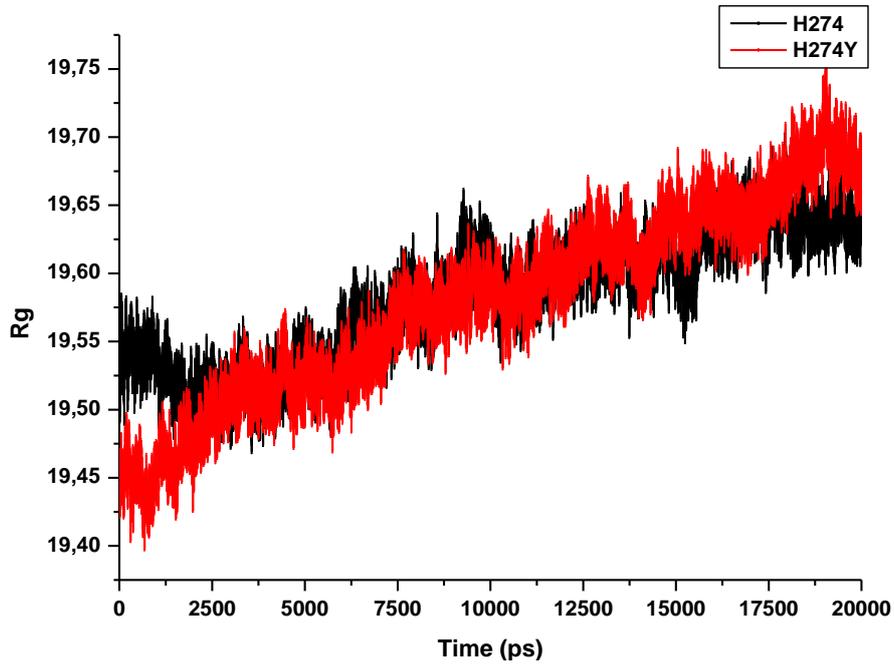
Figure 6: Radius of gyration average comparison across the 20 ns MD simulation of WT$_{H5N1}$ and H274Y$_{H5N1}$: T1, T2, T3, T4, T5 and T$_{avg}$ presenting the 5 individual 20 ns MD trajectories and overall average, respectively.
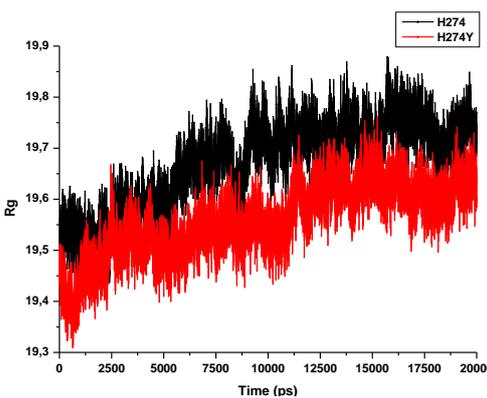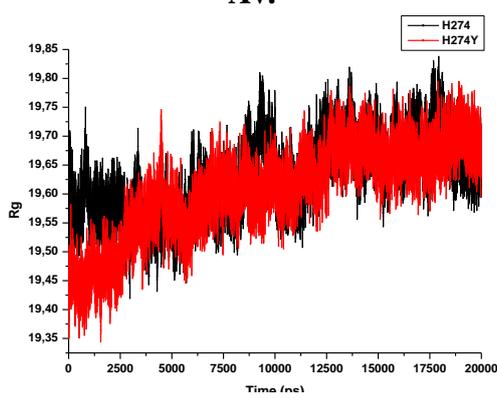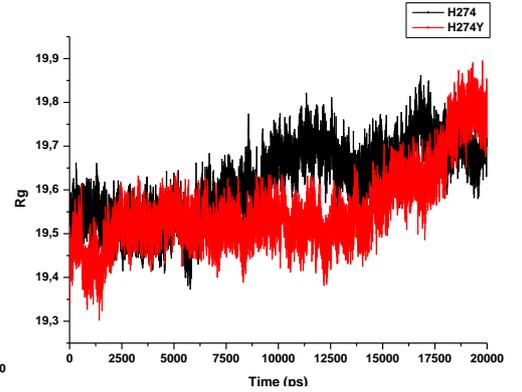
Figure 7: Radius of gyration average comparison across the 20 ns MD simulation of WT$_{H5N1}$ and H274Y$_{H1N1}$: T1, T2, T3, T4, T5 and T$_{avg}$ presenting the 5 individual 20 ns MD trajectories and overall average, respectively.
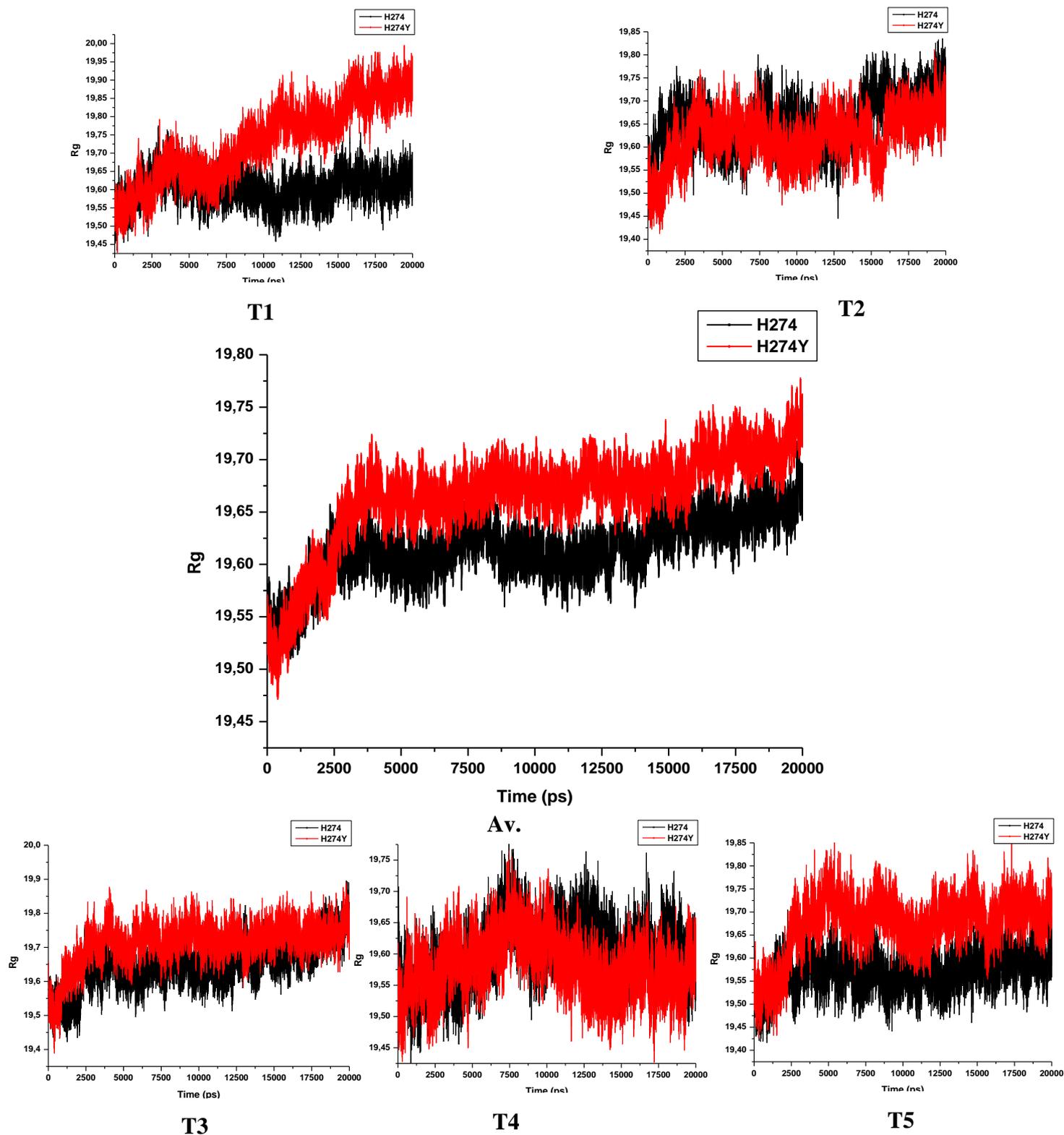
Calculation of MM/GBSA binding free energy

Evaluation of the free binding energies provides a thorough insight into the total energy of each of the systems. Table 2 summarizes the average molecular mechanics/generalized Born surface area (MM/GBSA) thermodynamic energy contributions elicited from the structural data of the five 20 ns MD simulations for the H5N1 and H1N1 systems.

The terms ΔE_{ele} and ΔE_{vdW} represent the electrostatic and van der Waals intermolecular interacting components, respectively, between the protein and inhibitor.[69–70] Systems WT$_{H5N1}$ and I222K$_{H5N1}$ showed a difference of -6.1302 kcal/mol (ΔE_{ele}) and -2.5055 kcal/mol (ΔE_{vdW}). H274Y$_{H5N1}$ and H274Y-I222K$_{H5N1}$ showed a difference of -21.2395 kcal/mol (ΔE_{ele}) and -4.1821 kcal/mol (ΔE_{vdW}). The energy differences advocate a more stable protein–ligand interaction of WT and H274Y$_{H5N1}$. This was measured in terms of effective binding distance and positioning of amino acid residues for optimum interaction between charged entities. The ΔG_{sol}, difference defined between systems WT$_{H5N1}$ and I222K$_{H5N1}$ was estimated to 1.2626 kcal/mol and for systems H274Y$_{H5N1}$ and H274Y-I222K$_{H5N1}$ was estimated to be 13.3466 kcal/mol. The change to polar amino acids (e.g. from Ile to Lys and from His to Tyr) in systems I222K$_{H5N1}$ and H274Y-I222K$_{H5N1}$ contributed to an improved ΔG_{sol} energy difference. Closer inspection of the free binding energy, ΔG_{bind}, quantified the interaction between the protein and ligand. The difference in free binding energy between systems WT$_{H5N1}$ and I222K$_{H5N1}$ was -7.1732 kcal/mol, and that between systems H274Y$_{H5N1}$ and H274Y-I222K$_{H5N1}$ was -11.0751 kcal/mol. The trend in binding free energy difference was observed to be as follows: WT$_{H5N1}$, I222K$_{H5N1}$~H274Y$_{H5N1}$, H274Y-I222K$_{H5N1}$, indicating that the protein-ligand interaction became progressively thermodynamically unfavourable in the presence of mutations.

Table 3 corresponds to the calculated energies for the H1N1 neuraminidase systems. The amine group of oseltamivir was kept positively charged during the MD simulation. According to the literature, binding of drug in a positively charged state to the active site residues is falsely represented by an ameliorated free binding energy when comparing the mutant sequence with the WT sequence.[71] Based on a study by Le *et al* it was proposed that a predominantly negatively charged column of residues at the binding pocket electrostatically funnels oseltamivir to the active site of N1 neuraminidases.[1] This greatly impacts the binding pose, whereby a positively charged drug is drawn toward the epicentre of the active site. The energy contribution ΔG_{GB} of H274Y$_{H1N1}$ and I222K$_{H1N1}$ is testament to the proposed binding

instruction as both systems highlighted an improved binding, with an energy difference from the WT of -4.1784 kcal/mol and -2.4447 kcal/mol, respectively. Surprisingly, the double mutation species H274Y-I222K$_{H1N1}$ demonstrated a similar energy profile trend to WT$_{H1N1}$. The electrostatic energy of the I222K$_{H1N1}$ and H274Y$_{H1N1}$ species differed significantly from the WT$_{H1N1}$ system by -8.3000 kcal/mol and -15.9730 kcal/mol, respectively. However, a remarkable drop in electrostatic energy was observed in the H274Y-I222K$_{H1N1}$ system compared with H274Y$_{H1N1}$. This phenomenon could relate to the conformation of the double mutant system, as the residues interacting with the solvent direct themselves inwardly, interacting with neighbouring amino acid residues. The ΔE_{vdW} and ΔG_{sol} differences between H274Y$_{H1N1}$ and double mutation H274Y-I222K$_{H1N1}$ confirm the funnelling mechanism, as both shared an improvement in the ΔE_{vdW} value over the WT. However, a significant decline in ΔG_{sol} of 37.4933 kcal/mol indicated a lack of solvent interaction. The van der Waals contributions for I222K$_{H1N1}$ suggested a slight decline in hydrophobic interaction, with an energy difference from WT of -1.0594 kcal/mol. A ΔG_{sol} difference between the systems of 4.7960 kcal/mol implied enhanced solvent interaction of the I222K$_{H1N1}$ complex. The H274Y$_{H1N1}$ offered a similar van der Waals contribution to WT$_{H1N1}$, as the aromatic group replaced a linear hydrocarbon chain. Despite this structural feature, a solvation energy difference of 13.1386 kcal/mol indicated that the hydroxyl group of Tyr imposed an increased solvent exposure.

Table 2: Energy contribution derived from MM/GBSA technique corresponding to structural entities of H5N1 system

Complexes	Trajectory	ΔG_{bind}	ΔE_{ele}	ΔE_{vdW}	ΔG_{gas}	ΔG_{sol}
H274	**T1**	-27.80 ± 6.57	-54.97 ± 15.38	-30.25 ± 3.27	-85.21± 17.19	57.42 ± 12.03
	T2	-28.56 ± 3.63	-57.65 ± 7.07	-33.19 ± 2.43	-90.85 ± 7.05	62.29 ± 6.22
	T3	-38.19 ± 4.36	-71.84 ± 6.42	-36.14 ± 2.87	-107.97 ± 6.83	69.79 ± 7.34
	T4	-29.29 ± 7.13	-60.08 ± 8.79	-31.52 ± 3.71	-91.63 ± 10.82	61.34 ± 6.48
	T5	-31.89 ± 5.17	-62.10 ± 10.66	-33.42 ± 2.90	-95.52 ± 11.96	63.64 ± 7.78
	T$_{avg}$	**-31.14 ± 4.34**	**-61.33 ± 9.16**	**-32.91 ± 4.59**	**-94.24 ± 8.70**	**62.89 ± 7.45**
I222K	**T1**	-24.22 ± 6.65	-59.82 ± 11.74	-28.20 ± 4.20	-88.03 ± 12.78	63.80 ± 7.97
	T2	-23.98 ± 5.44	-41.06 ± 12.39	-31.92 ± 3.75	-72.98 ± 14.35	48.99 ± 9.56
	T3	-23.08 ± 4.42	-56.77 ± 11.18	-31.16 ± 3.60	-87.93 ± 12.45	64.85 ± 9.48
	T4	-26.51 ± 7.93	-69.03 ± 8.79	-31.44 ± 3.22	-100.47 ± 9.68	73.96 ± 7.84
	T5	-22.05 ± 8.68	-49.31 ± 17.41	-29.29 ± 4.90	-78.60 ± 20.95	56.55 ± 23.77

	T$_{avg}$	**-23.97 ± 4.89**	**-55.20 ± 11.68**	**-30.40 ± 6.44**	**-85.60 ± 14.61**	**61.63 ±8.71**
H274Y	T1	-30.80 ± 4.62	-60.00 ± 8.22	-31.74 ± 2.66	-91.75 ± 8.57	60.95 ± 5.98
	T2	-28.53 ±3.97	-56.79 ± 8.77	-29.99 ± 3.23	-86.78 ± 9.48	58.24 ± 7.31
	T3	-30.80 ± 4.62	-60.00 ± 8.22	-31.74 ± 2.66	-91.75 ±8.57	60.95 ±5.98
	T4	-28.53 ± 3.97	-56.79 ± 8.77	-29.99 ± 3.23	-86.78 ± 9.48	58.24 ± 7.31
	T5	-28.53 ± 3.97	-56.79 ± 8.77	-29.99 ± 3.23	-86.78 ± 9.48	58.24 ± 7.31
	T$_{avg}$	**-29.44 ± 3.44**	**-58.07 ± 6.90**	**-30.69 ± 4.11**	**-88.76 ± 8.68**	**59.33 ± 6.11**
H274Y-I222K	T1	-16.86 ± 4.85	-39.61 ± 9.57	-28.15 ± 3.75	-67.76 ± 9.47	50.90 ± 8.06
	T2	-19.70 ± 5.25	-45.19 ± 11.75	-24.53 ± 3.98	-69.72 ± 13.85	50.02 ± 9.90
	T3	-15.41± 6.14	-21.27 ± 8.69	-25.64 ± 7.34	-46.91 ± 13.87	31.50 ± 9.50
	T4	-20.22 ± 3.68	-50.65 ±8.41	-23.39 ± 2.77	-79.04 ± 8.68	58.83 ± 7.00
	T5	-19.63 ± 3.26	-27.45 ± 9.69	-30.82 ± 2.68	-58.27 ± 9.66	38.64 ± 8.05
	T$_{avg}$	**-18.36 ± 3.98**	**-36.84 ± 9.32**	**-26.51 ± 5.51**	**-64.34 ± 10.78**	**45.98 ± 7.54**

Table 3: Energy contribution derived from MM/GBSA technique corresponding to structural entities of H1N1 system

Complexes	Trajectory	ΔG$_{bind}$	ΔE$_{ele}$	ΔE$_{vdW}$	ΔG$_{gas}$	ΔG$_{sol}$
H274	T1	-20.30 ± 4.34	-144.88 ± 10.13	-31.31 ± 2.52	-176.19 ± 10.21	155.90 ± 9.36
	T2	-16.48 ± 9.03	-137.06 ± 22.53	-28.17 ± 5.26	-165.22 ± 24.78	148.74 ± 17.50
	T3	-25.81 ± 4.11	-151.45 ± 8.54	-32.68 ± 3.16	-184.14 ± 8.37	158.32 ± 6.69
	T4	-30.52 ± 4.24	-167.85 ± 10.35	-30.61 ± 3.50	-198.46 ± 10.11	167.94 ± 8.49
	T5	-23.27 ± 4.21	-148.70 ± 12.83	-30.35 ± 2.82	-179.04 ± 12.23	155.78 ± 9.73
	T$_{avg}$	**-23.28 ± 5.19**	**-149.99 ± 12.88**	**-30.63 ± 3.45**	**-180.61 ± 13.14**	**157.34 ± 10.35**
I222K	T1	-28.43 ± 4.90	-171.09 ±13.93	-31.26 ± 3.42	-202.34 ± 13.73	173.91 ± 11.24
	T2	-20.18 ± 6.53	-144.76 ± 15.31	-25.08 ± 4.85	-169.84 ± 16.07	149.66 ± 12.94
	T3	-27.13 ±6.00	-156.67 ± 14.22	-31.64 ± 3.28	-188.31 ± 15.30	161.18 ± 11.58
	T4	-26.87 ± 6.57	-165.80 ± 19.73	-31.10 ± 3.41	-196.90 ± 20.91	170.04 ± 16.10
	T5	-26.01 ± 5.15	-153.12 ± 12.22	-28.76 ± 3.28	-181.88 ± 13.21	155.88 ± 10.14
	T$_{avg}$	**-25.72 ± 5.83**	**-158.29 ± 15.08**	**-29.57 ± 3.65**	**-187.85 ± 15.84**	**162.13 ± 12.40**
H274Y	T1	-26.22 ± 3.25	-131.63 ± 16.23	-32.37 ± 2.62	-164.00 ± 15.71	137.78 ± 14.03
	T2	-36.56 ± 4.76	-175.26 ± 15.34	-34.47 ± 3.93	-209.72 ± 14.01	173.16 ± 13.30
	T3	-27.71 ± 3.24	-179.67 ± 11.45	-31.25 ± 2.58	-210.91 ± 11.38	183.21 ± 10.88
	T4	-24.26 ± 5.61	-173.30 ± 16.41	-31.80 ± 2.60	-205.10 ± 16.83	180.84 ± 13.83
	T5	-22.52 ± 5.12	-169.94 ± 22.45	-29.96 ± 2.93	-199.91 ± 21.39	177.39 ± 19.32
	T$_{avg}$	**-27.46 ± 4.40**	**-165.96 ± 16.37**	**-31.97 ± 2.93**	**-197.93 ± 15.87**	**170.47 ± 14.27**

H274Y-I222K	T1	-20.35 ± 5.40	-139.80 ± 15.52	-27.49 ± 4.01	-167.29 ± 15.07	146.94 ± 12.68
	T2	-24.92 ± 4.13	-126.91 ± 17.15	-32.09 ± 2.96	-158.99 ± 15.94	134.08 ± 13.95
	T3	-25.68 ± 3.55	-112.27 ± 12.68	-33.69 ± 2.87	-145.96 ± 11.80	120.28 ± 10.14
	T4	-26.03 ± 3.60	-115.52 ± 15.64	-32.90 ± 2.48	-148.42 ± 15.24	122.39 ± 13.43
	T5	-20.62 ± 5.29	-132.49 ± 22.51	-29.35 ± 3.11	-161.85 ± 22.20	141.22 ± 19.16
	T$_{avg}$	**-23.52 ± 4.39**	**-125.40 ± 16.70**	**-31.10 ± 3.09**	**-156.50 ± 16.05**	**132.99 ± 13.88**

Hydrogen bond formation between amino acid residues

The dimensionality of a protein is a vector metric governed by the direction of interaction and number of hydrogen bonds. Comparative analysis of WT$_{H5N1}$ and H274Y$_{H5N1}$ verified a decrease in the overall number of hydrogen bonds in the mutant species (Figure 8). This phenomenon was attributed to a reduction of hydrophobic interactions within the protein, evident from the van der Waals energy contribution. The residues of neuraminidase interact to a greater extent with the surrounding solvent, hence a depletion in the number of internal protein hydrogen bonds. System I222K$_{H5N1}$ showed a similar trend to the H274Y$_{H5N1}$ system (Figure S21). The double mutation H274Y-I222K$_{H5N1}$ showed an increase in the number of hydrogen bonds (Figure S22), offering a more compact structure that is supported by the radius of gyration.

Mutation H274Y$_{H1N1}$ (Figure S23) showed a decline in the overall average of calculated hydrogen bonds as compared with the WT$_{H1N1}$ system. This finding supports the idea of an unfolding of the protein due to depletion of interactions facilitating α-helices or β-strands capable of compacting the enzyme. Similarly, system I222K$_{H1N1}$ (Figure S24), compared to the parent protein, WT$_{H1N1}$, has a recorded decrease in the number of internal hydrogen bonds. A comparison of systems H274Y$_{H1N1}$ and H274Y-I222K$_{H1N1}$ (Figure S25) revealed a near equivalent hydrogen bonding pattern. This could be due to torsional or steric stress of the macromolecule, such that thermodynamically no further unravelling of the protein can be accommodated before an irreversible deformation of the protein takes effect.
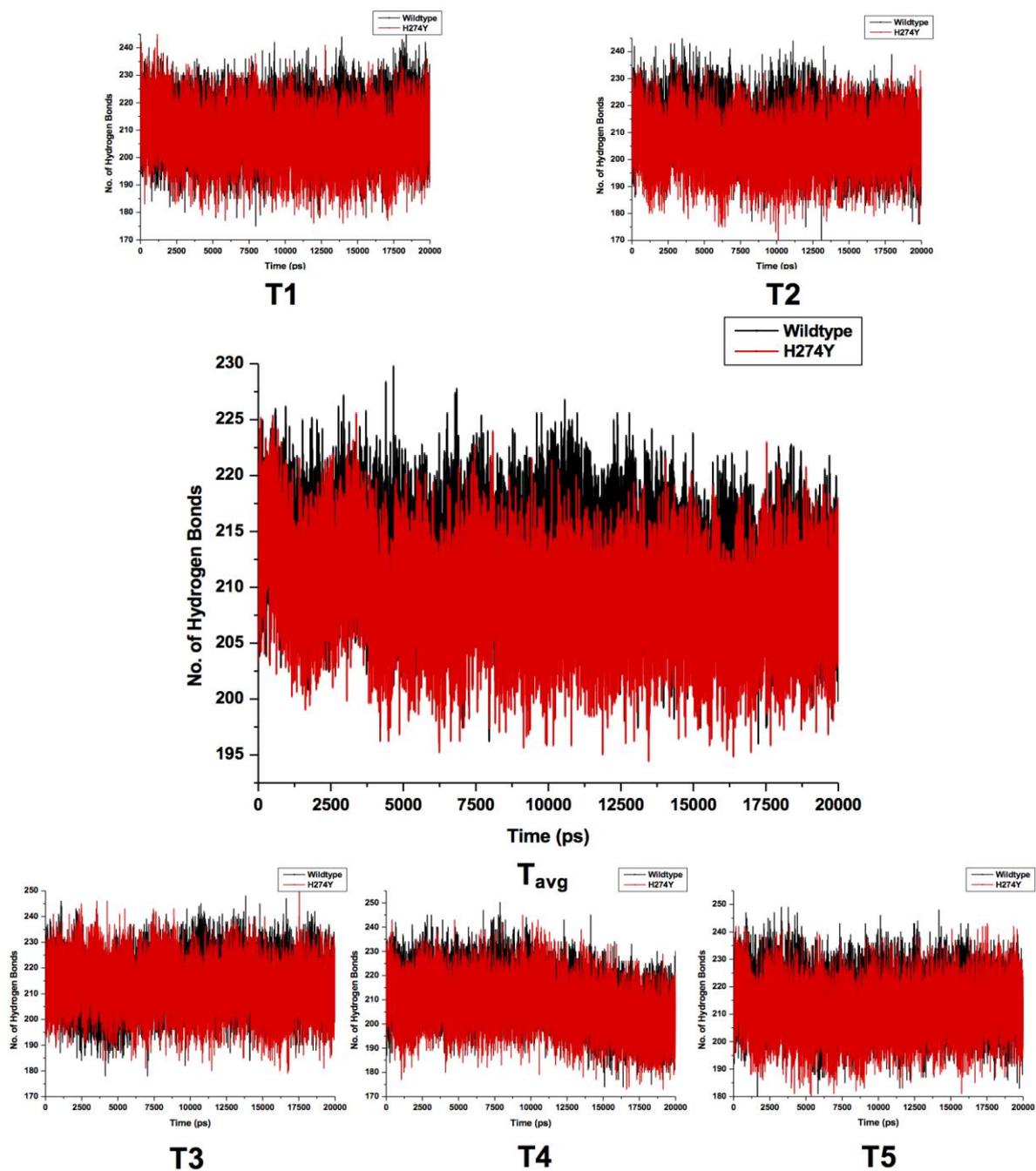
Figure 8: Averaged no. of hydrogen bonds in WT$_{H5N1}$ and H274Y$_{H5N1}$ across the 20 ns MD simulation: T1, T2, T3, T4, T5 and T$_{avg}$ presenting the 5 individual 20 ns MD trajectories and overall average, respectively.

Principal component analysis

PCA is used to assess the impact of mutations on the conformational dynamics of neuraminidase to solicit an impartial resistance mechanism.[72] Collation of systems WT_{H5N1} and $H274Y_{H5N1}$ (Figure 9) associated $H274Y_{H5N1}$ with restricted motion in relation to its α-carbon backbone. The displacements were reduced, minimizing the spatial occupancy of $H274Y_{H5N1}$ with a covariance of 0.3191. The magnitude of covariance for WT_{H5N1} was estimated to be -3.6367, advocating a more flexible framework with an inversely proportionate motion (Figure S26). System $H274Y_{H5N1}$ has a disproportionate motion of correlation ($R2$) value of 0.1814 in that the vectors commission dynamism in a single cooperative direction (Figure S27). The $R2$ value for WT_{H1N1} of -0.6526 indicated a more proportionate but antagonistic motion. Examination of system $I222K_{H5N1}$ (Figure S28) suggested that the neuraminidase enzyme tolerated an abridged flexibility of its α-carbon backbone in comparison with the WT enzyme. This corresponded with the determinations surmised from the RMSF values. The single mutation has a covariance of -3.1127, which is marginally smaller in magnitude than that for the WT, with an $R2$ of -0.3032. The direction of motion of $I222K_{H5N1}$ is antiparallel, the associations predominate by an antonymous relationship of large and small variables (Figure S29). The double mutation, $H274Y\text{-}I222K_{H5N1}$, has improved flexibility in the α-carbon backbone in comparison with $H274Y_{H5N1}$. This is demonstrated in the PCA scatter and covariance value estimated at 0.8148 (Figure S30). The improvement in flexibility coincided with the RMSD values, suggesting a greater number of conformational possibilities of the $H274Y\text{-}I222K_{H5N1}$ system. An improved $R2$ of 0.2337, in contrast with $H274Y_{H5N1}$, described a unified direction of motion of disproportionate magnitude (Figure S31).

The WT_{H1N1} enzyme (Figure 10) appeared to have reduced internal motion of its α-carbon backbone in comparison with $H274Y_{H1N1}$, a restriction which corresponded to the compactness observed from calculation of the radius of gyration. The covariance and $R2$ values were estimated to be -1.2249 and -0.4616, respectively, suggesting a disproportionate antithetic motion of the enzyme (Figure S32). A similar effect was observed with the double mutation system $H274Y\text{-}I222K_{H1N1}$, where the overall flexibility of the protein was rigid in comparison with $H274Y_{H1N1}$ (Figure S33). The covariance and $R2$ values of -0.4099 and -0.1539 support these findings (Figure S34). System $H274Y_{H1N1}$ demonstrated a more flexible α-carbon backbone which moved in a disorganized motion with a covariance of -0.7200 and an $R2$ value of -0.05218 (Figure S35). System $I222K_{H1N1}$ exhibited a continuous linear motion along its α-

carbon backbone, with a covariance of 0.6457 (Figure S36). This phenomenon was supported by an $R2$ of 0.3271 (Figure S37).
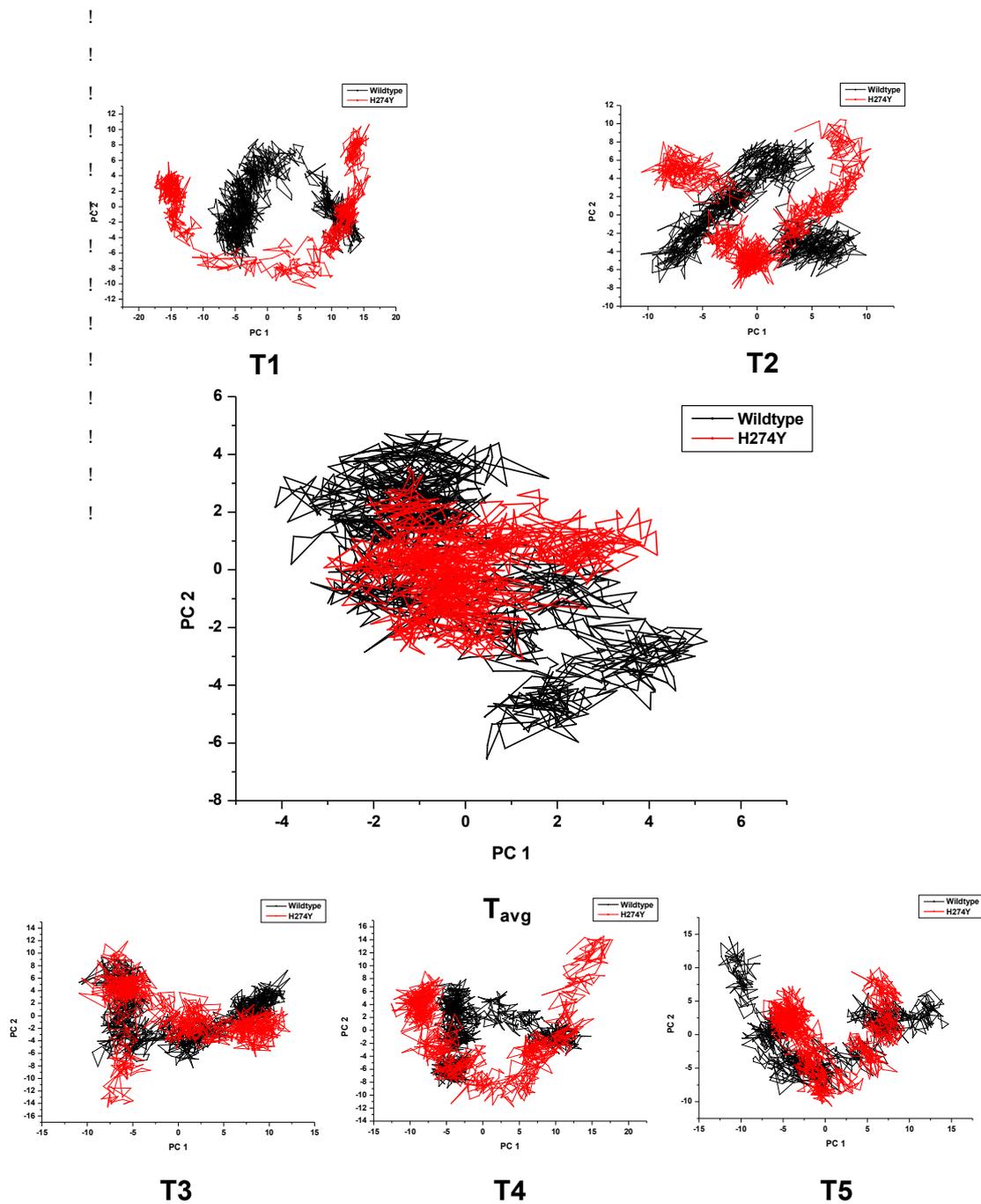

Figure 9: PCA scatter plots of 1000 frames of the distribution along two planes, PC1 and PC2 for; WT_{H5N1} and $H274Y_{H5N1}$ illustrating differences in eigenvectors of T1, T2, T3, T4, T5 and T_{avg} presenting the 5 individual 20 ns MD trajectories and overall average, respectively.

Figure 10: PCA scatter plots of 1000 frames of the distribution along two planes, PC1 and PC2 for; WT_{H1N1} and $H274Y_{H1N1}$ illustrating differences in eigenvectors of T1, T2, T3, T4, T5 and T_{avg} presenting the 5 individual 20 ns MD trajectories and overall average, respectively.
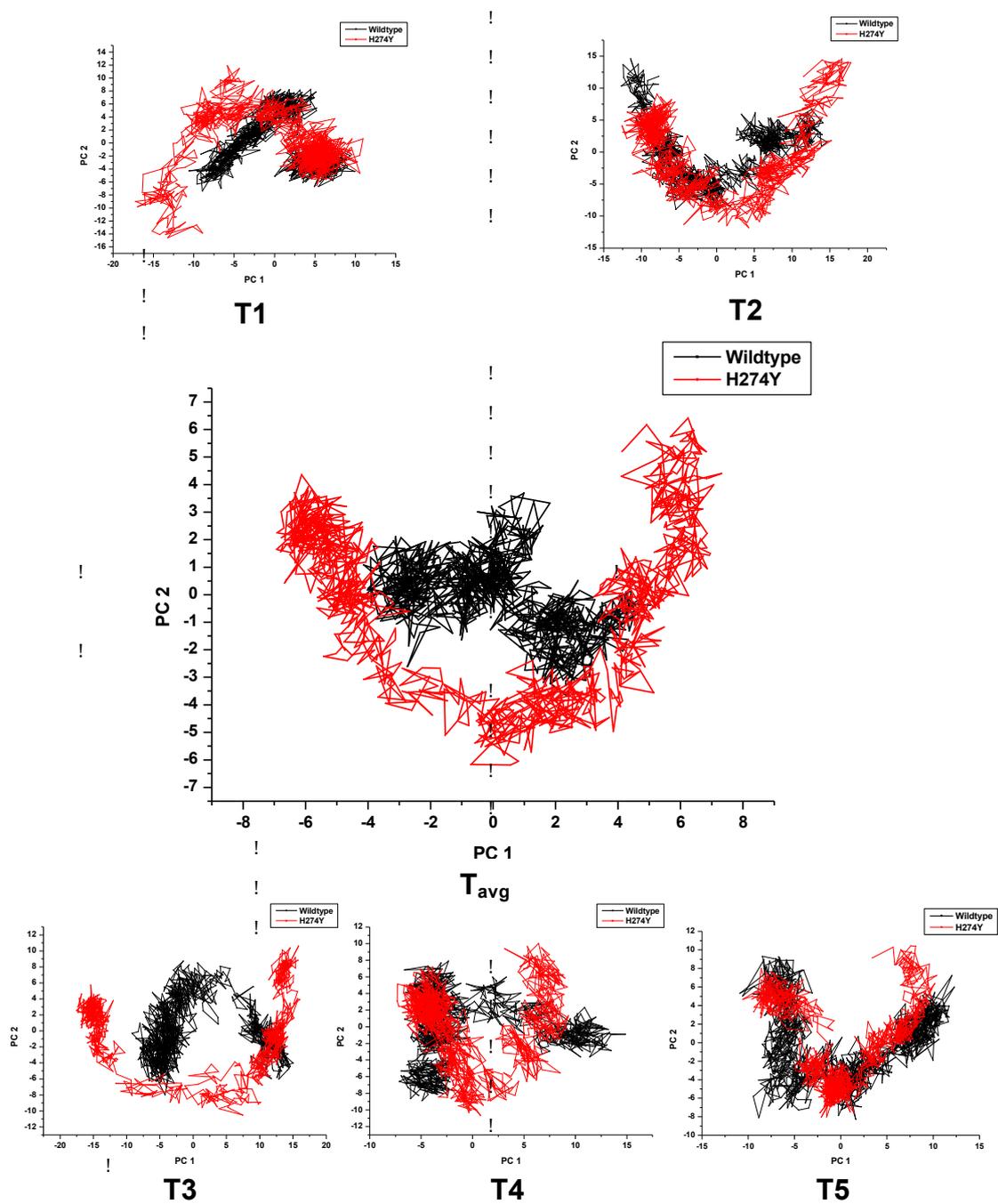
Solvent accessible surface area

The SASA imparts information relative to the compactness of the structure as well as the extent of hydrophobicity in the interior of the folded protein.[73] System $H274Y_{H5N1}$ appeared to expand over time (Figure S38), whereas WT_{H5N1} remained unchanged. Both systems reflected a distinct and relatively stable complexed conformation, which is affirmed by the radius of gyration estimations. System $I222K_{H5N1}$ (Figure S39) and $H274Y\text{-}I222K_{H5N1}$ (Figure S40) demonstrated a nominally lower surface area than the WT and $H274Y_{H5N1}$, respectively. This implied that contact between the van der Waals partitions and the solvent was diminished. A dehydron species may be present in systems $I222K_{H5N1}$ and $H274Y\text{-}I222K_{H5N1}$ due to the contributions of ΔE_{vdW} and ΔG_{sol}, resulting in a reduced interaction between the solvent and van der Waals region.[74] The preservation of the hydrophobic regions results in a loss in volume of the active site.

The WT_{H1N1} system has a reduced exposed surface area when compared with $H274Y_{H1N1}$ (Figure S41). This observation corresponded to the PCA of $H274Y_{H1N1}$, facilitating a structure with greater potential for interaction with solvent. System $I222K_{H1N1}$ exhibited a slightly different trend such that initially both the WT and single mutant shared similar SASA. Over time, at ~10,000 ps onward, a change in conformation occurred, with $I222K_{H1N1}$ appearing to become more accessible to the solvent (Figure S42). A similar effect was found between $H274Y_{H1N1}$ and $H274Y\text{-}I222K_{H1N1}$ (Figure S43). During the initial 10,000 ps of the simulation, there appeared to be a difference in the SASA of the two systems, with $H274Y_{H1N1}$ having greater solvent accessibility than the double mutant. In the later 10,000 ps, both systems begin to overlap. This anomaly suggests a conformational change in the presence of the double mutation H274Y and I222K, whereby these residues pointed outward, encouraging solvent interaction and disrupting the three-dimensional compactness.

Conclusion

Significant findings emerged in our endeavour to gain insight into the binding mode and origin of resistance of oseltamivir in H1N1 and H5N1. It was noted that the mechanism of resistance is entirely dependent on the type of mutation, relative positioning of such occurrences, and the repercussions thereof on the active site of the enzyme under analysis. Our results indicate that resistance of neuraminidase was inherited by the expression of H274Y and/or I222K mutations. The neuraminidase of H1N1 has known resistance against oseltamivir in the presence of mutations H274Y and/or I222K. Despite originating from the same host by reassortment,

H5N1 has proven to be more susceptible to oseltamivir than H1N1 in the presence of mutations. Comparatively, neuraminidase of H5N1 appears to experience greater structural fluctuations in acquiring resistance against oseltamivir as opposed to the H1N1 system. H5N1 gains resistance by a loss in volume of the active site due to compaction of the neuraminidase, while H1N1 develops resistance by solvent exposure of active site residues. Subsidiary analyses of RMSD, RMSF and potential energy confirm the progressive development of resistance from WT through to mutants (for both H1N1 and H5N1).

The free binding energy established from the MM/GBSA algorithm offers the first sign of evidence of resistance. Based on the electrostatic funnel mechanism of neuraminidase (for H1N1), an increase in binding energy was observed. This was an artefact of the active site residues and drug interacting with solvent, rather than with one another. The presence of the H274Y mutation has a compensatory effect to ensure the survival of the viral species by inhibiting enzyme interaction with the drug, without compromising the integrity of the enzyme. Mutation I222K, through distance, transposes a putative constriction in the volume of the active site of H5N1, whereas in H1N1 the protein unfolds, promoting solvent interaction. As a result, the active site inherently cannot support the bulky 1-ethylpropoxy hydrophobic moiety of oseltamivir. It is reasonable to suppose that the I222K mutation has a potent effect in discriminating between binding of the substrate and binding of the drug oseltamivir. The H274Y-I222K double mutant exhibits even greater drug resistance. Overall resistance in systems H5N1 and H1N1 occurs as a result of loss of hydrophobicity within the binding pocket of the active site. The lack of hydrophobicity leads to structural collapse of the available active site, which in turn diminishes the integrity of the binding landscape. This poses a threat, given that H5N1/avian influenza has been found to have an increasing incidence of human transmissibility. The existence of a resistant strain is having a severe negative impact on human health. Therefore, it is imperative to design and identify unique, innovative, and selective chemotherapeutic agents to adapt to the newly defined binding pocket.

Disclosure

The authors report no conflicts of interest in this work.

References

1. Le L, Lee EH, Hardy DJ, Truong TN, Schulten K. Molecular dynamics simulations suggest that electrostatic funnel directs binding of Tamiflu to influenza N1 neuraminidases. *PLoS Comput Biol*. 2010;6(9):e1000939.
2. World Health Organization. Pandemic preparedness. Available from: http://web.archive.org/web/20030202145905/http://www.who.int/csr/disease/influenza/pandemic/en. Accessed April 21, 2015.
3. Cohen E. When a pandemic isn't a pandemic. Available from: http://edition.cnn.com/2009/HEALTH/05/04/swine.flu.pandemic/index.html. Accessed April 22, 2015.
4. Doshi P. The elusive definition of pandemic influenza. *Bull World Health Organ*. 2011;89(7):532–538.
5. Patel RB, Mathur MB, Gould M, et al. Demographic and clinical predictors of mortality from highly pathogenic avian influenza A (H5N1) virus infection: CART analysis of international cases. *PLoS One*. 2014;9(3):e91630.
6. Smith GJ, Vijaykrishna D, Bahl J, et al. Origins and evolutionary genomics of the 2009 swine – origin H1N1 influenza A epidemic. *Nature*. 2009;459(7250):1122–1125.
7. Yarris L. New biological route for swine flu to human infections. Berkeley Lab, 2009. Available from: https://newscenter.lbl.gov/2009/12/09/swine-flu-to-human/. Accessed June 12, 2015.
8. Mostaço-Guidolin LC, Bowman CS, Greer AL, Fisman DN, Moghadas SM. Transmissibility of the 2009 H1N1 pandemic in remote and isolated Canadian communities: a modelling study. *BMJ Open*. 2012;2(5):e001614.
9. Sifferlin A. H1N1's death toll: 15 times higher than previously thought. Available from: http://healthland.time.com/2012/06/26/h1n1s-death-toll-15-times-higher-than-previously-thought/. Accessed April 23, 2015.
10. Davis CP. Swine flu (swine influenza A [H1N1 and H3N2v] virus. Available from: http://www.medicinenet.com/swine_flu/article.htm. Accessed April 23, 2015.

11. Baz M, Abed Y, Papenburg J, Bouhy X, Hamelin ME, Boivin G. Emergence of oseltamivir-resistant pandemic H1N1 virus during prophylaxis. *N Engl J Med.* 2009;361(23):2296–2297.

12. Centers for Disease Control and Prevention. Influenza A (H3N2) variant virus. Available from: http://www.cdc.gov/flu/swineflu/h3n2v-cases.htm. Accessed April 25, 2015.

13. Adabala PJ, LeGresley EB, Bance N, Niikura M, Pinto BM. Exploitation of the catalytic site and 150 cavity for design of influenza A neuraminidase inhibitors. *J Org Chem.* 2013;78(21):10867–10877.

14. Sylte M, Suarez D. Influenza neuraminidase as a vaccine antigen. In: Compans RW, Orenstein WA, editors. *Vaccines for Pandemic Influenza.* Volume 333. Berlin, Germany: Springer; 2009.

15. Russell RJ, Haire LF, Stevens DJ, et al. The structure of H5N1 avian influenza neuraminidase suggests new opportunities for drug design. *Nature.* 2006;443(7107):45–49.

16. Li Q, Qi J, Zhang W, et al. The 2009 pandemic H1N1 neuraminidase N1 lacks the 150-cavity in its active site. Nat Struct Mol Biol. 2010;17(10):1266–1268.

17. Woods CJ, Malaisree M, Long B, McIntosh-Smith S, Mulholland AJ. Analysis and assay of oseltamivir-resistant mutants of influenza neuraminidase via direct observation of drug unbinding and rebinding in simulation. *Biochemistry.* 2013;52(45):8150–8164.

18. Han N, Liu X, Mu Y. Exploring the mechanism of zanamivir resistance in a neuraminidase mutant: a molecular dynamics study. *PLoS One.* 2012;7(9):e44057.

19. Shobugawa Y, Saito R, Sato I, et al. Clinical effectiveness of neuraminidase inhibitors – oseltamivir, zanamivir, laninamivir, and peramivir – for treatment of influenza A(H3N2) and A(H1N1)pdm09 infection: an observational study in the 2010–2011 influenza season in Japan. *J Infect Chemother.* 2012;18(6):858–864.

20. Takashita E, Ejima M, Itoh R, et al. A community cluster of influenza A(H1N1)pdm09 virus exhibiting cross-resistance to oseltamivir and peramivir in Japan, November to December 2013. *Euro Surveill.* 2014;19(1):20666.

21. Kim CU, Lew W, Williams MA, et al. Influenza neuraminidase inhibitors possessing a novel carbocyclic sialic acid analogues with potent anti-influenza activity. *J Am Chem Soc.* 1997;119(4):681–690.

22. [No authors listed]. Global monitoring of antiviral resistance in currently circulating human influenza viruses, November 2011. *Wkly Epidemiol Rec.* 2011;86(45):497–501.

23. Meijer A, Rebelo-de-Andrade H, Correia V, et al. Global update on the susceptibility of human influenza viruses to neuraminidase inhibitors, 2012–2013. *Antiviral Res*. 2014;110:31–41.

24. Karthick V, Shanthi V, Rajasekaran R, Ramanathan K. Exploring the cause of oseltamivir resistance against mutant H274Y neuraminidase by molecular simulation approach. *Appl Biochem Biotechnol*. 2012;167(2):237–249.

25. Karplus M, McCammon JA. Molecular dynamics simulationa of biomolecules. *Nature*. 2002;9(9):646.

26. Li L, Li Y, Zhang L, Hou T. Theoretical studies on the susceptibility of oseltamivir against variants of 2009 A/H1N1 influenza neuraminidase. *J Chem Inf Model*. 2012;52(10):2715–2729.

27. Karthick V, Ramanathan K. Insight into the oseltamivir resistance R292K mutation in H5N1 influenza virus: a molecular docking and molecular dynamics approach. *Cell Biochem Biophys*. 2014;68(2):291–299.

28. Operario DJ, Moser MJ, St George K. Highly sensitive and quantitative detection of the H274Y oseltamivir resistance mutation in seasonal A/H1N1 influenza virus. *J Clin Microbiol*. 2010;48(10):3517–3524.

29. Malaisree M, Rungrotmongkol T, Nunthaboot N, et al. Source of oseltamivir resistance in avian influenza H5N1 virus with the H274Y mutation. *Amino Acids*. 2009;37(4):725–732.

30. Rungrotmongkol T, Intharathep P, Malaisree M, et al. Susceptibility of antiviral drugs against 2009 influenza A (H1N1) virus. *Biochem Biophys Res Commun*. 2009;385(3):390–394.

31. Wang NX, Zheng JJ. Computational studies of H5N1 influenza virus resistance to oseltamivir. *Protein Sci*. 2009;18(4):707–715.

32. Yang Z, Yang G, Zhou L. Mutation effects of neuraminidases and their docking with ligands: a molecular dynamics and free energy calculation study. *J Comput Aided Mol Des*. 2013;27(11):935–950. 33. Pan D, Sun H, Bai C, et al. Prediction of zanamivir efficiency over the possible 2009 influenza A (H1N1) mutants by multiple molecular dynamics simulations and free energy calculations. *J Mol Model*. 2011;17(10):2465–2473.

34. Takano R, Kiso M, Igarashi M, et al. Molecular mechanisms underlying oseltamivir resistance mediated by an I117V substitution in the neuraminidase of subtype H5N1 avian influenza A viruses. *J Infect Dis*. 2013;207(1):89–97.

35. Bloom JD, Nayak JS, Baltimore D. A computational-experimental approach identifies mutations that enhance surface expression of an oseltamivir-resistant influenza neuraminidase. *PLoS One*. 2011;6(7):e22201.

36. McKimm-Breschkin JL, Rootes C, Mohr PG, Barrett S, Streltsov VA. In vitro passaging of a pandemic H1N1/09 virus selects for viruses with neuraminidase mutations conferring high-level resistance to oseltamivir and peramivir, but not to zanamivir. *J Antimicrob Chemother*. 2012;67(8):1874–1883.

37. Nguyen HT, Fry AM, Gubareva LV. Neuraminidase inhibitor resistance in influenza viruses and laboratory testing methods. *Antivir Ther*. 2012;17(1 Pt B):159–173.

38. Huang L, Cao Y, Zhou J, et al. A conformational restriction in the influenza A virus neuraminidase binding site by R152 results in a combinational effect of I222T and H274Y on oseltamivir resistance. *Antimicrob Agents Chemother*. 2014;58(3):1639–1645.

39. Hayden FG, de Jong MD. Emerging influenza antiviral resistance threats. *J Infect Dis*. 2011;203(1):6–10.

40. Wagh K, Bhatia A, Greenbaum BD, Bhanot G. Bird to human transmission biases and vaccine escape mutants in H5N1 infections. *PLoS One*. 2014;9(7):e100754.

41. Mirjalili V, Noyes K, Feig M. Physics-based protein structure refinement through multiple molecular dynamics trajectories and structure averaging. *Proteins*. 2014;82 Suppl 2:196–207.

42. Altis A, Nguyen PH, Hegger R, Stock G. Dihedral angle principal component analysis of molecular dynamics simulations. *J Chem Phys*. 2007;126(24):244111.

43. Shlens J. *A Tutorial on Principal Component Analysis*. San Diego, CA, USA: University of California; 2005.

44. van der Vries E, Collins PJ, Vachieri SG, et al. H1N1 2009 pandemic influenza virus: resistance of the I223R neuraminidase mutant explained by kinetic and structural analysis. *PLoS Pathog*. 2012;8(9):e1002914.

45. Collins PJ, Haire LF, Lin YP, et al. Crystal structures of oseltamivir-resistant influenza virus neuraminidase mutants. *Nature*. 2008;453(7199):1258–1261.

46. *UCSF Chimera – an Extensible Molecular Modeling System*. Resource for Biocomputing, Visualization, and Informatics, University of California, San Francisco; National Institute of General Medical Sciences; National Institutes of Health. Available from: http://www.cgl.ucsf.edu/chimera/. Accessed January 11, 2015.

47. Kanibolotsky DS, Novosyl'na OV, Abbott CM, Negrutskii BS, El'skaya AV. Multiple molecular dynamics simulation of the isoforms of human translation elongation factor 1A

reveals reversible fluctuations between "open" and "closed" conformations and suggests specific for eEF1A1 affinity for Ca2+-calmodulin. *BMC Struct Biol*. 2008;8:4.

48. Götz AW, Williamson MJ, Xu D, Poole D, Le Grand S, Walker RC. Routine microsecond molecular dynamics simulations with AMBER on GPUs. 1. Generalized Born. *J Chem Theory Comput*. 2012;8(5):1542–1555.

49. Case DA, Berryman JT, Betz RM, et al. AMBER 12, University of California, 2012. Available from: http://ambermd.org/doc12/Amber12.pdf. Accessed June 11, 2015.

50. Salomon-Ferrer R, Götz AW, Poole D, Le Grand S, Walker RC. Routine microsecond molecular dynamics simulations with AMBER on GPUs. 2. Explicit solvent particle mesh Ewald. *J Chem Theory Comput*. 2013;9(9):3878–3888.

51. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. *J Chem Phys*. 1983;79(2):926.

52. Essmann U, Perera L, Berkowitz M, Darden T, Lee H, Pedersen L. A smooth particle mesh Ewald method. *J Chem Phys*. 1995;103:8577.

53. Berendsen HJC, Postma JPM, Van Gunsteren WF, DiNola A, Haak JR. Molecular dynamics with coupling to an external bath. *J Chem Phys*. 1984;81:3684–3690.

54. Ryckaert JPC, Berendsen HJC. Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J Comput Phys*. 1977;23:327–341.

55. Le Grand S, Gotz AW, Walker RC. SPFP: speed without compromise – a mixed precision model for GPU accelerated molecular dynamics simulations. *Comput Phys Commun*. 2013;184:374–380.

56. Sun H, Li Y, Tian S, Xu L, Hou T. Assessing the performance of MM/PBSA and MM/GBSA methods. 4. Accuracies of MM/PBSA and MM/GBSA methodologies evaluated by various simulation protocols using PDBbind data set. *Phys Chem Chem Phys*. 2014;16(31):16719–16729.

57. Greenidge PA, Kramer C, Mozziconacci JC, Wolf RM. MM/GBSA binding energy prediction on the PDBbind data set: successes, failures, and directions for further improvement. *J Chem Inf Model*. 2013;53(1):201–209.

58. Godschalk F, Genheden S, Söderhjelm P, Ryde U. Comparison of MM/GBSA calculations based on explicit and implicit solvent simulations. *Phys Chem Chem Phys*. 2013;15(20):7731–7739.

59. Mitomo D, Fukunishi Y, Higo J, Nakamura H. Calculation of protein-ligand binding free energy using smooth reaction path generation (SRPG) method: a comparison of the explicit

water model, GB/SA model and docking score function. *Genome Inform*. 2009;23(1):85–97.

60. Hou T, Wang J, Li Y, Wang W. Assessing the performance of the MM/PBSA and MM/GBSA methods. 1. The accuracy of binding free energy calculations based on molecular dynamics simulations. *J Chem Inf Model*. 2011;51(1):69–82.

61. Cocco S, Monasson R, Weigt M. From principal component to direct coupling analysis of coevolution in proteins: low-eigenvalue modes are needed for structure prediction. *PLoS Comput Biol*. 2013;9(8):e1003176.

62. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. *J Mol Graph*. 1996;14(1):33–38.

63. Bakan A, Meireles LM, Bahar I. ProDy: protein dynamics inferred from theory and experiments. *Bioinformatics*. 2011;27(11):1575–1577.

64. Hess B. Convergence of sampling in protein simulations. *Phys Rev E Stat Nonlin Soft Matter Phys*. 2002;65(3 Pt 1):031910.

65. Wassenaar T. A molecular dynamics hands-on session II. Available from: http://www.nmr.chem.uu.nl/~tsjerk/course/md-tutorial/analysis.html. Accessed October 31, 2014.

66. [No authors listed]. From sequence to structure. New Science Press Ltd; 2004. Available from: http://www3.uah.es/farmamol/New_Science_Press/nsp-protein-1.pdf. Accessed June 11, 2015.

67. Lobanov MI, Bogatyreva NS, Galzitskaia OV. [Radius of gyration as an indicator of protein structure compactness]. *Mol Biol (Mosk)*. 2008;42(4):623–628. Russian.

68. Caves LS, Evanseck JD, Karplus M. Locally accessible conformations of proteins: multiple molecular dynamics simulations of crambin. *Protein Sci*. 1998;7(3):649–666.

69. Caravella JA, Carbeck JD, Duffy DC, Whitesides GM, Tidor B. Long-range electrostatic contributions to protein-ligand binding estimated using protein charge ladders, affinity capillary electrophoresis, and continuum electrostatic theory. *J Am Chem Soc*. 1999;121(18):4340–4347.

70. Cottrell D, Tupper PF. Energy drift in molecular dynamics simulations. *BIT Numerical Mathematics*. 2007;47(3):507–523.

71. Rungrotmongkol T, Intharathep P, Malaisree M, et al. Susceptibility of antiviral drugs against 2009 influenza A (H1N1) virus. *Biochem Biophys Res Commun*. 2009;385(3):390–394.

72. Jolliffe IT. *Principal Component Analysis*. New York, NY, USA: Springer; 2002.

73. Richmond TJ. Solvent accessible surface area and excluded volume in proteins: analytical equations for overlapping spheres and implications for the hydrophobic effect. *J Mol Biol*. 1984;178(1):63–89.

74. Chaplin M. Protein folding and denaturation. Available from: http://www1.lsbu.ac.uk/water/protein_denatured.html. Accessed October 31, 2014.

Chapter 6

Per-residue Free Binding Energy Profiled Pharmacophore Modelling as an Enhanced Approach in Drug Discovery: A Case Study for *in silico* Screening and Validation of Potential Influenza A Neuraminidase Inhibitors

Ashona Singh[a], Sarentha Chetty[a], Mahmoud E.   Soliman[a*]

[a] School of Health Sciences, University of KwaZulu-Natal, Westville, Durban 4000, South Africa

[*] Corresponding author: Mahmoud E.   Soliman, email: soliman@ukzn.ac.za

Telephone: +27 31 260 7413, Fax: +27 31 260 7413

Abstract

The rapid development of resistance to existing anti-influenza agents has prompted the urgent need for the discovery of novel potential neuraminidase inhibitors.   The use of computer-aided drug design has changed remarkably the perspective on the discovery and development of new drug candidates.   In this report, we propose an improved pharmacophore modelling based on the binding free energy contributions from the amino acid residues that are critically involved in drug binding rather than arbitrary set of amino acid residues as widely adopted in literature.   The robustness of this approach lies in the fact that the calculated Molecular Mechanics/ Generalised Born Surface Area (MM-GBSA) per-residue binding free energies used to map the pharmacophoric features were obtained from molecular dynamics ensembles as opposed to docking scores as commonly adopted.   Prototype Influenza A Neuraminidase Inhibitors, oseltamivir, zanamivir, laninamivir and peramivir were used to validate our approach.   A wide range of simulation and analytical protocols such as, but not limited to,

ligand-based and structure-based pharmacophores, molecular dynamics and per-residue free energy footprints were also reported herein to substantiate the proposed approach.

We believe that the enhanced pharmacophore-modelling approach presented in this report would serve as a useful tool in drug discovery workflows.

Keywords

Virtual screening, neuraminidase, free binding energy, molecular dynamics, influenza virus A, ligand-based pharmacophore, structure-based pharmacophore

1. Introduction

The natural evolution of the influenza virus presents a continuous challenge to the human immune system. Evolution of the virus occurs primarily via antigenic variations through two distinct mechanisms: 1) antigenic shifts, and 2) antigenic drifts. Antigenic drifts are the result of point mutations that occur as the virus replicates and is often the cause of seasonal flu. Antigenic shift occurs when two different strains of the influenza virus combine to create a novel influenza subtype, which can lead to epidemics and pandemics [1-3].

Genetic mutations, together with viral compensatory evasion mechanisms in response to current therapeutic agents, have led to the rapid build-up of resistance. Influenza A is a prime example of natural selection through an adaptive survival mechanism. It has been documented throughout history as contributing to high fatality rates. The most significant incidence of influenza infection in the past century include the '*Spanish flu*' pandemic which extended from 1918 – 1919 with an estimated mortality of 50 million worldwide [4]. In more recent times, the 2009 pandemic strain of H1N1 claimed the lives of approximately 284,500 people [5]. To date, there have been accounts of outbreaks of the newer influenza strains including swine (H1N1)

107

flu and the highly-pathogenic avian (H5N1) flu [6-8].   These strains have already shown

resistance to current antivirals; oseltamivir (Tamiflu®) [9], zanamivir (Relenza®) [10], peramivir

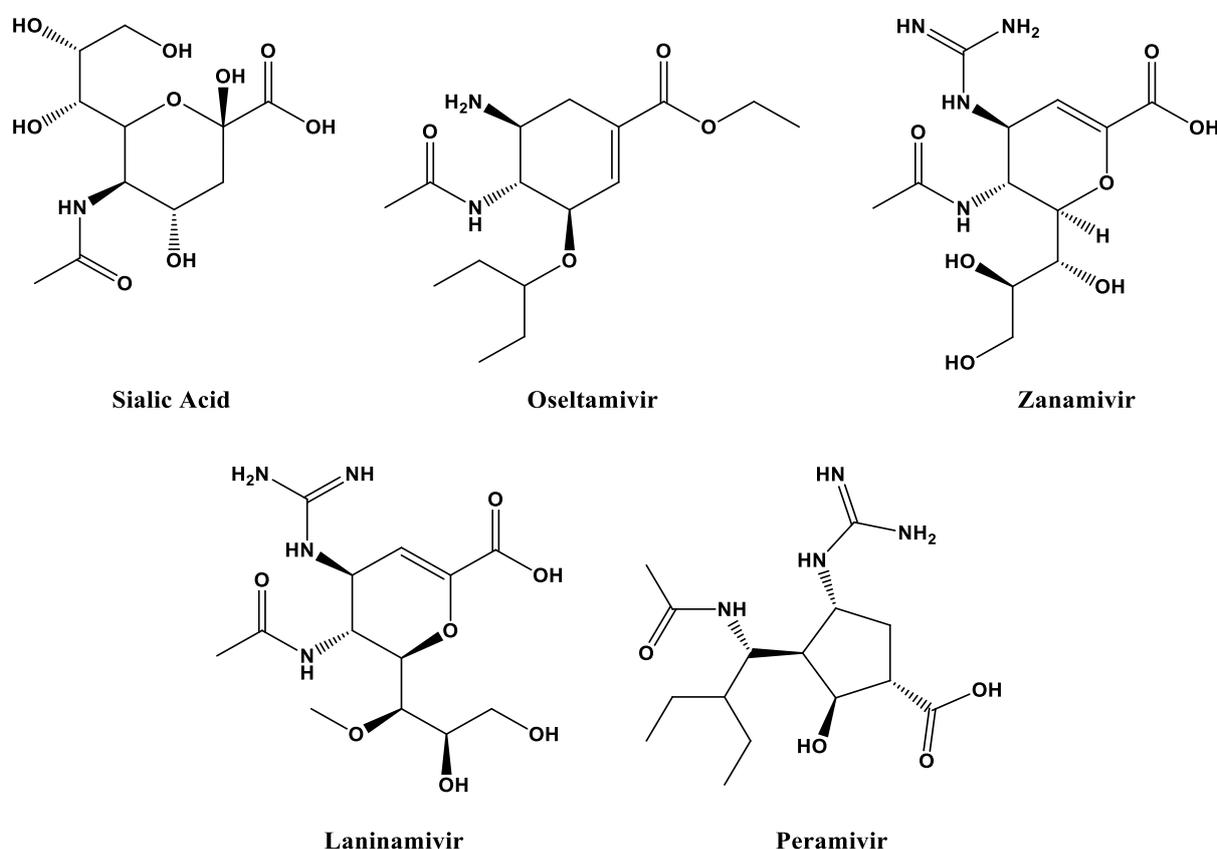(Rapivab®) [11] and laninamivir (Inavir®) [12].   These drugs work by neuraminidase inhibition.

Sialic Acid   **Oseltamivir**   **Zanamivir**

Laninamivir   **Peramivir**

Figure 1: 2-D Chemical structure of sialic acid (substrate) and active inhibitors (oseltamivir, zanamivir, laninamivir and peramivir)

Neuraminidase is one of three surface antigens which has become the primary target for

potential pharmacological interventions [13-16].   It is an exosialidase enzyme responsible for the

cleavage of the α-ketosidic bond between sialic acid and its adjacent sugar residue from the

viral envelope.   Cleavage occurs during the final stage of infection resulting in the release of

virus progeny from the host cell [17-18].   It has been suggested that in the early stages of infection,

neuraminidase may catalyse the removal of "decoy" receptors on mucins, cilia and cellular

glycocalyx present on human airway epithelium thus promoting access to target cells [19-20].

There are nine neuraminidase subtypes (N1-N9).   The group 1 neuraminidase enzymes *i.e.*

N1, N4 and N8; are phylogenic subtypes which share a conserved active site [21-24]. The appearance of resistance to current neuraminidase inhibitors has emphasized the urgency in the discovery, design and development process of alternate chemotherapeutic agents to prevent future pandemics [25-30].
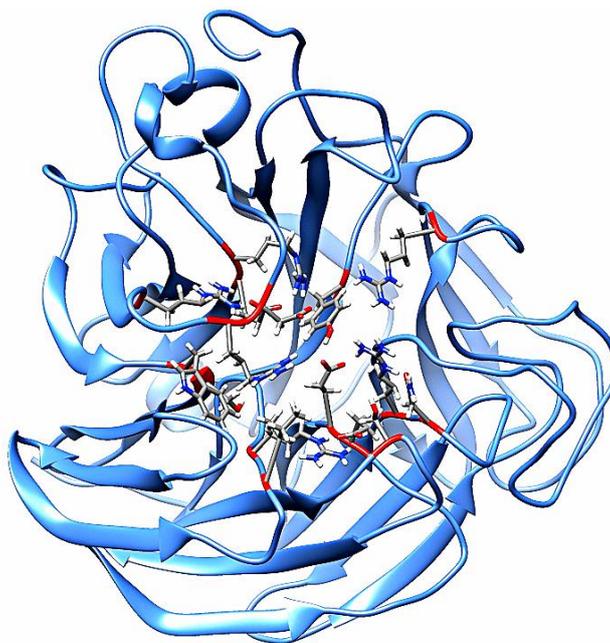

Figure 2: 3-D Structure of unbound H1N1 neuraminidase enzyme (PDB ID: 3TI6)

Classical methods of drug development against specified targets estimate a turnaround time (*i.e.* time from the initiation of work to final registration) of 12 – 15 years [31]. The use of computational methodology such as virtual screening (VS) can vastly improve efficiency. It is also more cost-effective by being less resource and labour intensive [32]. Pharmacophores can be used as a filtering tool in virtual screening. The pharmacophore model is generated from a three dimensional (3-D) scaffold, mapping critical chemical and functional moieties, such as; hydrogen bond acceptors (HBA), hydrogen bond donors (HBD), negatively and positively ionisable species, metal chelating groups and hydrophobic areas [33-34]. There are two main classes of pharmacophore VS models 1) Ligand-based, and 2) Structure-based [33].

Ligand-based pharmacophore modelling is based on known actives. The bioactive conformation allows for library generation of pharmacophore models through similarity searches. A number of 3-D inhibitors are aligned to isolate common pharmacophoric descriptors in order to build a pharmacophore model [32]. Structure-based pharmacophore modelling exploits a high resolution 3-D structure of a drug-protein complex for the assimilation and construction of common interaction patterns of a number of chemically diverse and similar functional moieties. Structural/feature mapping can be used directly to conceptualise a pharmacophore ensemble [35]. A disadvantage of ligand-based pharmacophore modelling is that the molecules discovered are strictly based on information derived from existing ligands (*i.e.* templates), which result in limited scaffold diversity in comparison to structure-based pharmacophore modelling [32-33, 36-37].

In this study we wish to propose an additional step to ligand-based pharmacophore VS by using information derived from molecular dynamic simulations of actives complexed with their respective receptors, to build a conformationally-enhanced pharmacophore model. The design strategy centres on obtaining a binding free energy profile of high contributing active site residues and their corresponding interacting ligand chemical features. Our proof of concept implements molecular dynamics (MD), where a thermodynamically favourable ensemble (*i.e.* energetically optimised drug-protein complex) is used to map pharmacophoric regions. This method proves highly advantageous as the MD ensemble closely mimics a biological environment. MD is a more dependable tool compared to receptor-ligand docking as it utilizes improved force fields and scoring functions [38]. The chemo-structural bias of the pharmacophore is reduced as the selection criteria of chemical features is governed by common interaction patterns of bioactive conformers with active site residues exhibiting a high free binding energy contribution. The free binding energy takes into account more chemical details in structural design, such as: local molecular interactions and environmental effects (e.g.

Protonation states).   Our ligand-based pharmacophore selection is three-pronged: 1) biological activity, 2) structural diversity/similarity and 3) energetically favourable interactions.   There have been previous energy based ranking studies to determine the significance of lead compounds using the ligand- and structure-based models.   However, the modus operandi utilized docking energies [39-43].   Unfortunately, the use of docking energies incurs artefacts, which could reflect inflated ligand-protein interactions.   MD would therefore give a more accurate depiction.

The energy mapping of chemical features for the pharmacophore corresponding to critical ligand-protein interactions can be used to limit false positives.   Our methodology proposes reduced randomisation by the generation of a more focussed library.   Coupled with the validation of MM-GBSA results against experimentally active compounds, this technique can prove a useful tool in the discovery of potential hits/leads of innovative and unique neuraminidase inhibitors.

2.   Methods

The flow chart in **Figure 3**, offers a basic overview of the critical steps in this study.   Each process is discussed in detail in the subsections below (2.1 to 2.6) and Results section (3.0).
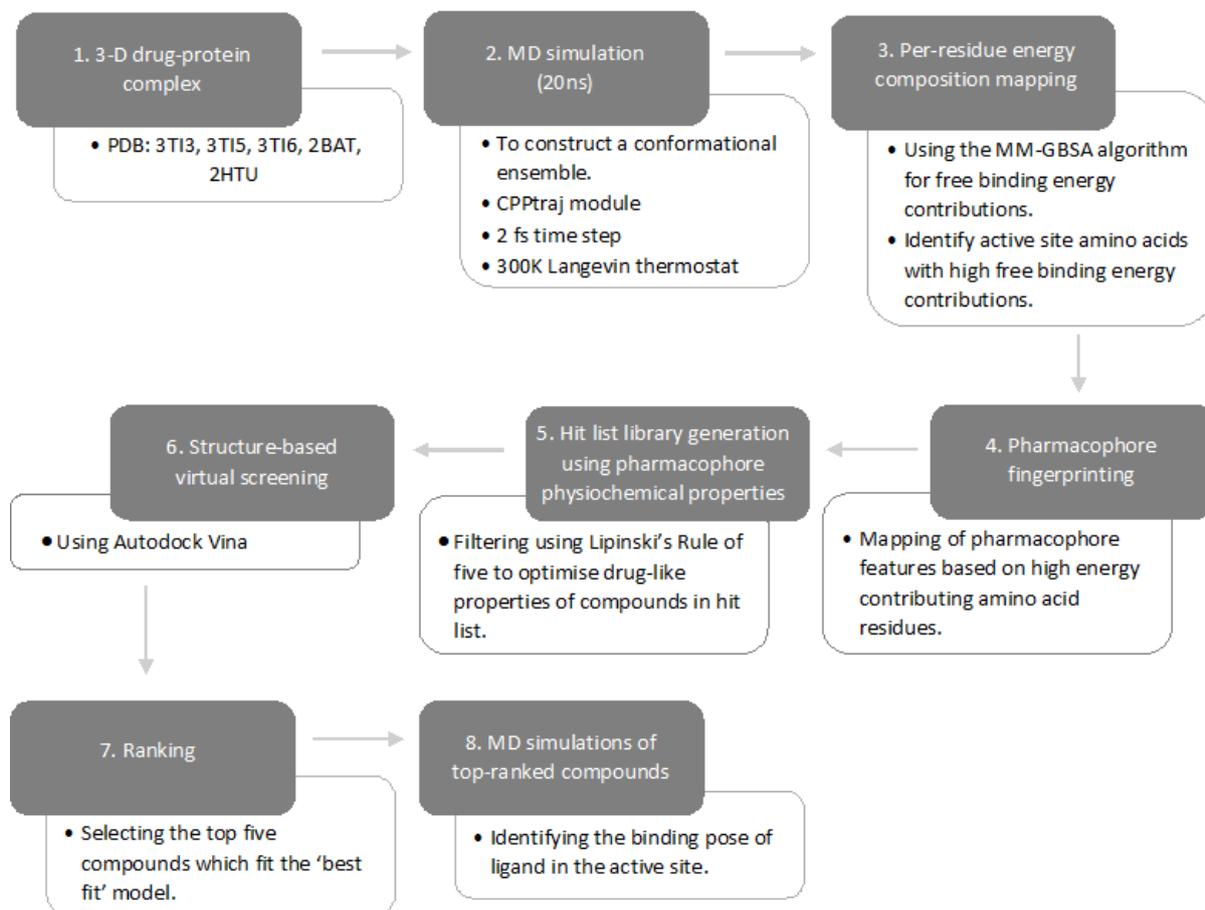
Figure 3: Pipeline of processes involves in the *in silico* screening of the generated MD ensemble pharmacophores

2.1 System preparation

Crystallographic structures of H1N1 neuraminidase bound with sialic acid (PDB ID: 2BAT) [44], oseltamivir (PDB ID: 3TI6) [45], zanamivir (PDB ID: 3TI5) [45], laninamivir (PDB ID: 3TI3) [45] and peramivir (PDB ID: 2HTU) [22], were obtained from the RCSB Protein Data Bank (PDB). Using the Chimera 1.8.1 software package [46-47] a monomer was selected followed by the correction of protonation states of amino acid residues as well as the removal of water, heavy metal atoms, polar hydrogens and non-bonding species.

2.2 MD simulation

Amber 14.0 GPU version of the PMEMD engine was used to perform a 20 ns molecular dynamics simulation. The protein was described with the FF99SB variant force field. The LEaP module embedded within Amber was use for the addition of hydrogen atoms and neutralisation of the protein by the addition of Na^+ counter ions. The system was suspended within a TIP3P [48484848] water box such that all protein atoms were within 8 Å from the box edge. The drug-protein complex was energetically minimised in three stages; 1) 1000 step restraint gradient minimisation with selective boundary conditions of the particle-mesh Ewald method which has a parameter of direct space and a van der Waals cut-off of 12 Å. Stage 2, a 2500 of deepest descent step with a solute harmonic potential of 500 kcal/mol. $Å^2$; and stage 3, a 1000 step unrestrained conjugate gradient energy minimisation of the complete system. The system was heated using a canonical ensemble (NVT) from 0 K to 300 K over 50 ps. The collision frequency of solutes was 1.0 ps^{-1} with a harmonic potential of 10 kcal/mol.$Å^2$. A final equilibration step of 500 ps was performed using the SPFP precision model of the SHAKE algorithm to constrict hydrogen bonds at 2 fs intervals. The simulations were conducted in a randomised seeding of an isobaric-isothermal ensemble (NPT). A constant pressure of 1 bar was maintained by the Berendsen barostat with a pressure coupling constant of 2 ps. A temperature of 300 K was set using a Langevin thermostat with collision frequency of 1.0 ps^{-2}. Co-ordinates were saved every 1 ps and the trajectories were analysed every 1 ps using a CPPTRAJ module in Amber 14.0.

2.3 Pharmacophore model

The construction of the models was a multi-step process and encompassed the principles of structure-based pharmacophore modelling. Pharmacophore generation is reliant primarily on the quality of 3-D molecules used in the training set. The choice of complexes was based on good crystallographic and activity data.

2.3.1 Per-residue free binding energy profiling

$$\Delta G_b = \Delta E_{MM} + \Delta G_{solv} + \Delta G_{SA}$$

The study emphasizes the specific interaction pattern of the ligand complexed with protein, highlighting constructive per-residue energy contributions. The per-residue free binding free energy is a post-dynamic calculation estimated from the Molecular Mechanics, Generalised Born, Surface Area (MM-GBSA) algorithm embedded within the Amber package. MM-GBSA method is capable of accurately ranking inhibitors based on binding free energies. The term ΔE_{MM} is the difference in energy between the protein-ligand complex and the unbound receptor and free ligand. ΔG_{solv} is the difference in the solvation energies. ΔG_{SA} is the estimated difference in the surface area energies of the unliganded receptor and free ligand. The critical binding site amino acid residues corresponding to defined chemical moieties of the ligand are profiled in **Figure 4** and the Supplementary Information.
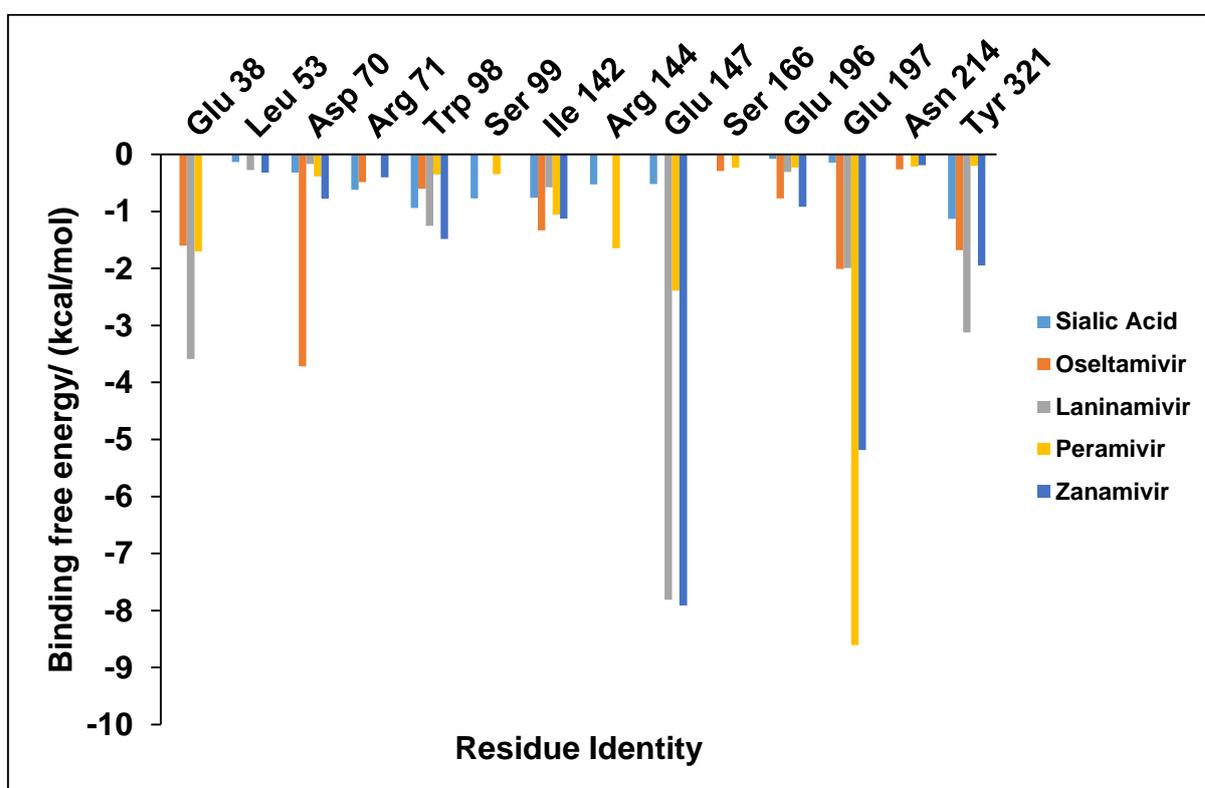

Figure 4: High contributing per-residue binding free energy interaction pattern of sialic acid, oseltamivir, laninamivir, peramivir and zanamivir complexed with neuraminidase

2.3.2 Ligand-based pharmacophore

The Espresso algorithm of LigandScout 3.12 software was used to generate the pharmacophore from the 3-D training set containing oseltamivir, laninamivir, sialic acid, peramivir and zanamivir. The model was predicted from the alignment of the five ligands and the overlap of similar chemical features. A three-feature limit was used for the creation of the model.

2.3.3 Structure-based pharmacophore

LigandScout is capable of interpreting critical interactions between the binding site residues and ligand. The predicted model was generated using the automated LigandScout pipeline which considers the 3-D macromolecular structure of the complexed ligand. The pharmacophoric ensemble is a culmination of the four active inhibitor complexes: oseltamivir (PDB ID: 3TI6), laninamivir (PDB ID: 3TI3), peramivir (PDB ID: 2HTU) and zanamivir (PDB ID: 3TI5). The minimum feature requirement was set at three. The model was a measure of dihedral angles and complementary chemical features with specific distance interaction filters as well as exclusion volumes.

2.4 Pharmacophore & Docking selectivity and evaluation

2.4.1 Pharmacophores

The pharmacophore models obtained from the ligand- and structure-based techniques were evaluated for the ability to detect actives from inactives (decoys). A decoy test set of 698 molecules was extracted from the electronic directory of useful decoys (E-DUD) [49]. The molecules exhibit similar physical properties but dissimilar chemical topology to the inhibitors. The active set contained 17 compounds, ranging from high (0.1 nM) to moderate activity (20 nM). (See Supplementary Information for a list of actives, **Table S1**).

2.4.2 Docking system

AutoDock Vina version was used to carry out docking. Validation of the docking software was conducted by re-docking of ligands back into their respective binding pockets and cross-docking of ligands peramivir, laninamivir, sialic acid and zanamivir into the receptor of the oseltamivir (PDB ID: 3TI6). The crystal structure of PDB ID: 3TI6 (resolution: 1.69 Å) was selected for the virtual screening as it had the best overall crystallography data. The docking software was further assessed for its ability to pick out actives from inactives using decoy molecules.

2.5 *In silico* virtual screening

Two *in silico* screening methodologies were employed, 1) pharmacophore screening and 2) molecular docking. LigandScout version 3.12 was used for the pharmacophore screen. The predicted ligand- and structure-based pharmacophores models were screened against a library of 5189 drug-like compounds obtained from the ZINC database. A hit list of 23 compounds and five compounds was retrieved from the ligand-based and structure-based pharmacophore models, respectively. A combined hit list of these 28 potential inhibitors was further docked into the receptor of PDB ID: 3TI6 using Autodock Vina employing an enhanced loop docking protocol. Autodock uses a Lamarckian algorithm to predict docked conformations. Prior to docking, MGL tools were used to define atom types and assign Gasteiger partial charges. The files were then converted to a pdbqt format using the Raccoon version software. A gridbox of dimensions of 34 x 24 x 36 (nm^3) was used. The docking calculations were performed in triplicate.

3. Results and Discussion

The use of CADD strategies proved to be an invaluable tool in the landmark discoveries of antiviral neuraminidase inhibitors (zanamivir, oseltamivir and peramivir). By the construction

of a 3-D model using per-residue binding free energies, our work serves to provide an enriched perspective on pharmacophore generation.

Alignment of receptors N1 and N8, confirm a conserved active site with similarity of sequence and structure of 0.54 and 56%, respectively.   This allowed the inclusion of peramivir complexed with an N8 receptor, in the construction of the pharmacophores.   The catalytic cleavage of sialic acid occurs at pH 4.   The system was therefore optimised to mimic the slightly acidic biological environment by protonation of the arginine residues.

Neuraminidase's mechanism of action occurs in four distinctive steps.   The first step referred to as the 'binding event' is solvent-mediated.   A charge-charge interaction (*i.e.* salt bridge) between the pseudo-equatorial carboxylate moiety and Arg 37, 212, 287 ensues.   The conformation creates a steric constraint which is accommodated by residue, Tyr 321.   The next step is the catalytic reaction, in which hydrogen bond donation from the cationic ligand to the solvent and enzymatic residues is facilitated.   A network of interactions with residues Asp 70, Arg 71 and Glu 196 and Glu197 stabilizes the cationic intermediate [4].

The active site of N1 neuraminidase can be divided into four critical binding pockets, **Figure 5** [50].   Pocket C1 (**Figure 5**, green residues) encompasses the positively charged guanidino groups of Arg 37, Arg 212 and Arg 287 which interact with the carboxylate group of substrate and inhibitor.   Residue Arg 71 in the C5 (**Figure 5**, orange residues) pocket functions as a hydrogen bond donor and comprises residues Trp 98, Arg 144 and Ile 142 which create a small hydrophobic region.   In the C4 (**Figure 5**, cyan residues) pocket, a guanidine or amine group of the substrate or inhibitor participates in a charge-charge interaction and hydrogen bonds to Glu 38, Asp 70 and Glu 147.   Within the C6 pocket (**Figure 5**, yellow residues), Glu 197, the side chain of Arg 71, the amidic carbonyl of Trp 98 and Asp 70 forms a new hydrophobic binding pocket.
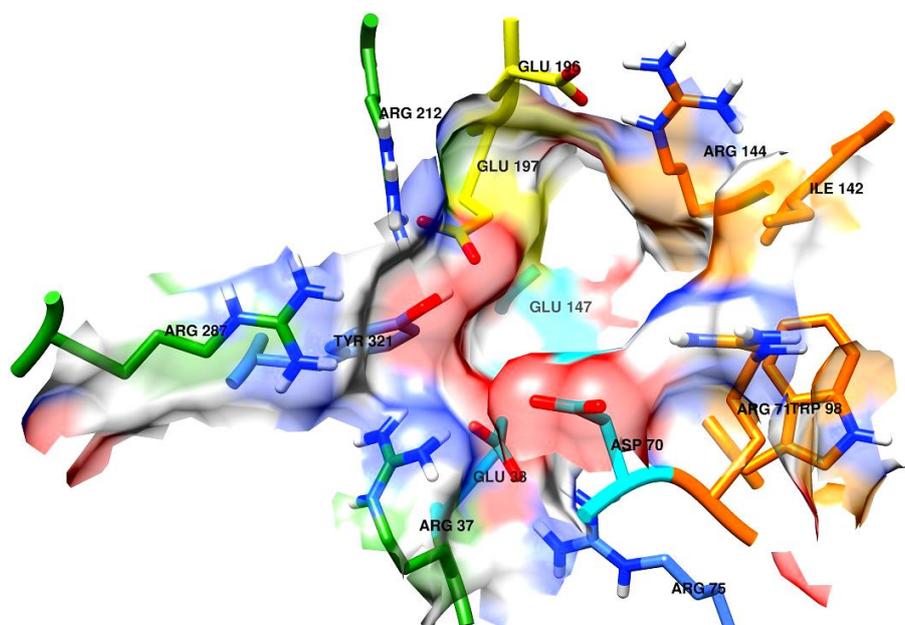
Figure 5: Outlining the residues of the four binding site pockets; (Pocket C1: green), (Pocket C4: cyan), (Pocket C5: orange), (Pocket C6: yellow)

3.1 Ligand-protein interaction pattern and per-residue free binding energy profile

Post-dynamic analysis of the substrate (sialic acid) and active inhibitors (oseltamivir, laninamivir, peramivir and zanamivir) complexed with neuraminidase, revealed that each system achieved convergence (see RMSD and potential energy plots in Supplementary Information, **S1** and **S2**). The MM-GBSA algorithm was used to describe the per-residue binding free energies. The ranking of contributions of interacting residues highlighted critical pharmacophoric features of each inhibitor used in this study; see **Table 1**. A noticeable trend in the system energetics of all complexes was observed. Hydrogen bond donation from protonated arginine residues 37, 71, 144, 212 and 287; to the ligand had unfavourable energy contributions, whereas the hydrogen bond acceptor residues have an appreciable per-residue binding free energy contribution. A possible explanation for this occurrence is that salt bridges have a distinct dual function: participating in both hydrogen bonding and ionic interactions. The energy term for residues engaged in salt bridges, is largely unfavourable due to desolvation, as the interaction endures significant entropic loss. The ubiquitous nature

of hydrogen bonds in an aqueous environment sometimes requires desolvation of the donor and acceptor prior to formation. Thus, the per-residue free binding energies of hydrogen bond donors of the enzyme, reflected in the MM-GBSA profiles are in a positive scale due to enhanced polar solvation properties.

Table 2: Common active site residues in the binding pattern of each inhibitor

Inhibitor	Common Residues												
	Arg 37	**Glu 38**	**Asp 70**	**Arg 71**	**Trp 98**	**Ile 142**	Arg 144	**Glu 147**	**Glu 196**	**Glu 197**	Arg 212	Arg 287	**Tyr 321**
Oseltamivir	√	√	√	√	√	√			√	√	√	√	√
Zanamivir		√	√	√	√	√	√	√	√	√	√		√
Laninamivir	√	√	√		√	√	√	√	√	√	√	√	√
Peramivir	√	√	√	√	√	√	√	√	√	√	√	√	√

The interaction pattern of crystallographic structures of the four inhibitors corresponds with reported interactions in literature. **Table 1**, lists the common interactions observed in the inhibitor complexes. Residues Glu 38, Asp 70, Arg 71, Trp 98, Ile 142, Glu 147, Glu 196, Glu 197 and Tyr 321 of **Table 1**; represent residues with a high binding free energy contribution. The two-dimensional binding of inhibitors and substrate; oseltamivir (**Figure 6**), zanamivir (**Figure S3**), laninamivir (**Figure S4**), peramivir (**Figure S5**) and sialic acid (**Figure S6**) highlight the residues with high contributing binding free energies.
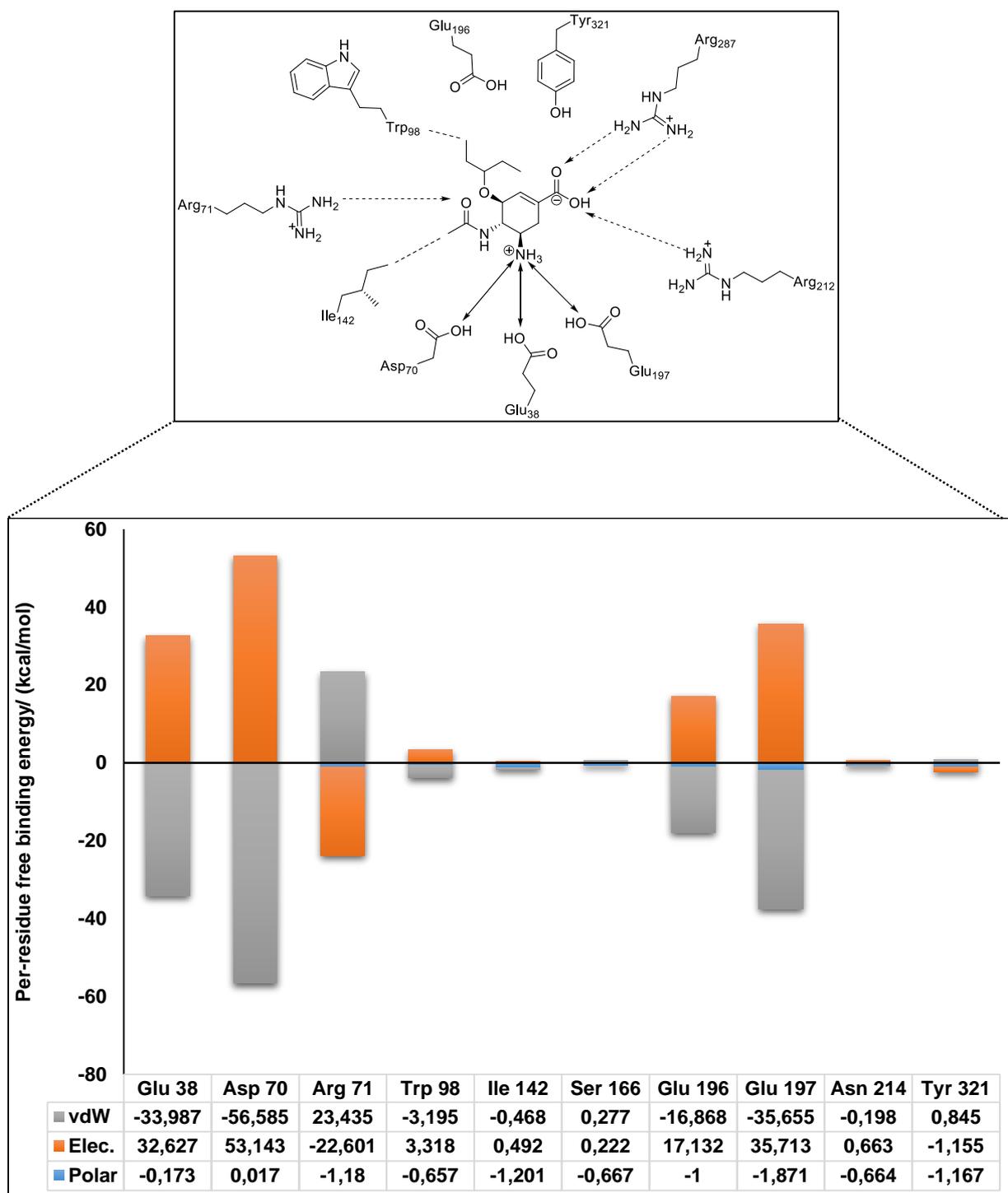
Figure 6: 2-D interaction profile of oseltamivir with residues of high binding free energies

3.2 Pharmacophore model construction, evaluation and database screening

3.2.1 Ligand-based pharmacophore model

The Espresso algorithm, a built-in heuristics modelling tool of LigandScout was used to predict

the ten hypothesised models, as shown in **Table 2**.   The calculation implements a parallel

alignment of pharmacophoric moieties of the training set molecules. This was done by the automatic assignment and overlaying of distinct binding modes of the ligands until a geometrically adequate representation of all five training set molecules was achieved. Boolean minimisation refined the pattern matching resulting in a near optimal solution of all possible conformations. Each model was evaluated using two screening sets: 1) active set and 2) decoy set of molecules, as described in **Section 3.4.1**. This methodology properly assesses the ability of the pharmacophore models to adequately detect active compounds over decoys. Pharmacophore model 1 expressed the greatest specificity for detecting actives over inactives. The predicted model also aligned well with the profile of the per-residue binding free energy pharmacophoric pattern. The LigandScout package calculates the area under the curve (AUC) as well as the enrichment factor (EF), which was estimated at 0.82 and 35.1% respectively. The pharmacophoric model qualities are represented in the receiver operating characteristic (ROC) curve, **Figure 7**.

Table 3: Hypothesised models 1-10

Model number	Scoring function	Pharmacophoric Features
1	0.8342	5 x HBA, 4 x HBD, NI, PI
2	0.8301	4 x HBA, 3 x HBD, NI, PI
3	0.8251	5 x HBA, 3 x HBD, PI
4	0.8242	5 x HBA, 3 x HBD, PI
5	0.8224	5 x HBA, 3 x HBD, PI
6	0.8017	5 x HBA, 4 x HBD, NI, PI
7	0.7997	6 x HBA, 3 x HBD, NI
8	0.7996	6 x HBA, 3 x HBD, NI
9	0.7996	6 x HBA, 3 x HBD, NI
10	0.7774	6 x HBA, 3 x HBD, PI

*HBA- hydrogen bond acceptor, HBD- hydrogen bond donor, NI- negative ionisable species, PI- positive ionisable species.
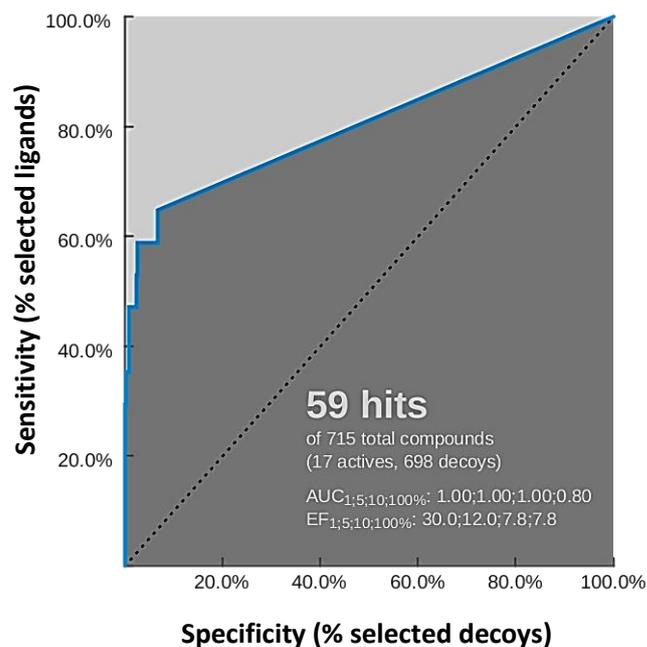
Figure 7: ROC curve of model 1 of the ligand-based pharmacophore

Validation of pharmacophore model 1 (**Figure 8**) prompted its use in a ligand based virtual screen. A library of molecules with drug-like properties was extracted from the ZINC database for use in the *in silico* screening against the selected ligand-based pharmacophore model. The minimum feature requirement was set to three and possible hits were detected using the best matching conformation method and scored based on their pharmacophore-fit. A total of 5879 compounds was screened of which only 23 compounds satisfied the predicted pharmacophore model. These molecules were then subjected to a structure based virtual screening which was used as a further filtering tool.
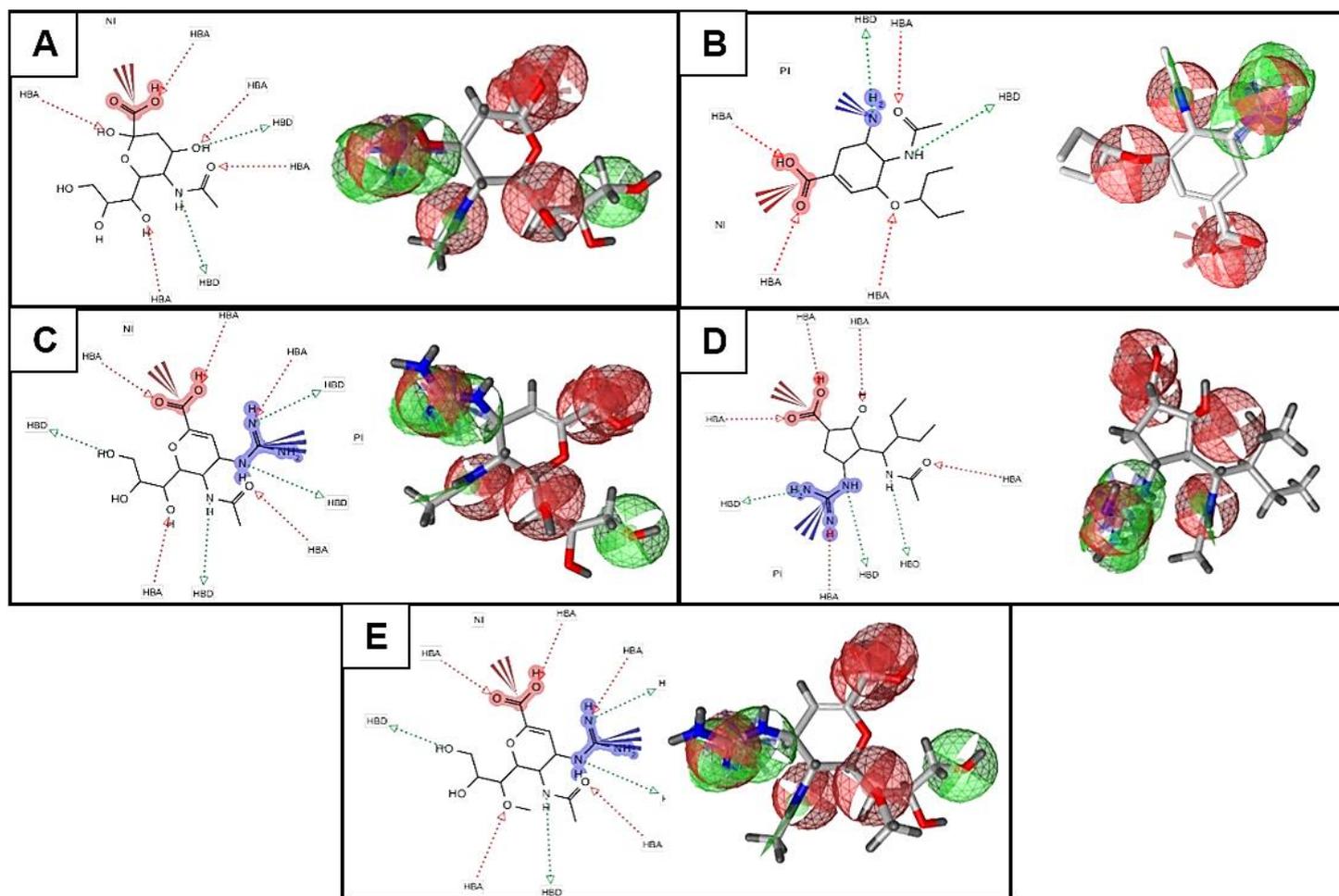
Figure 8: 2-D and 3-D representations of Model-1 with; A) sialic acid, B) oseltamivir, C) zanamivir, D) peramivir and E) laninamivir overlay.

3.2.2 Structure-based pharmacophore model

A dynamic pharmacophore was generated by superimposing four hypothesised models of each inhibitor. With the aid of LigandScout software the interactions between protein and ligand, as well as some of the excluded volume spheres corresponding to 3-D protein structures, were interpreted. The protein-ligand complexes of laninamivir, peramivir, zanamivir and oseltamivir displayed a significant overlap in their binding patterns. Common interaction patterns were observed with catalytic residues Arg 37, Arg 70, Arg 71, Arg 212 and Arg 286. Glu 196 was also highlighted and it is known to facilitate the electrostatic funnelling of ligands into the active site [4]. Other important residues included amino acids: Glu 38, Arg 75, Trp 98,

Ile 142, Arg 144, Glu 147, Glu 197, Arg 287, Tyr 321; all of which contribute to the 3-D integrity of the binding pockets.

Overlapping the chemical features of the complexed inhibitors resulted in a structure-based pharmacophore model with four HBA's, a PI and NI species as well as exclusion volumes, as seen in **Figure 9**. The inclusion of exclusion volumes allows the pharmacophore to accurately describe the spatial interface of the ligand bound in the binding pocket, such that steric regions that are not occupied by the active ligand are mapped. The model generated was in close approximation to the high per-residue contributors of binding free energy described in **Section 3.1**. Enrichment metrics of the model estimated 0.7 (AUC) and 47.1 % (EF), (**Figure 10**). The predicted model was screened against the 5879 drug-like molecules from the ZINC database. The resulting hit-list of five molecules which suitably satisfied the model's criteria was subjected to further study by molecular docking.
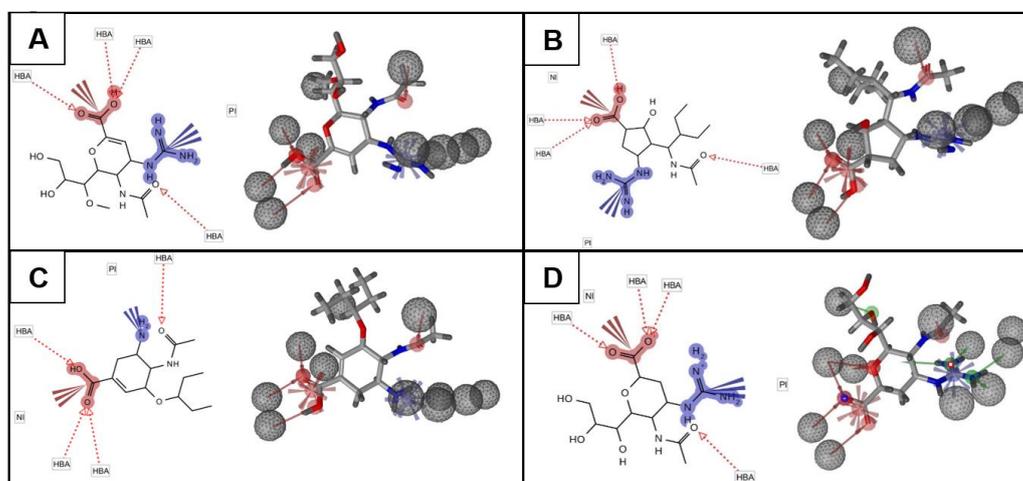

Figure 9: 3-D and 2-D representations of the hypothesised model with; A) oseltamivir, B) peramivir, C) laninamivir and D) zanamivir overlay
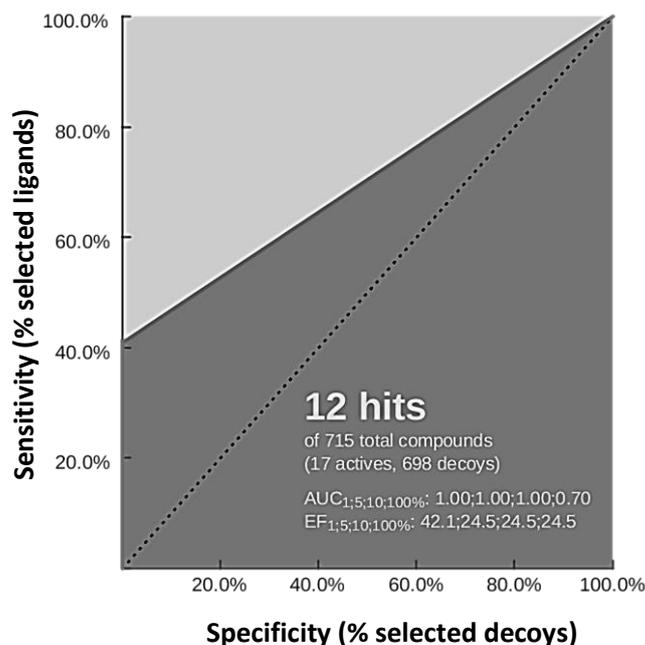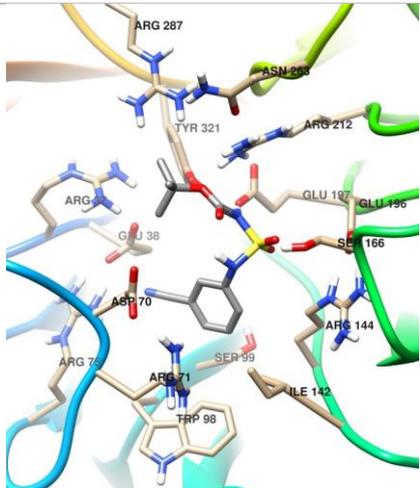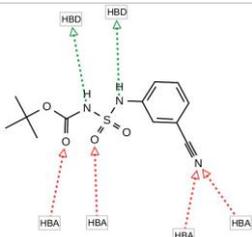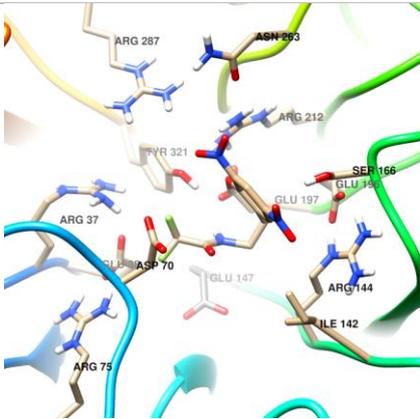
Figure 10: ROC curve of the structure-based pharmacophore

3.3 Molecular Docking

The crystallographic structure of oseltamivir (PDB ID: 3TI6) complexed with neuraminidase, was selected for the molecular docking study. Validation of the docking system by a re-docking (see **Table S2**) and cross-docking (refer to **Table S3**) procedure established that the docking software was capable of placing the ligand in the correct conformation within the binding site. Docking a combined library of known inhibitors with decoy compounds established that AutoDock was capable of selecting actives from inactives. The list of hit compounds detected from the screening against the two predicted pharmacophore models was dock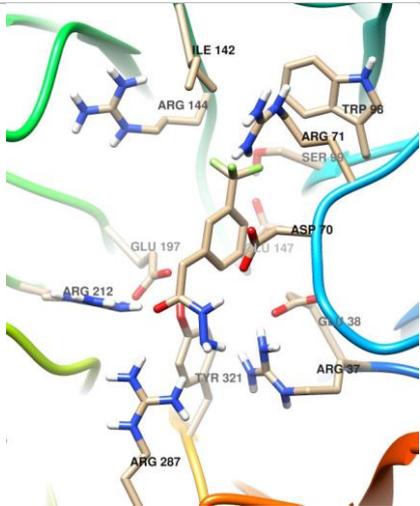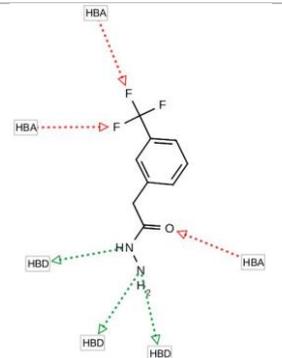ed into the neuraminidase binding site. A total of 28 molecules was docked and ranked according to binding free energy values. The five highest ranked ligands from each of the hit lists of the ligand-based pharmacophore and structure-based pharmacophore, were analysed to interpret their significant ligand characteristics and binding interaction patterns with the receptor. **Table 3**, lists the selected compounds for further analysis.

Table 4: Top five docked ligands of LB and SB pharmacophore hit lists

Ligand-based			
3-D structure	**2-D structure**	**Docking Binding free energy/ (kcal/mol)**	**MD Binding free energy/ (kcal/mol)**
	Ligand 1 (ZINC 12336793)	-7.00	-26.68 +/- 4.21
	Ligand 2 (ZINC 43828061)	-7.20	-19.05 +/- 2.87
	Ligand 3 (ZINC 89224608)	-6.90	-16.79 +/- 5.33
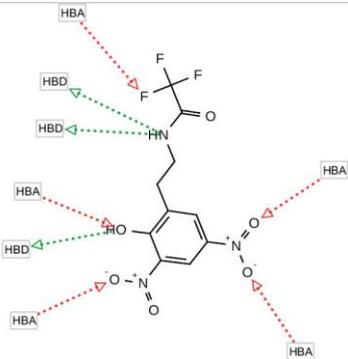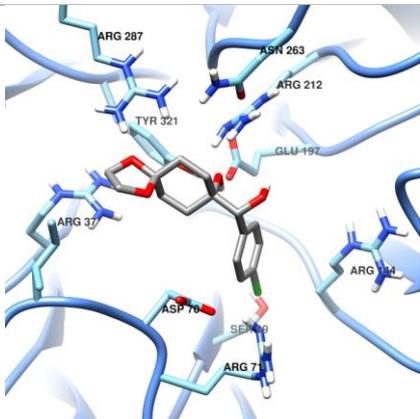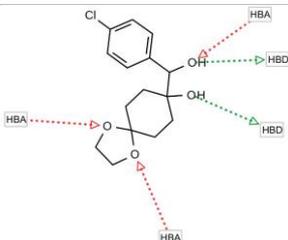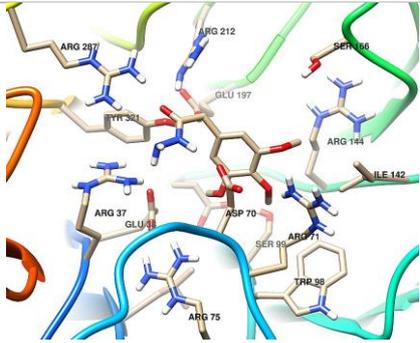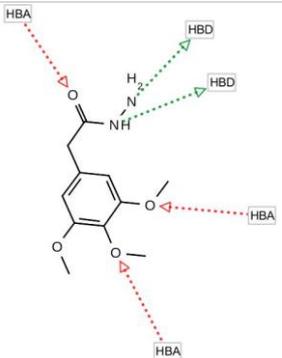
	Ligand 4 (ZINC 34491253)	-7.00	-12.62 +/- 6.60
	Ligand 5 (ZINC 37387871)	-7.20	-8.80 +/- 8.31
Structure-based			
	Ligand 6 (ZINC 32911334)	-6.20	-24.08 +/- 3.31
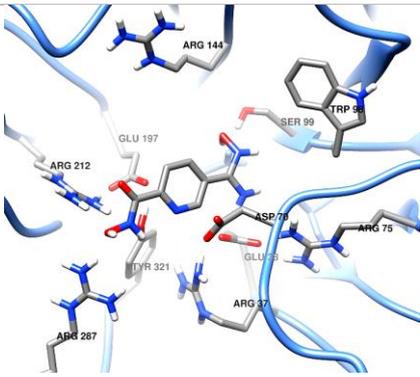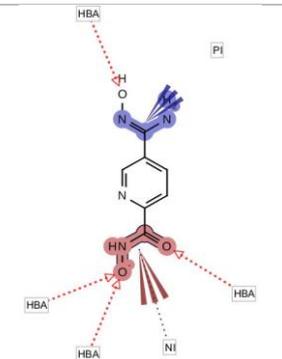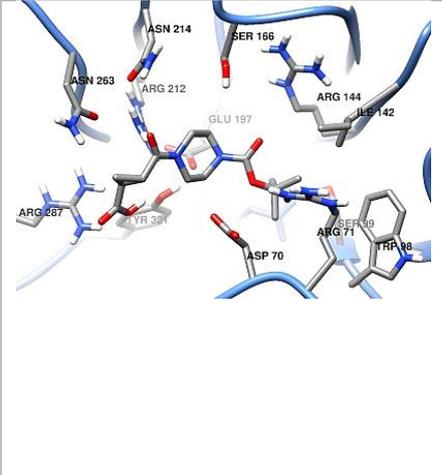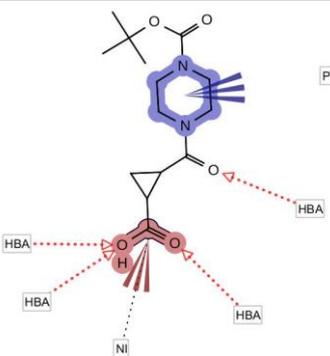
	 Ligand 7 (ZINC 36532822)	-6.70	-23.45 +/- 3.37
	 Ligand 8 (ZINC 29753816)	-6.90	-23.30 +/- 6.17
	 Ligand 9 (ZINC 08729923)	-8.10	-20.00 +/- 3.46
	 Ligand 10 (ZINC 00158858)	-6.40	-17.66 +/- 2.15
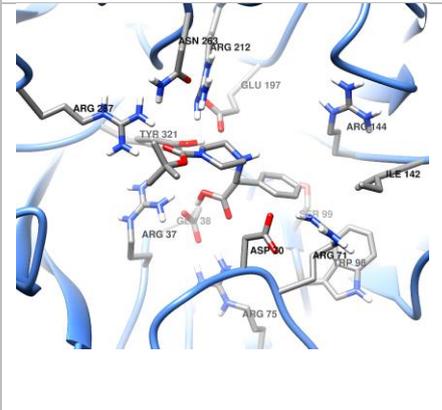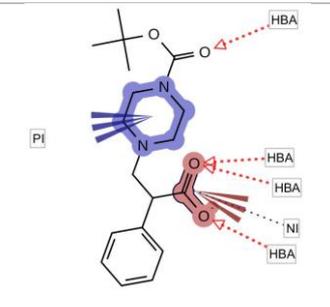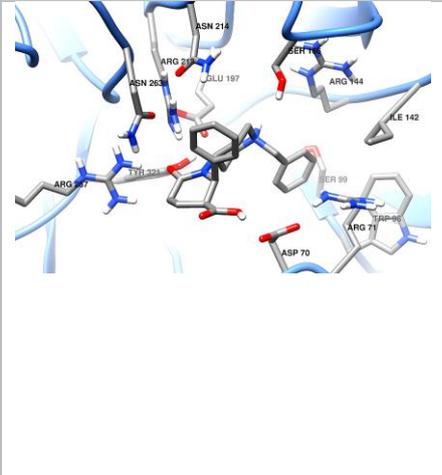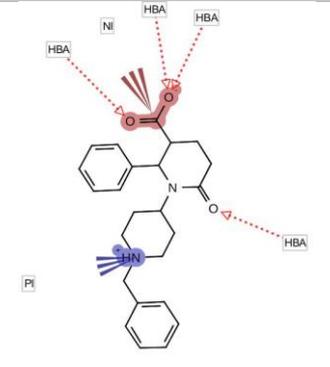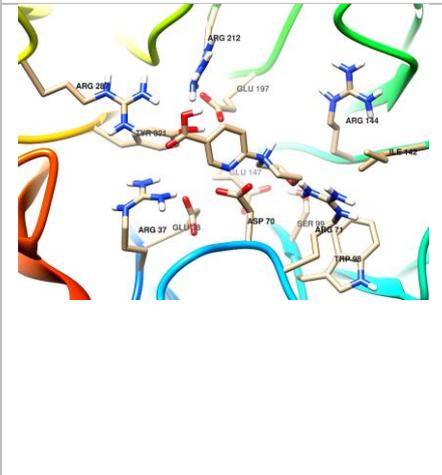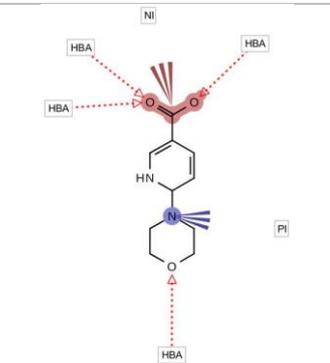
These ligands were further subjected to a 20 ns MD simulation. The root mean square deviation (RMSD) and potential energy plots of ligands 1-10 confirmed the convergence of the MD protein-ligand complexes at approximately 25 ns. See supplementary information (**Figure S7 to S10**). The root mean square fluctuation (RMSF) plots closely resemble the residue contribution profile of inhibitor oseltamivir.

Based on the predicted binding free energy (ΔG_{bind}), ligand 1 is ranked as having the most favourable binding interaction with the receptor followed by ligands 6, 7, 8, are comparable to the binding free energy of active inhibitor oseltamivir. Ligands 2, 3, 4, 5, 9 and 10, however, exhibit weaker interaction contributions with the receptor. However, the correlation of the docked energies with calculated binding free energies of the ligand-based pharmacophore hits are better than the structure-based pharmacophore. The total binding free energy was further decomposed into its components. The four inhibitors (oseltamivir, zanamivir, laninamivir and peramivir), shared a similar energy contribution trend, the electrostatic energy (ΔE_{ele}) contribution was counter-balanced by the polar solvation energy (ΔG_{GB}) contribution. A significant proportion of the binding free energy comes from the van der Waals energy (ΔE_{vdW}) contributor. The same trend is observed for the ligands extracted from the virtual screen. Structurally the ligands are different to the current drug inhibitors and, as previously mentioned, the more basic the ligand, the more potent it is in inhibiting the enzyme [50].

The binding pattern of ligand 1 displayed significant hydrophobic interactions between the conjugated ring and residues Arg 71, 144 and Ile 142, **Table 3**. The hydrogen bond donation from the protein receptor residues Arg 75, 144, 212 with the sulphonyl amide group, was similar to the substrate and active inhibitors. Further to this, Arg 71 expressed a weak π - cation interaction between the face of the electron rich π-system of the aromatic ring. The total binding free energy of ligand 1 is -26.68 (kcal/mol), **Table 4**. The molecule has three distinctive structural features surrounding the phenyl anchor. The cyano group is

accommodated in the C4 pocket. The sulfamoylcarbamate and *tert*-butyl group extends from the C1 to the C6 binding pockets. Overall, the molecule offers many electron pairs which can attribute to the basicity of the ligand. The quasi-harmonic entropic approximation (ΔS) of ligand 1 comprised of three aspects translational, rotational and vibrational contributions. The measure of disorder for each estimated -12.80, -10.10 and -49.54 (kcal/mol), respectively. Collectively giving a ΔS of -72.43 (kcal/mol) suggesting an ordered compact three-dimensional protein complex. Hydrophobic interactions of ligand 2 were observed between Asp 70 and Arg 144 with the hydroxyl linker and aromatic ring functionalities, respectively. The ligand structure appears to be interlocked and rigid due to the fused ring complement, resulting in a restricted hydrogen bond network with residues Glu 197, Arg 212 and Arg 287. The unique ringed structure of 1,3-dioxolane sits in the C1 pocket, decanol occupies the C6 and C4 pockets and a chlorophenyl group resides in the C5 pocket with an overall ΔG_{bind} of -19.05 (kcal/mol). The estimated ΔS of the system was measured to be -71.84 (kcal/mol). With a translational contribution of -12.87 (kcal/mol), a rotational contribution of -10.14 (kcal/mol) and a vibrational contribution of -48.83 (kcal/mol). This suggests an ordered three-dimensional protein complex. Ligand 2 acts as a weak acid *i.e.* a hydrogen bond donor, but has improved lipophilicity compared to ligand 1.

Ligand 3 is an electronegative rich molecule with a hydrophobic interaction between Glu 197 and the trifluoromethyl moiety. The hydrogen bond network observes significant hydrogen bond donation from residues Ser 166, Arg 212 and Tyr 321, whilst Glu 197 accommodates a hydrogen interaction by donation from the amide group. A bridging water molecule mediates an interaction between Arg 287 and the trifluoromethyl group. The predicted ΔG_{bind} for ligand 3 estimated -16.79 (kcal/mol). The compound comprises four substituents extending from a central phenyl ring. There are two oxohydroxylammonium groups that occupy the C1 and C5 pockets as hydrogen bond acceptors. A hydroxyl group sits in the C6 pocket which acts

as an HBA and HBD, and a trifluoroacetamido ethyl group which contains a range of HBA and HBD moieties, is accommodated in the C4 pocket. The compound has an overall acidic profile. The entropic contribution estimated -42.00 (kcal/mol) comprising -12.80 (translational quasi-harmonic approximation (qha)), -10.29 (rotational qha) and -18.91 (vibrational qha) (kcal/mol), respectively. The trifluoromethyl moiety of ligand 4 expressed an electrostatic interaction with Lewis base residues Arg 71 and Trp 98. The dipole created by opposing side chain group's *i.e.* the trifluoromethyl group and acetohydrazide, is supported by Asp 70 *via* a bridging water and a hydrophobic interaction. Asp 70 also functions as both a HBA and HBD. Additional hydrogen bond contributions from Arg 212 and Arg 287 were established. Ligand 4 has two defined structural features attached to the phenyl anchor, the first being the HBA, trifluoromethyl group which partially occupies the C5 and C6 binding pocket, and the HBD, acetohydrazide which is accommodated in the C1 pocket. The molecule is observed to have a weak Lewis acid chemical profile with a predicted ΔG_{bind} of -12.62 (kcal/mol). Ligand 4, has a total entopic quasi-harmonic approximation of -65.37 (kcal/mol), translational qha of -12.61 (kcal/mol), vibrational qha of -9.87 (kcal/mol) and vibrational qha measuring -42.89 (kcal/mol). Ligand 5 reflected a defined hydrogen bond interaction pattern with acetohydrazide functional group and residues: Asp 70, Arg 71, Arg 212, and Arg 287. The structure of ligand 5 comprises three identical methyl ethers and are chemically recognised as HBA's that occupy and overlap with the binding pockets C4, C5 and C6. The molecule also contains an acetohydrazide that possesses an HBA and HBD moiety which interacts with residues of the C1 binding pocket. The compound has an overall basicity with a calculated binding free energy estimating -8.80 (kcal/mol). The measured ΔS reflects an ordered protein profile of -43.14 (kcal/mol), and a translational, rotational and vibrational qha of -12.52, -9.51 and -21.10 (kcal/mol), respectively.

Table 5: Calculated binding free energies based on the MM-GBSA method

System	ΔE_{ele}	ΔE_{vdW}	ΔG_{SA}	ΔG_{GB}	ΔE_{gas}	ΔG_{sol}	ΔG_{bind}
Oseltamivir	-9.44 +/- 3.52	-30.29 +/- 5.45	-3.88 +/- 0.42	21.22 +/- 3.63	-39.73 +/- 6.30	17.34 +/- 3.43	-22.40 +/- 5.44
Laninamivir	-169.73 +/- 17.48	-32.95 +/- 3.62	-5.12 +/- 0.42	183.20 +/- 14.00	-202.68 +/- 16.94	178.08 +/- 13.95	-24.60 +/- 5.70
Peramivir	-153.79 +/- 12.78	-35.42 +/- 2.95	-5.33 +/- 0.27	163.22 +/- 10.75	-189.21 +/- 12.71	157.89 +/- 10.70	-31.32 +/- 4.26
Zanamivir	-189.20 +/- 18.38	-34.62 +/- 3.24	-5.52 +/- 0.32	198.29 +/- 13.75	-223.82 +/- 18.43	192.78 +/- 13.64	-31.04 +/- 6.87
Ligand 1	-4.09 +/- 5.28	-32.86 +/- 4.46	-4.36 +/- 0.64	14.63 +/- 4.33	-36.95 +/- 6.79	10.27 +/- 4.26	-26.68 +/- 4.21
Ligand 2	-8.11 +/- 4.46	-27.75 +/- 3.56	-3.65 +/- 0.46	20.45 +/- 4.42	-35.85 +/- 5.22	16.80 +/- 4.35	-19.05 +/- 2.87
Ligand 3	-2.79 +/- 3.39	-23.67 +/- 7.10	-3.21 +/- 0.97	12.87 +/- 5.31	-26.45 +/- 8.75	9.66 +/- 4.56	-16.79 +/- 5.33
Ligand 4	-13.22 +/- 10.76	-19.30 +/- 5.63	-2.95 +/- 1.02	22.84 +/- 11.31	-32.52 +/- 14.72	19.89 +/- 10.58	-12.63 +/- 6.60
Ligand 5	-1.50 +/- 3.15	-13.02 +/- 12.09	-2.11 +/- 1.95	7.83 +/- 7.69	-14.52 +/- 13.75	5.72 +/- 5.85	-8.80 +/- 8.31
Ligand 6	-209.42 +/- 15.96	-12.48 +/- 2.93	-3.13 +/- 0.21	200.95 +/- 13.84	-221.90 +/- 15.24	197.82 +/- 13.81	-24.08 +/- 3.31
Ligand 7	-5.48 +/- 2.45	-30.21 +/- 3.41	-3.78 +/- 0.49	16.02 +/- 3.33	-35.69 +/- 5.00	12.24 +/- 2.98	-23.45 +/- 3.37
Ligand 8	-16.27 +/- 5.64	-33.29 +/- 4.55	-4.75 +/- 0.55	31.01 +/- 3.96	-49.56 +/- 8.31	26.25 +/- 3.73	-23.30 +/- 6.17
Ligand 9	-55.34 +/- 14.05	-27.31 +/- 3.22	-3.74 +/- 0.39	66.39 +/- 13.50	-82.66 +/- 13.83	62.65 +/- 13.42	-20.00 +/- 3.46
Ligand 10	-0.57 +/- 2.99	-24.74 +/- 1.85	-3.14 +/- 0.20	10.80 +/- 2.75	-25.31 +/- 3.52	7.65 +/- 2.71	-17.66 +/- 2.15

*the energy values reported are measured in kcal/mol: $\Delta G_{bind} = \Delta E_{gas} + \Delta G_{sol} - T\Delta S$; $\Delta E_{gas} = \Delta E_{int} + \Delta E_{vdW} + \Delta E_{ele}$; $\Delta G_{sol} = \Delta G_{GB} + \Delta G_{SA}$.

An extensive hydrogen bonding network of ligand 6 with HBD's: Arg 37, Asp 70, Arg 75, Ser 99, Arg 212, Arg 287 and HBA's: Glu 38, Trp 98, Glu 147 and Tyr 321 exists.   A water bridge

created by Arg 287 supplements an electrostatic interaction.   Ligand 6 was ranked with the most favourable ΔG_{bind} measuring -24.08 (kcal/mol).   The ligand has three key structural features; 1) the pyridine anchor, 2) the hydroxamic acid, and 3) the hydroxycarbamimidoyl. The hydroxamic acid acts as an HBA, residing in the C1 binding pocket.   The hydroxycarbamimidoyl group's interaction is shared with the C4 and C6 pockets.   The molecule assumes a very weak acid profile perceiving it to be a potential inhibitor.   Ligand 7 had a more conserved binding pattern, with a hydrophobic interaction between Ile 142 and tert-butyl group.   Arg 212 and Tyr 321 are delegated as a hydrogen bond donor and acceptor, respectively.   Salt bridges exerted by Arg 37 and Arg 287 allow for the accommodation of the carboxylate group.   Ligand 7 comprises a keto-cyclopropanecarboxylic acid which interacts with the C1 binding pocket residues.   A piperazine moiety anchors the molecule and engages in a charge-charge interaction in the C4 binding pocket.   The C5 pocket accommodates the *tert*-butoxycarbonyl group.   The molecule has an overall ΔG_{bind} of -23.45 (kcal/mol).   Ligand 8 shared a similar binding profile to ligand 7.   A distinct hydrophobic interaction with residue Arg 144 exists.   A hydrogen bond network between residues Glu 38, Arg 75 and Arg 287 and the carboxylate group exists.   Salt bridges created by Arg 37, Asp 70 and Glu 197 provide valuable contributions to electrostatic interactions.   The composites of ligand 8, contrast with its counterparts.   It is composed of piperazine with a *tert*-butoxycarbonyl group that facilitates an HBA interaction with the C1 pocket.   The C4, C5 and C6 binding pockets promote an interaction with the carboxylic acid linker, HBA and phenyl ring.   The molecule mimics weak acid properties with a total binding free energy of -23.30 (kcal/mol).   The complexation of ligand 9 offered three distinct salt bridges: Arg 37, Arg 212 and Arg 287 to the carboxylate group.   Ligand 9 has an estimated ΔG_{bind} of -20.00 (kcal/mol).   It has steric bulk attributed by the chain of two phenyl rings and two piperadine groups which reside in the C1 and C5 binding pockets.   The carboxylic acid group is a

significant HBA contributor that interacts with the C6 pocket residues.   The composition of ligand 10 incorporates a pyridine, morpholine and a carboxylic acid which reside in the C6, C4 and C1 binding pockets, respectively.   The ligand has a strong affinity predicted ΔG_{bind} of -17.66 (kcal/mol).   The aromatic ring on ligand 10 demonstrated appreciable hydrophobic interactions between residues Asp 70, Arg 71, Ile 142 and Arg 144.   Arg 71 formed a π-cation interaction with the face of the benzene ring electron cloud.   Hydrogen bond donation from residues Arg 212 and 287 with the pyridine rings was observed.

3.4 Similarity

The novelty of the ligands identified was confirmed using PubMed.   Ligand 1 and 3 share no similarity with any reported inhibitor in literature.   Ligand 2 displayed no biological activity in literature, although the 2,4-dinitrophenolate group has become the basis of the design of innovative anti-cancer agents (CID 229329).   Ligand 5, has not been previously reported as a potential neuraminidase inhibitor.   However, analogues of the 2-phenylacetohydrazide have been evaluated for their inhibition of Microbial tyramine oxidase (AID: 215960).   The 2-(3,4,5-trimethoxyphenyl)acetohydrazide of ligand 4, serves as an important framework for active inhibitors against human AChE and human plasma BChE.   Ligand 6 (5-[amino(hydroxyimino)methyl]-N-hydroxy-2-pyridinecarboxamide), 7 (2-{[4-(tert-butoxycarbonyl)piperazino]carbonyl}cyclopropanecarboxylic acid) and 8 (3-[4-(tert-butoxycarbonyl)piperazino]-2-phenylpropanoic acid), share no similarity.   The piperazine of ligand 7 and 8, is a heterocycle which influences the biological activity by conferring metabolic stability and potentiating interactions with macromolecules [51-53].   The compound, 6-morpholinonicotinic acid (ligand 10) has previously been tested in six bioassays (SID11536988); 1) influenza A RNA, 2) DnaK protein chaperone, 3) homo sapien peptidyl-prolyl cis-trans isomerase FKBP1A isoform, 4) influenza A virus M2 ion channel, 5) homo sapien ubiquitin-conjugating proteasome E2, and 6) the FRB domain of mTOR.   The molecule

has demonstrated no activity in each of these systems. The 1-benzylpiperidine structure of ligand 9 serves as a template for human betaine/GABA transporter 1 inhibitor *N*-(1-benzyl-4-piperidinyl)-2,4-dichlorobenzamide (BPDBA) [54].

4. Conclusion

The emergence of resistance in neuraminidase of the influenza A virus has encouraged research methods incorporating high-throughput screening. Current methodologies propose that the structure-based pharmacophore predicted model as being superior to the ligand-based pharmacophore. Our proof of concept in using per-residue binding free energies obtained from an MD ensemble using the MM-GBSA calculation to generate the ligand-based pharmacophore, has proven to be an invaluable additional technique for the generation of a suitable pharmacophore model. The structure-based pharmacophore generated in the study has six pharmacophoric features, whereas, the ligand-based pharmacophore has 11 pharmacophoric features that overlap with the interaction pattern mapped by favourable contributions of per-residue binding free energy. The greater number of pharmacophoric features expands the screen for similar structured molecules. It also introduces improved selectivity with greater structural diversity. The five highest ranking molecules of the ligand-based pharmacophore hit-list demonstrated a better correlation of its docking energies with the calculated MM-GBSA binding free energy. Despite the predicted ΔG_{bind} of the ligand-based docked complexes having a reduced affinity, a defined interaction of the chemical specific side-chains with the specific binding pockets is observed. The ease of drugability of these compounds based on lipophilicity makes them even more attractive as potential inhibitors.

Introducing MD simulation into the workflow in pharmacophore generation, creates an energetically enhanced 3-D profile of the ligand-protein complex. The interactions are thus based on an optimised conformation. The decomposition of the binding free energy into the per-residue contributions using MM-GBSA calculations allows us to organise and rank the

interacting residues and subsequent chemical features, by their favourable energetics. The potential inhibitors filtered through LBP virtual screens are more structurally specific for the target. Thus the use of this technique would vastly improve the efficacy of *in silico* screening methods which, in turn expedites the design and development of innovative inhibitors to prevent future disease. The results of this study form a strong basis for further investigation of the identified high ranking ligands 1-10 from the ligand- and structure-based virtual screen as potential inhibitors of influenza neuraminidase.

5. References

1. Shao, T.-j.; Li, J.; Yu, X.-f.; Kou, Y.; Zhou, Y.-y.; Qian, X., Progressive antigenic drift and phylogeny of human influenza A(H3N2) virus over five consecutive seasons (2009–2013) in Hangzhou, China. *International Journal of Infectious Diseases* **2014,** *29*, 190-193.

2. Air, G. M.; West, J. T., Antigenic Variation☆. In *Reference Module in Biomedical Sciences*, Elsevier: 2014.

3. Carrat, F.; Flahault, A., Influenza vaccine: The challenge of antigenic drift. *Vaccine* **2007,** *25* (39–40), 6852-6862.

4. Le, L.; Lee, E. H.; Hardy, D. J.; Truong, T. N.; Schulten, K., Molecular Dynamics Simulations Suggest that Electrostatic Funnel Directs Binding of Tamiflu to Influenza N1 Neuraminidases. *PLoS Comput Biol* **2010,** *6* (9), e1000939.

5. Patel, R. B.; Mathur, M. B.; Gould, M.; Uyeki, T. M.; Bhattacharya, J.; Xiao, Y.; Khazeni, N., Demographic and clinical predictors of mortality from highly pathogenic avian influenza A (H5N1) virus infection: CART analysis of international cases. *PloS one* **2014,** *9* (3), e91630.

6. Organization, W. H. Pandemic preparedness. http://web.archive.org/web/20030202145905/http://www.who.int/csr/disease/influenza/pandemic/en/ (accessed 21 April).

7. Cohen, E. When a pandemic isn't a pandemic. http://edition.cnn.com/2009/HEALTH/05/04/swine.flu.pandemic/index.html (accessed 22 April).

8. Doshi, P., The elusive definition of pandemic influenza. *Bull World Health Organ* **2011,** *89* (7), 532-8.

9. Kim, C. U.; Lew, W.; Williams, M. A.; Liu, H.; Zhang, L.; Swaminathan, S.; Bischofberger, N.; Chen, M. S.; Mendel, D. B.; Tai, C. Y.; Laver, W. G.; Stevens, R. C., Influenza neuraminidase inhibitors possessing a novel hydrophobic interaction in the enzyme active site: design, synthesis, and structural analysis of carbocyclic sialic acid analogues with potent anti-influenza activity. *Journal of the American Chemical Society* **1997,** *119* (4), 681-90.

10. Han, N.; Liu, X.; Mu, Y., Exploring the mechanism of zanamivir resistance in a neuraminidase mutant: a molecular dynamics study. *PloS one* **2012,** *7* (9), e44057.

11. Takashita, E.; Ejima, M.; Itoh, R.; Miura, M.; Ohnishi, A.; Nishimura, H.; Odagiri, T.; Tashiro, M., A community cluster of influenza A(H1N1)pdm09 virus exhibiting cross-resistance to oseltamivir and peramivir in Japan, November to December 2013. *Euro surveillance : bulletin Europeen sur les maladies transmissibles = European communicable disease bulletin* **2014,** *19* (1).

12. Shobugawa, Y.; Saito, R.; Sato, I.; Kawashima, T.; Dapat, C.; Dapat, I. C.; Kondo, H.; Suzuki, Y.; Saito, K.; Suzuki, H., Clinical effectiveness of neuraminidase inhibitors--oseltamivir, zanamivir, laninamivir, and peramivir--for treatment of influenza A(H3N2) and A(H1N1)pdm09 infection: an observational study in the 2010-2011 influenza season in Japan. *Journal of infection and chemotherapy : official journal of the Japan Society of Chemotherapy* **2012,** *18* (6), 858-64.

13. Sifferlin, A. H1N1's Death Toll: 15 Times Higher than Previously Thought. http://healthland.time.com/2012/06/26/h1n1s-death-toll-15-times-higher-than-previously-thought/ (accessed 23 April).

14. Charles Patrick Davis, M., PhD Swine Flu (Swine Influenza A [H1N1 and H3N2v] Virus. http://www.medicinenet.com/swine_flu/article.htm (accessed 23 April).

15. Baz, M.; Abed, Y.; Papenburg, J.; Bouhy, X.; Hamelin, M.-È.; Boivin, G., Emergence of Oseltamivir-Resistant Pandemic H1N1 Virus during Prophylaxis. *New England Journal of Medicine* **2009,** *361* (23), 2296-2297.

16. Prevention, C. f. D. C. a. Influenza A (H3N2) variant virus. http://www.cdc.gov/flu/swineflu/h3n2v-cases.htm (accessed 25 April).

17. Huang, L.; Cao, Y.; Zhou, J.; Qin, K.; Zhu, W.; Zhu, Y.; Yang, L.; Wang, D.; Wei, H.; Shu, Y., A conformational restriction in the influenza A virus neuraminidase binding site by R152 results in a combinational effect of I222T and H274Y on oseltamivir resistance. *Antimicrobial agents and chemotherapy* **2014,** *58* (3), 1639-45.

18.   Adabala, P. J.; LeGresley, E. B.; Bance, N.; Niikura, M.; Pinto, B. M., Exploitation of the catalytic site and 150 cavity for design of influenza A neuraminidase inhibitors. *The Journal of organic chemistry* **2013,** *78* (21), 10867-77.

19.   Lu, X.; Liu, F.; Zeng, H.; Sheu, T.; Achenbach, J. E.; Veguilla, V.; Gubareva, L. V.; Garten, R.; Smith, C.; Yang, H.; Stevens, J.; Xu, X.; Katz, J. M.; Tumpey, T. M., Evaluation of the antigenic relatedness and cross-protective immunity of the neuraminidase between human influenza A (H1N1) virus and highly pathogenic avian influenza A (H5N1) virus. *Virology* **2014,** *454–455*, 169-175.

20.   Matrosovich, M. N.; Matrosovich, T. Y.; Gray, T.; Roberts, N. A.; Klenk, H.-D., Neuraminidase Is Important for the Initiation of Influenza Virus Infection in Human Airway Epithelium. *Journal of Virology* **2004,** *78* (22), 12665-12667.

21.   Sylte, M.; Suarez, D., Influenza Neuraminidase as a Vaccine Antigen. In *Vaccines for Pandemic Influenza*, Compans, R. W.; Orenstein, W. A., Eds. Springer Berlin Heidelberg: 2009; Vol. 333, pp 227-241.

22.   Russell, R. J.; Haire, L. F.; Stevens, D. J.; Collins, P. J.; Lin, Y. P.; Blackburn, G. M.; Hay, A. J.; Gamblin, S. J.; Skehel, J. J., The structure of H5N1 avian influenza neuraminidase suggests new opportunities for drug design. *Nature* **2006,** *443* (7107), 45-9.

23.   Li, Q.; Qi, J.; Zhang, W.; Vavricka, C. J.; Shi, Y.; Wei, J.; Feng, E.; Shen, J.; Chen, J.; Liu, D.; He, J.; Yan, J.; Liu, H.; Jiang, H.; Teng, M.; Li, X.; Gao, G. F., The 2009 pandemic H1N1 neuraminidase N1 lacks the 150-cavity in its active site. *Nature structural & molecular biology* **2010,** *17* (10), 1266-8.

24.   Woods, C. J.; Malaisree, M.; Long, B.; McIntosh-Smith, S.; Mulholland, A. J., Analysis and assay of oseltamivir-resistant mutants of influenza neuraminidase via direct observation of drug unbinding and rebinding in simulation. *Biochemistry* **2013,** *52* (45), 8150-64.

25.   Global monitoring of antiviral resistance in currently circulating human influenza viruses, November 2011. *Releve epidemiologique hebdomadaire / Section d'hygiene du Secretariat de la Societe des Nations = Weekly epidemiological record / Health Section of the Secretariat of the League of Nations* **2011,** *86* (45), 497-501.

26.   Meijer, A.; Rebelo-de-Andrade, H.; Correia, V.; Besselaar, T.; Drager-Dayal, R.; Fry, A.; Gregory, V.; Gubareva, L.; Kageyama, T.; Lackenby, A.; Lo, J.; Odagiri, T.; Pereyaslov, D.; Siqueira, M. M.; Takashita, E.; Tashiro, M.; Wang, D.; Wong, S.; Zhang, W.; Daniels, R. S.; Hurt, A. C., Global update on the susceptibility of human influenza viruses to neuraminidase inhibitors, 2012-2013. *Antiviral research* **2014,** *110*, 31-41.

27.   Karthick, V.; Shanthi, V.; Rajasekaran, R.; Ramanathan, K., Exploring the cause of oseltamivir resistance against mutant H274Y neuraminidase by molecular simulation approach. *Applied biochemistry and biotechnology* **2012,** *167* (2), 237-49.

28.   Smith, G. J. D.; Vijaykrishna, D.; Bahl, J.; Lycett, S. J.; Worobey, M.; Pybus, O. G.; Ma, S. K.; Cheung, C. L.; Raghwani, J.; Bhatt, S.; Peiris, J. S. M.; Guan, Y.; Rambaut, A., Origins and evolutionary genomics of the 2009 swine-origin H1N1 influenza A epidemic. *Nature* **2009,** *459* (7250), 1122-1125.

29.   Yarris, L., New Biological Route for Swine Flu to Human Infections.   20 October 2014, 2009.

30.   Mostaço-Guidolin, L. C.; Bowman, C. S.; Greer, A. L.; Fisman, D. N.; Moghadas, S. M., Transmissibility of the 2009 H1N1 pandemic in remote and isolated Canadian communities: a modelling study. *BMJ Open* **2012,** *2* (5).

31.   Danishuddin, M.; Khan, A. U., Structure based virtual screening to discover putative drug candidates: Necessary considerations and successful case studies. *Methods* **2015,** *71*, 135-145.

32.   Xia, J.; Tilahun, E. L.; Reid, T.-E.; Zhang, L.; Wang, X. S., Benchmarking methods and data sets for ligand enrichment assessment in virtual screening. *Methods* **2015,** *71*, 146-157.

33.   Vuorinen, A.; Schuster, D., Methods for generating and applying pharmacophore models as virtual screening filters and for bioactivity profiling. *Methods* **2015,** *71*, 113-134.

34.   Ke, Y.-Y.; Coumar, M. S.; Shiao, H.-Y.; Wang, W.-C.; Chen, C.-W.; Song, J.-S.; Chen, C.-H.; Lin, W.-H.; Wu, S.-H.; Hsu, J. T. A.; Chang, C.-M.; Hsieh, H.-P., Ligand efficiency based approach for efficient virtual screening of compound libraries. *European Journal of Medicinal Chemistry* **2014,** *83*, 226-235.

35.   Du, H.; Brender, J. R.; Zhang, J.; Zhang, Y., Protein structure prediction provides comparable performance to crystallographic structures in docking-based virtual screening. *Methods* **2015,** *71*, 77-84.

36.   Chen, Y.-C., Beware of docking! *Trends in Pharmacological Sciences* **2015,** *36* (2), 78-95.

37.   Kumar, A.; Zhang, K. Y. J., Hierarchical virtual screening approaches in small molecule drug discovery. *Methods* **2015,** *71*, 26-37.

38.   Kurauchi, R.; Watanabe, C.; Fukuzawa, K.; Tanaka, S., Novel type of virtual ligand screening on the basis of quantum-chemical calculations for protein–ligand complexes and

extended clustering techniques. *Computational and Theoretical Chemistry* **2015,** *1061*, 12-22.

39.   Amaro, R.; Baron, R.; McCammon, J. A., An improved relaxed complex scheme for receptor flexibility in computer-aided drug design. *Journal of computer-aided molecular design* **2008,** *22* (9), 693-705.

40.   Rastelli, G., Emerging Topics in Structure-Based Virtual Screening. *Pharm Res* **2013,** *30* (5), 1458-1463.

41.   Khanna, M.; Wang, F.; Jo, I.; Knabe, W. E.; Wilson, S. M.; Li, L.; Bum-Erdene, K.; Li, J.; G, W.   S.; Khanna, R.; Meroueh, S.   O., Targeting multiple conformations leads to small molecule inhibitors of the uPAR.uPA protein-protein interaction that block cancer cell invasion. *ACS Chem Biol* **2011,** *6* (11), 1232-43.

42.   Wang, L.; Gu, Q.; Zheng, X.; Ye, J.; Liu, Z.; Li, J.; Hu, X.; Hagler, A.; Xu, J., Discovery of New Selective Human Aldose Reductase Inhibitors through Virtual Screening Multiple Binding Pocket Conformations. *Journal of chemical information and modeling* **2013,** *53* (9), 2409-2422.

43.   Okimoto, N.; Futatsugi, N.; Fuji, H.; Suenaga, A.; Morimoto, G.; Yanai, R.; Ohno, Y.; Narumi, T.; Taiji, M., High-performance drug discovery: computational screening by combining docking and molecular dynamics simulations. *PLoS Comput Biol* **2009,** *5* (10), e1000528.

44.   Varghese, J. N.; McKimm-Breschkin, J. L.; Caldwell, J. B.; Kortt, A. A.; Colman, P. M., The structure of the complex between influenza virus neuraminidase and sialic acid, the viral receptor. *Proteins* **1992,** *14* (3), 327-32.

45.   Vavricka, C. J.; Li, Q.; Wu, Y.; Qi, J.; Wang, M.; Liu, Y.; Gao, F.; Liu, J.; Feng, E.; He, J.; Wang, J.; Liu, H.; Jiang, H.; Gao, G.   F., Structural and functional analysis of laninamivir and its octanoate prodrug reveals group specific mechanisms for influenza NA inhibition. *PLoS pathogens* **2011,** *7* (10), e1002249.

46.   Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E., UCSF Chimera—A visualization system for exploratory research and analysis. *Journal of Computational Chemistry* **2004,** *25* (13), 1605-1612.

47.   UCSF Resource for Biocomputing, V. a. I. *UCSF Chimera--an Extensible Molecular Modeling System*, University of California.

48.   Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., Comparison of simple potential functions for simulating liquid water. *The Journal of chemical physics* **1983,** *79* (2), 926.

49.   Mysinger, M. M.; Carchia, M.; Irwin, J. J.; Shoichet, B. K., Directory of Useful Decoys, Enhanced (DUD-E): Better Ligands and Decoys for Better Benchmarking. *Journal of Medicinal Chemistry* **2012,** *55* (14), 6582-6594.

50.   Liu, Y.; Jing, F.; Xu, Y.; Xie, Y.; Shi, F.; Fang, H.; Li, M.; Xu, W., Design, synthesis and biological activity of thiazolidine-4-carboxylic acid derivatives as novel influenza neuraminidase inhibitors. *Bioorganic & Medicinal Chemistry* **2011,** *19* (7), 2342-2348.

51.   Huang, S.-X.; Li, H.-Y.; Liu, J.-Y.; Morisseau, C.; Hammock, B. D.; Long, Y.-Q., Incorporation of Piperazino Functionality into 1,3-Disubstituted Urea as the Tertiary Pharmacophore Affording Potent Inhibitors of Soluble Epoxide Hydrolase with Improved Pharmacokinetic Properties. *Journal of Medicinal Chemistry* **2010,** *53* (23), 8376-8386.

52.   Todorovic, A.; Haskell-Luevano, C., *Peptides* **2005,** *26*, 2026.

53.   Kundu, B., *Curr. Opin. Drug Discovery Dev.* **2003,** *6*, 815.

54.   Kragholm, B.; Kvist, T.; Madsen, K. K.; Jørgensen, L.; Vogensen, S. B.; Schousboe, A.; Clausen, R. P.; Jensen, A. A.; Bräuner-Osborne, H., Discovery of a subtype selective inhibitor of the human betaine/GABA transporter 1 (BGT-1) with a non-competitive pharmacological profile. *Biochemical Pharmacology* **2013,** *86* (4), 521-528.

Chapter 7

A molecular dynamics description of the conformational binding of a potential influenza A endonuclease inhibitor

Ashona Singh[a‡], Sarentha Chetty[a] and M. E. S. Soliman[a,b*]

[a]School of Health Sciences, University of KwaZulu-Natal, Westville, Durban 4001, South Africa

[b]Department of Pharmaceutical Organic Chemistry, Faculty of Pharmacy, Zagazig University, Zagazig 44519, Egypt

[*]Corresponding author: Mahmoud E. S. Soliman, email: soliman@ukzn.ac.za

Telephone: +27 31 260 8048, Fax: +27 31 260 7872

[‡]Main contributor

Abstract

The threat of impending influenza pandemics, due to the rapid development of resistance of the influenza virus toward existing chemotherapeutic defences, expedites the need for innovative and effective drugs against selective targets. In this study, we provide a comprehensive molecular dynamics perspective of the bonding mode of a designed inhibitor in the active site of the 2009 pH1N1 endonuclease. Using computational techniques we were able to decipher the capacity of the active site by comparing the conformations of a bound enzyme to an apo-enzyme. Inhibitors of endonuclease require three features:

1) a greater number of hydrophilic moieties which are situated on terminal ends of the inhibitor,

2) long, flexible planar side-chains which are anchored to a hydrophobic or neutral base,

3) steric bulk, to block the binding pocket.

The structural features outlined will offer tremendous insight when constructing and developing future anti-viral agents.

Introduction

The influenza virus is an important human pathogen, being the primary cause of many epidemics and pandemics worldwide. There have been several documented pandemics, such as the '1918' *Spanish flu* which claimed approximately 40 million lives [1], in 1957 and 1968 the *Asian* and *Hong Kong* influenza viruses were reported each claiming about one million lives [2]. The most recent pandemic identified by the World Health Organisation (WHO) was the 2009 H1N1/swine flu, however, the virus still maintains its pathogenicity across the eastern border specifically in India. The unpredictable nature of H1N1 to gain resistance to current chemotherapies through rapid evolution and its ease of human-human transmissibility poses an eminent threat of a new pandemic, to which, the human population has no defence[3-5]. Thus it is crucial to understand the mechanism of resistance of influenza A viruses, as well continually research the development and design of new effective anti-viral agents.

Influenza vaccines are the most widely used prophylactic measure, however, due to the rapid spread of infection the need for effective antiviral drugs is imperative [6]. The influenza A virus genome is comprised of eight segments each being viable drug targets. These include: 1) the antigen haemagglutinin [7]; 2) the M2 ion channel which, has two main active inhibitors that are commercially available Symmetrel/amantadine and Flumadine/rimantadine [8]; and 3) neuraminidase which, has four main active drug candidates; Relenza/zanamivir, Tamiflu/oseltamivir, Rapivab/peramivir and laninamivir [9]. However, the rapid development of resistance of the influenza A virus toward available prophylaxis is of growing concern. Resistance is often acquired by mutation in the target protein, thus making our efforts to try and keep up with evolution more difficult. Therefore it is essential to explore and investigate alternate targets. An example would be to target the viral replication process.

The mechanics of viral replication lie in eight negative-stranded RNA segments which encode at least 17 viral proteins, a major component of these segments is the RNA-dependent RNA polymerase (RdRp) [10]. The crystal structure of the RNA polymerase reveals a large protein complex (~250 kDa) composed of three subunits, which are highly conserved among influenza A and B viruses; these include polymerase basic protein 1 and 2 (PB1 and PB2), and polymerase acidic protein (PA) [11, 12]. During viral transcription, the RdRp cleaves host pre-mRNA's at a distance of 10-13 nucleotides from their 5`-capped terminus [13]. The binding of the cap is performed by PB2, whilst the endonuclease activity resides in the N-terminal domain of $PA_N$ (PA-Nter; containing residues 1 to ~195). After endonuclease cleavage, the short 5`-capped RNA serves as primer for the initiation of viral mRNA synthesis, performed by the PB1 unit with viral mRNA's being subsequently translated by the host cell machinery [14]. The process of binding and cleavage of the 5`-cap *i.e.* 'cap snatching mechanism', is not a normal cellular function thus it remains an exclusive and essential step in the influenza virus lifecycle [13, 15].

There has been further evidence to suggest the existence of an alternate encoded protein, PA-X, which has an overlapping sequence and similar enzymatic activity as PA_N, may contribute to the host protein shutdown observed during influenza A infection [16]. Thus the acidic component of the polymerase offers a viable strategy to inhibit the influenza viral replication. Over the last 20 years several structurally diverse small molecule PA inhibitors have been discovered to have strong activity toward the PA-Nter enzyme. Such as flutimide and its structural analogues, tetrameric acids, green tea catechins, *N*-hydroxamic acid, epigallocatechin gallate, *N*-hydroxypyridazinones and hydroxypyri(mi)dinones. (most potent antiviral L-742,001) [17-20].

The catalytic site of endonuclease is similar to the (P)DX$_N$(D/E)XK nuclease motifs formed by D108, E119, a proline (influenza A) or alanine (influenza B) at position 107, and K134 or

K137[21]. It comprises a histidine (H41) and a cluster of three acidic residues (E80, D108, E119), which are conserved in all influenza viruses and coordinate (together with I120) to one, two or three divalent metal ions (Mg^{2+} or Mn^{2+}, with Mg^{2+} being the prominent cofactor *in vivo*) [22-24]. There are approximately 35 crystal structures related to the influenza virus PA endonuclease deposited in the RCSB Protein Data Bank, with over 20 enzymes in complex with an inhibitor. Multiple biochemical studies reveal critical information of the structural composition of the enzyme[17, 25]. The PAIs identified inhibit the PA endonuclease through chelation of its metal cofactor(s) with the active site. The architectural characterisation of PA_N allows the design and development of novel and rational PAI chelating structures and enhanced enzyme binding properties to improve antiviral activity in cell culture.

Bauman *et al*, created a new class of influenza endonuclease inhibitors *via* structure-based drug design[26]. It was identified that (compound 7's) had the greatest antiviral potency amongst the endonuclease inhibitors. In our study we wish to supply a molecular dynamic comparative perspective of the apo-enzyme of endonuclease as well as a bound ligand. This will provide significant structural binding markers that will allow us to *via in silico* screening identify an expanded class of endonuclease inhibitor. In turn, highlight critical features to be considered for *de novo* drug synthesis.

Methods

Preparation of System

Crystallographic structure of H1N1 endonuclease unbound (PDB ID: 4M5Q) and bound with inhibitor (compound 7) (PDB ID: 4M4Q), were obtained from the RCSB Protein Data Bank (PDB)[26]. Using the Chimera 1.8.1 software package a monomer was selected followed by the correction of protonation states of amino acid residues as well as the removal of water, polar

hydrogens and non-bonding species [27, 28]. The specific van der Waals radius and molecular mass of the Mn metal was accommodated for in mol2 format in preparation of the ligand.

Molecular Dynamics Simulation

Amber 14.0 GPU version of the PMEMD engine was used to perform a 100 ns molecular dynamics simulation. The protein was described with the FF99SB variant force field [29]. The LEaP module embedded within Amber was use for the addition of hydrogen atoms and neutralisation of the protein by the addition of Na^+ counter ions. The system was suspended within a TIP3P [303029] water box such that all protein atoms were within 8 Å from the box edge. The drug-protein complex was energetically minimised in three stages; 1) 1000 step restraint gradient minimisation with selective boundary conditions of the particle-mesh Ewald method which has a parameter of direct space and a van der Waals cut-off of 12 Å. Stage 2, a 2500 of deepest descent step with a solute harmonic potential of 500 kcal/mol. $Å^2$; and stage 3, a 1000 step unrestrained conjugate gradient energy minimisation of the complete system. The system was heated using a canonical ensemble (NVT) from 0 K to 300 K over 50 ps. The collision frequency of solutes was 1.0 ps^{-1} with a harmonic potential of 10 kcal/mol.$Å^2$. A final equilibration step of 500 ps was performed using the SPFP precision model of the SHAKE algorithm to constrict hydrogen bonds at 2 fs intervals. The simulations were conducted in a randomised seeding of an isobaric-isothermal ensemble (NPT). A constant pressure of 1 bar was maintained by the Berendsen barostat with a pressure coupling constant of 2 ps. A temperature of 300 K was set using a Langevin thermostat with collision frequency of 1.0 ps^{-2}. Co-ordinates were saved every 1 ps and the trajectories were analysed every 1 ps using a CPPTRAJ module in Amber 14.0.

Binding free energy calculation

An implicit solvent model was employed to describe the thermodynamic free binding energies of the bound conformation of endonuclease enzyme. The Molecular Mechanics/Generalised-Born Surface area method was used to evaluate the ligand-protein complex binding affinities.[31-34] To calculate the free binding energy contributions 10000 snapshots were extracted from the 100 ns trajectory. The following set of equations describe the calculation of the binding free energy:

$$\Delta G_{bind} = G_{complex} - G_{receptor} - G_{ligand} \qquad (1)$$

$$\Delta G_{bind} = E_{gas} + G_{sol} - TS \qquad (2)$$

$$E_{gas} = E_{int} + E_{vdw} + E_{ele} \qquad (3)$$

$$G_{sol} = G_{GB} + G_{SA} \qquad (4)$$

$$G_{SA} = \gamma SASA \qquad (5)$$

The term E_{gas} represents the gas-phase energy, and is directly estimated from the FF99SB force field terms which comprises the energy contributor's internal energy (E_{int}), van der Waals energy (E_{vdW}) and Coulomb energy (E_{ele}). The solvation energy, G_{sol} is estimated by the contribution of energies from polar states, G_{GB} and non-polar states, G_{SA}. The non-polar solvation energy is determined from the solvent accessible surface area (SASA) using a water probe of 1.4Å. The temperature and total solute entropy are represented by T and S, respectively[35].

Principle Component Analysis (PCA)

PCA reveals the structure of atomic fluctuations. PCA describes the motion of the system in terms of eigenvectors (planar of motion) and eigenvalue (magnitude of motion) [36].

The individual MD trajectories are stripped of solvent and ions using the PTRAJ and CPPTRAJ modules in Amber 12.0/14.0. The resulting trajectories were aligned against a fully minimised structure. PCA was performed on $C\alpha$ backbone with 5000 snapshots taken every 20 ps. The first two Eigen vectors (PC1 and PC2) corresponding to the first two modes of PCA covariance matrices were generated using in-house scripts. The corresponding PCA scatters were plotted using Origin software (http://www.originlab.com/), structural postscript diagrams were created using VMD [37]. Porcupine plots of the first and second modes developed by the normal mode wizard (NMW) using the ProDy interface of VMD were sketched for each of the systems [38].

Results and Discussion

The crystal structure of the apo- influenza endonuclease (4M5Q) and ligand-bound (4M4Q) enzymes contain a residue complement of 196 and 200 amino acids, respectively. The difference in the amino acid sequence lies in the *N*-terminus of the enzyme and is dependent on the resolution of the crystal structure. The difference of four amino acids does not alter the secondary or tertiary structures of the metalloprotein as is observed from the secondary structure analysis of the systems (see Supplementary information **S1** to **S5**). The root mean square deviation (RMSD) and potential energy plots attained convergence in both systems at 10 ns for the apo-enzyme and 25 ns for the ligand-bound PA. With a measured average RMSD of 1.5 Å and 1.2 Å, respectively. The energy difference between the systems estimates 10 000 kcal/mol in favour of the ligand-enzyme complex.

The intention of this study is to highlight the critical binding features of the unique inhibitor. Each PA_N, comprises two divalent Mn metal centers which are essential to the binding of the inhibitor in the active site. The binding mode of the highly potent molecule is unlike the

binding of other endonuclease inhibitors in the active site. The metal cofactor exists as +2 cation, which, governs the architecture of the potential chemotherapeutic.

According to the online protein-ligand interaction profiler (PLIP) [39], PA_N has an alpha-beta construction with five-mixed beta-strands forming a twisted plane surrounded by seven alpha-helices. A negatively charged cavity surrounded by helices α2-5 and strand β3 house the metal ion. The geometry of metal co-ordination is square planar and square pyramidal. Metal chelation occurs *via* amino acid interactions: Ile120, acidic residues Glu80, Asp108, three water molecules stabilised by residue H41, Glu119 and the carbonyl oxygens of Lys106 and Pro107. The co-ordinating residues have been found to be highly conserved between influenza A, B and C viruses.

The root mean square fluctuation no observable perturbations between the apo-enzyme and the ligand-metalloprotein complex of the residues implicated in the binding of the metal (Figure 1). Residues 50 to 75 of both systems, which, are assembled in a coil exert a high frequency. The ligand of system 4M4Q expresses two prominent hydrophobic interactions of the α-helix. The first being between with β-carbon of Ala42 and C_{10} of the ligand cyclohexyl group at a distance of 3.79 Å. The second non-polar interface occurs between the γ-carbons (*i.e.* $C_{γ1}$ and $C_{γ2}$) of Ile43 and C_9 and C_{26} of the cycloheyl (3.78 Å) and fluorocyclohexenyl (3.69 Å) groups of the ligand, respectively. A hydrogen bond network exists between the ligand and catalytic residues. These include interactions of hydrogen bond donors of the protein; ε-amino group of Lys39 (H-acceptor distance = 3.01 Å, donor-acceptor distance = 3.97 Å) and Lys139 (H-acceptor distance = 1.80 Å, donor-acceptor distance = 2.58 Å), as well as the δ-guanidino group of Arg129 (H-acceptor distance = 2.77 Å, donor-acceptor distance = 3.76 Å) with N_{14} and N_{15} of the aromatic pyran ring and C_5 *para* to the nitrogen of pyridine ring. The ε-O of the carboxyl group of residue Glu85 and the hydroxyl group of Tyr135 accept hydrogens from the ligand *via* C_{20} of the fluorocyclohexenyl group (H-acceptor distance = 2.27 Å, donor-acceptor

distance = 3.01 Å) and C_5 of the pyridine ring (H-acceptor distance = 3.06 Å, donor-acceptor distance = 3.65 Å).
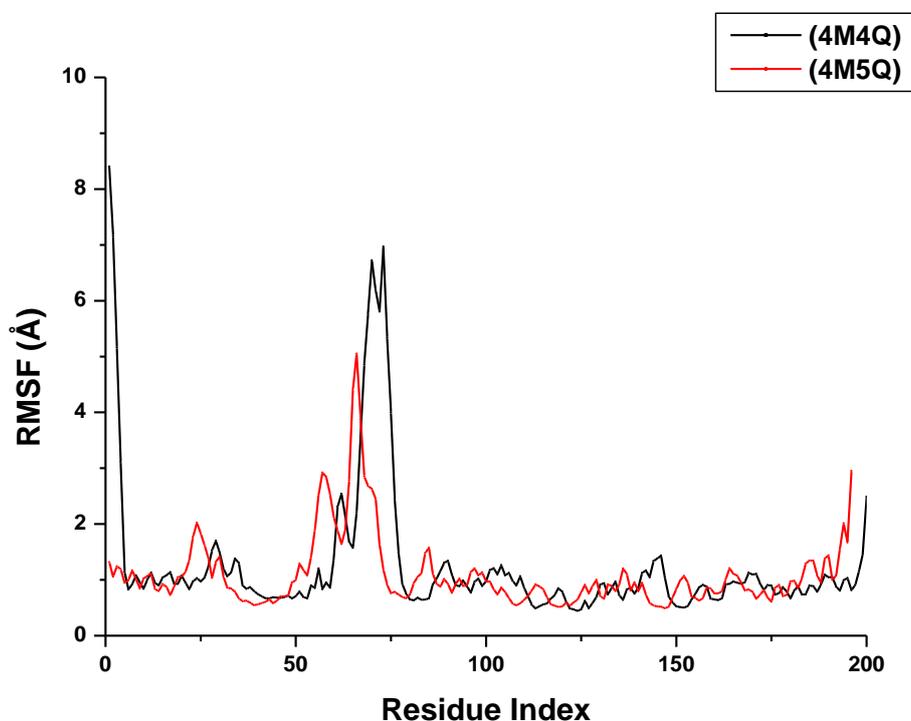

Figure 1: RMSF of the bound and unbound system of endonuclease enzyme

An estimation of the number of hydrogen bonds throughout the simulation affirmed that the apo-enzyme (4M5Q) and ligand-protein complex (4M4Q) were similar averaging between 90 − 130 hydrogen bonds (Figure 2). However, it was observed that the molecular dynamics of the apo-enzyme predisposes the enzyme to greater solvent exposure which leads to a 4-5 hydrogen bond difference of apo-enzyme and complex systems at any given point during the simulation. This is exacerbated in the apo-enzyme due to the high concentration of histidine residues which readily interact with adjacent residues in the α-helices and β-strands, as well as the metal centers. Binding of the ligand in the active site pocket is mediated by the divalent metal cations and specific hydrophobic interactions. Thus the number of hydrogen bonds when comparing the bound and unbound systems of the PA subunit remain essentially the same.
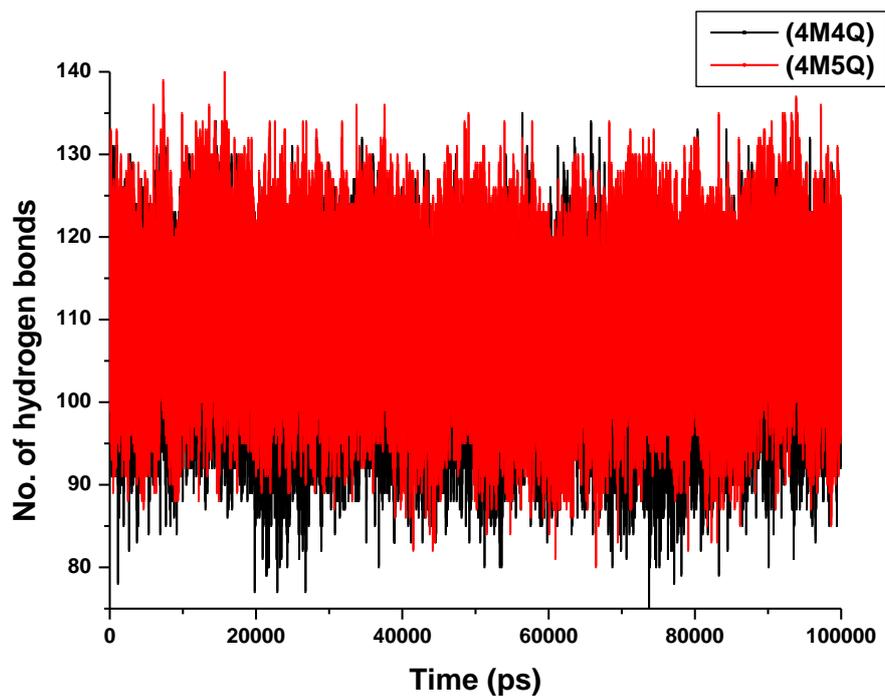
Figure 2: Graphical estimation of the number of hydrogen bonds in systems 4M4Q and 4M5Q
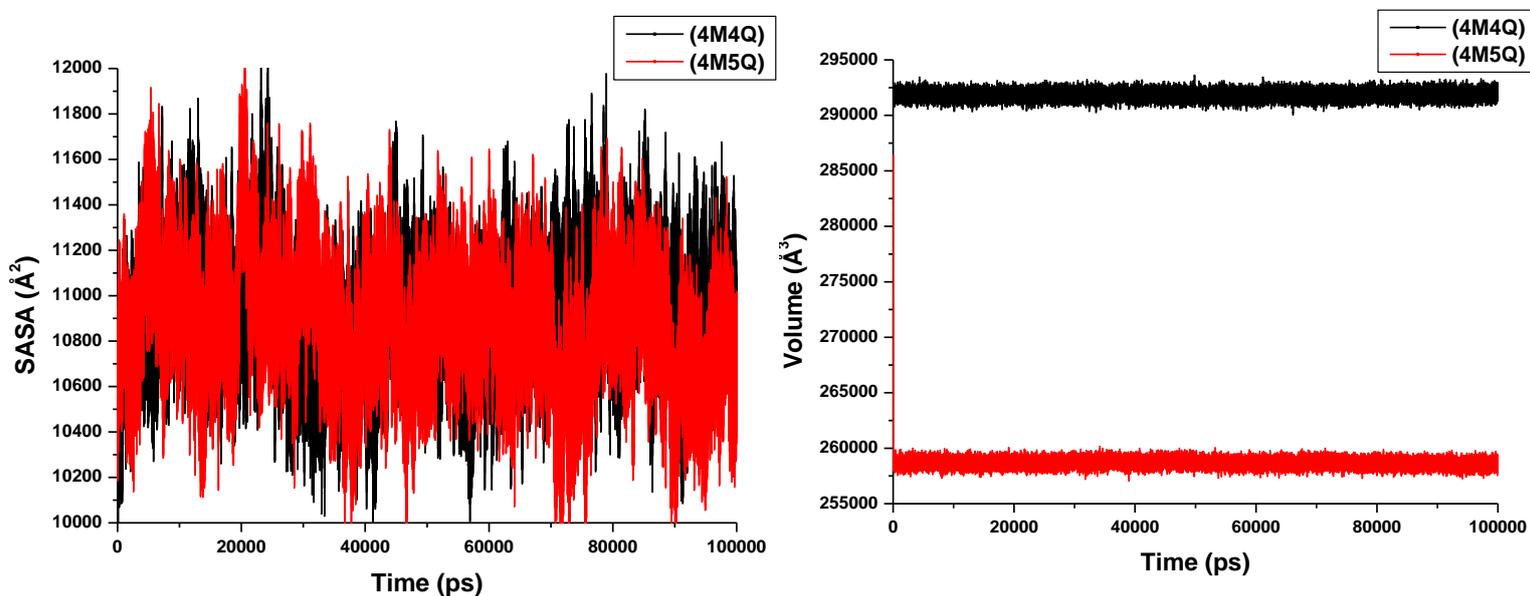

Figure 3: Solvent accessible surface area (left) and volume (right) analysis of the proteins bound (4M4Q) and unbound (4M5Q) over 100ns

Solvent accessible surface area (SASA) refers to the relative area exposed to solvent taking into consideration, the composition of van der Waals surface of the protein. The SASA for the

ligand-protein complex, 4M4Q is slightly larger that of the apo-enzyme, 4M5Q (Figure 3). The Mn^{+2} metal centers in their unbound state are co-ordinated to three waters. In the bound state the waters are extruded to accommodate the ligand. The structural moieties of the ligand are accommodated in the active site, the structure displays a significant number of hydrophilic pharmacophoric features. The metal-ligand interaction co-ordination geometry is by a square planar arrangement. The molecular size of the ligand is much larger than that of the three water molecules combined, which causes the binding cavity to expand. The overall volume of 4M4Q the system increases significantly (Figure 3). This distortion in conformation may be key to the inhibition of the endonuclease activity of the PA subunit.

The radius of gyration describes the dynamic motion of the amino acid residues bound in their secondary structure in relation to the α-carbon backbone. It indirectly quantifies the amount of energy required to overcome inertia. System comparison of 4M5Q and 4M4Q reveal that the apo-enzyme is more tightly bound than the ligand-metalloprotein complex (Figure 4). System 4M4Q achieves a stabilised conformation only after ~40 ns at which point the protein complex is least compact. This phenomenon is in line with the multiple biological functioning of the RNA-dependent enzyme. PA_N has been implicated in cap-binding, endonuclease activity, viral RNA binding and replication. The volume of the ligand-protein complex is considerably larger than the apo-enzyme. This proposes a constricted and relaxed conformation of the enzyme, where when selectively bound to the inhibitor the overall structure of the protein is more relaxed which increases the volume of the enzyme.
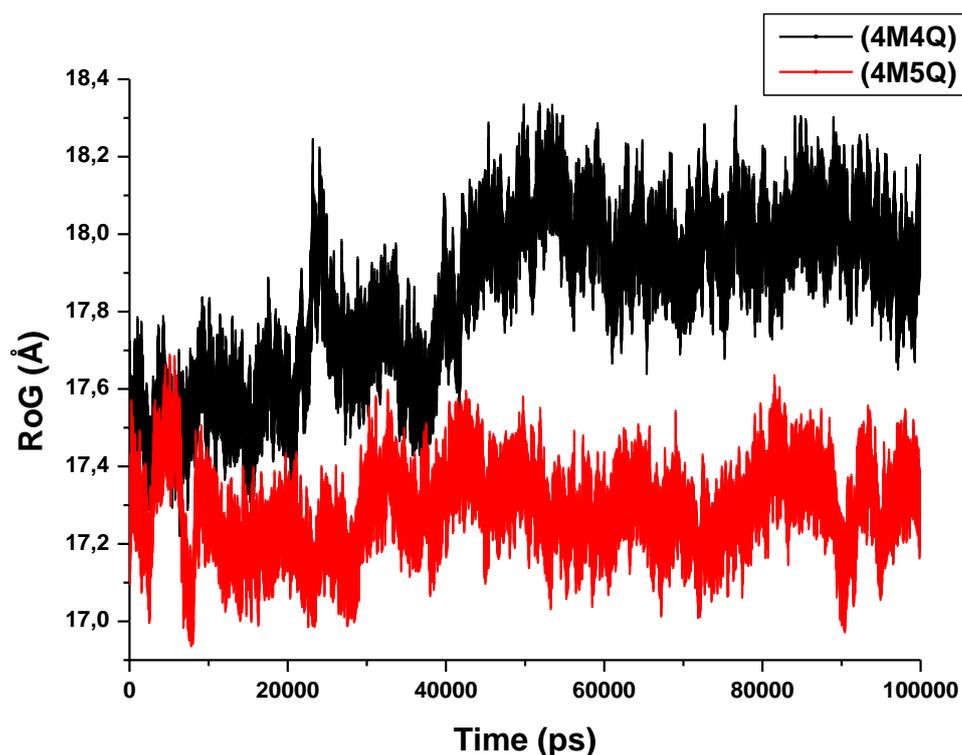
Figure 4: Radius of gyration of systems 4M4Q and 4M5Q over 100ns

Principal component analysis explores the dynamics of protein folding in terms of the diffusive properties of the polypeptide chain. System 4M5Q has a more compact three-dimensional structure than the ligand-protein complex, 4M4Q (Figure 5). With a correlation of -0.011 which is close to zero indicating that principal component 1 (PC 1) and principal component 2 (PC2) are independent of one another and the direction of motion is antithetic. The degree of motion although inversely proportional is also disproportionate in either direction. This is supported by a covariance value of -0.62. The correlation value of the bound complex 4M4Q, measured 0.0091 and the covariance value estimated 1.17. This indicates that the fluctuation of the α-carbon backbone of PC1 and PC2 is independent of one another and occurs in a uniform disproportionate direction.
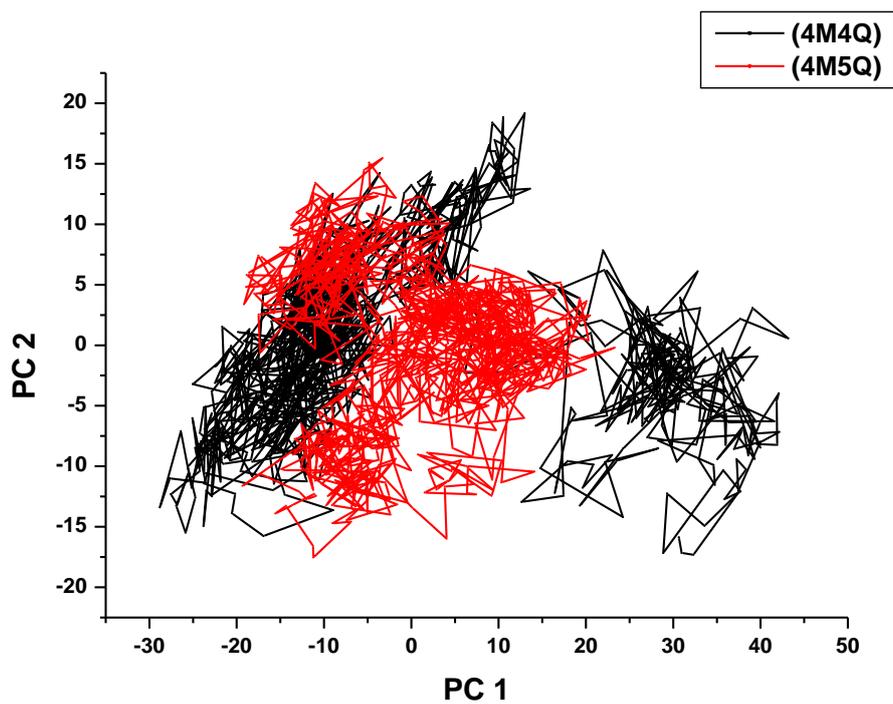
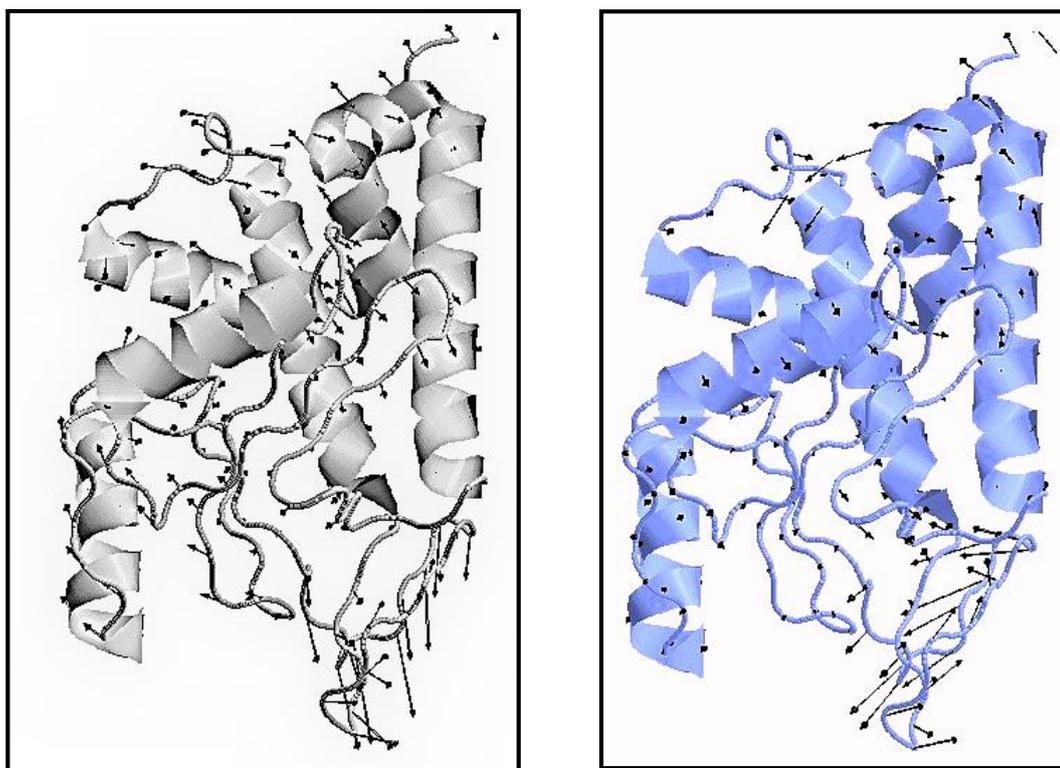**Figure 5**: PCA scatters of systems 4M4Q and 4M5Q averaged over 100 ns

Figure 6: Normal mode analysis of system 4M5Q (left) and 4M4Q (right)

Figure 6, describes the elastic motion of the alpha-carbon backbone of the PA subunit for both the 4M4Q and 4M5Q systems. The normal mode analysis demonstrates a greater disorder in the bound conformation of 4M4Q eigenvector to the order ~0.001.

Using MM/GBSA calculation it was found that the binding free energy of the bound conformation revealed significant contribution from the E_{vdw} of -29.64 (kcal/mol) to the total ΔG measuring -20.54 (kcal/mol). Van der Waals interactions are non-bonding entities and occur largely by the aggregation of non-polar residues forming hydrophobic pockets. The contribution from electrostatics and solvent are included in table 1.

Table 1: Energy contributions of the binding free energy term (ΔG)

Energy component	Energy contribution/ (kcal/mol)
E_{ele}	-6.70 +/- 8.12
G_{GB}	19.68 +/- 7.43
E_{gas}	-36.35 +/- 8.46
G_{SA}	-3.88 +/- 0.46
G_{sol}	15.81 +/- 7.23

The quasi-harmonic approximation of entropy (ΔS) describes the disorder in the protein sequence which impacts the folding of the protein. The composites of ΔS, include the effects of translation, rotation and vibration. In the bound conformation, system 4M4Q, these were estimated to be -12.93, -10.68 and -21.01 (kcal/mol) respectively. The overall measured disorder was -44.62 (kcal/mol).

Conclusion

The molecular dynamics analysis of bound and unbound endonuclease enzyme provided insight in the conformational change of the protein. Subtle changes in the active site are amplified in the overall protein folding structure. Using molecular dynamics simulation we were able to observe the effect of the inhibitor binding on the overall PA subunit. It is clear that in order for a molecule to inhibit or bind to the active site components and chelate to metal, the inhibitor would have to exhibit an inherently hydrophilic nature. Hydrophilic features would also need to be adaptive, in that the functional moiety would need to be ionisable. The co-ordination of the ligand to the metal centers is required to be resilient against any fluctuations in pH. The molecule should have two adjustable "arms" or side-chains in order to chelate to the metal center in a square planar geometry.

The study offered insight into the expandable limitations of the PA subunit. Thus future design of inhibitors may be constructed by active binding feature as well as steric bulk to block the active site.

References

[1] Taubenberger, J. K., and Morens, D. M. (2006) 1918 Influenza: the Mother of All Pandemics, *Emerging Infectious Diseases 12*.

[2] Masurel, N., and Marine, W. M. (1973) Recycling of Asian and Hong Kong influenza A virus hemagglutinins in man, *American journal of epidemiology 97*, 44-49.

[3] Smith, G. J. D., Vijaykrishna, D., Bahl, J., Lycett, S. J., Worobey, M., Pybus, O. G., Ma, S. K., Cheung, C. L., Raghwani, J., Bhatt, S., Peiris, J. S. M., Guan, Y., and Rambaut, A. (2009) Origins and evolutionary genomics of the 2009 swine-origin H1N1 influenza A epidemic, *Nature 459*, 1122-1125.

[4] Yarris, L. (2009) New Biological Route for Swine Flu to Human Infections, Berkeley Lab- University of California, California.

[5] Mostaço-Guidolin, L. C., Bowman, C. S., Greer, A. L., Fisman, D. N., and Moghadas, S. M. (2012) Transmissibility of the 2009 H1N1 pandemic in remote and isolated Canadian communities: a modelling study, *BMJ Open 2*.

[6] Prevention, C. f. D. C. a. (2015) Flublok Seasonal Influenza (Flu) Vaccine, In *Types of flu vaccine* (Prevention, C. f. D. C. a., Ed.).

[7] Fleury, D., Wharton, S. A., Skehel, J. J., Knossow, M., and Bizebard, T. (1998) Antigen distortion allows influenza virus to escape neutralization, *Nature structural biology 5*, 119-123.

[8] Jefferson, T., Deeks, J. J., Demicheli, V., Rivetti, D., and Rudin, M. (2004) Amantadine and rimantadine for preventing and treating influenza A in adults, *The Cochrane database of systematic reviews*, Cd001169.

[9] Administration, U. S. F. a. D. Influenza (Flu) Antiviral Drugs and Related Information, In *Drug Safety and Availability* (Administration, F. U. S. F. a. D., Ed.), US.

[10] Portela, A., and Digard, P. (2002) The influenza virus nucleoprotein: a multifunctional RNA-binding protein pivotal to virus replication, *J Gen Virol 83*, 723-734.

[11] Yuan, P., Bartlam, M., Lou, Z., Chen, S., Zhou, J., He, X., Lv, Z., Ge, R., Li, X., Deng, T., Fodor, E., Rao, Z., and Liu, Y. (2009) Crystal structure of an avian influenza polymerase PAN reveals an endonuclease active site, *Nature 458*, 909-913.

[12] Leahy, M. B., Dobbyn, H. C., and Brownlee, G. G. (2001) Hairpin loop structure in the 3' arm of the influenza A virus virion RNA promoter is required for endonuclease activity, *J Virol 75*, 7042-7049.

[13] Boivin, S., Cusack, S., Ruigrok, R. W. H., and Hart, D. J. (2010) Influenza A Virus Polymerase: Structural Insights into Replication and Host Adaptation Mechanisms, *Journal of Biological Chemistry 285*, 28411-28417.

[14] David B. Olsen, F. B., James L. Cole, Mark W. Stahlhut, Robert E. Dempski, Paul L. Darke, and Lawrence C. Kuo. (1996) Elucidation of Basic Mechanistic and Kinetic Properties of Influenza Endonuclease Using Chemically Synthesized RNAs, *THE JOURNAL OF BIOLOGICAL CHEMISTRY 271*, 7435-7439.

[15] Dias, A., Bouvier, D., Crepin, T., McCarthy, A. A., Hart, D. J., Baudin, F., Cusack, S., and Ruigrok, R. W. (2009) The cap-snatching endonuclease of influenza virus polymerase resides in the PA subunit, *Nature 458*, 914-918.

[16] Leahy, M. B., Pritlove, D. C., Poon, L. L., and Brownlee, G. G. (2001) Mutagenic analysis of the 5' arm of the influenza A virus virion RNA promoter defines the sequence requirements for endonuclease activity, *J Virol 75*, 134-142.

[17] DuBois, R. M., Slavish, P. J., Baughman, B. M., Yun, M.-K., Bao, J., Webby, R. J., Webb, T. R., and White, S. W. (2012) Structural and Biochemical Basis for Development of Influenza Virus Inhibitors Targeting the PA Endonuclease, *PLoS pathogens 8*, e1002830.

[18] Parkes, K. E. (2003) Use of a pharmacophore model to discover a new class of influenza endonuclease inhibitors, *J. Med. Chem. 46*, 1153-1164.

[19] Yamada, K., Koyama, H., Hagiwara, K., Ueda, A., Sasaki, Y., Kanesashi, S. N., Ueno, R., Nakamura, H. K., Kuwata, K., Shimizu, K., Suzuki, M., and Aida, Y. (2012) Identification of a novel compound with antiviral activity against influenza A virus depending on PA subunit of viral RNA polymerase, *Microbes Infect 14*, 740-747.

[20] Tomassini, J. E. (1996) A novel antiviral agent which inhibits the endonuclease of influenza viruses, *Antimicrob. Agents Chemother. 40*, 1189-1193.

[21] Zhao, C., Lou, Z., Guo, Y., Ma, M., Chen, Y., Liang, S., Zhang, L., Chen, S., Li, X., Liu, Y., Bartlam, M., and Rao, Z. (2009) Nucleoside monophosphate complex structures of the endonuclease domain from the influenza virus polymerase PA subunit reveal the substrate binding site inside the catalytic center, *J Virol 83*, 9024-9030.

[22] Xiao, S., Klein, M. L., LeBard, D. N., Levine, B. G., Liang, H., MacDermaid, C. M., and Alfonso-Prieto, M. (2014) Magnesium-dependent RNA binding to the PA endonuclease domain of the avian influenza polymerase, *J Phys Chem B 118*, 873-889.

[23] Doan, L., Handa, B., Roberts, N. A., and Klumpp, K. (1999) Metal ion catalysis of RNA cleavage by the influenza virus endonuclease, *Biochemistry 38*, 5612-5619.

[24] Kowalinski, E., Zubieta, C., Wolkerstorfer, A., Szolar, O. H. J., Ruigrok, R. W. H., and Cusack, S. (2012) Structural Analysis of Specific Metal Chelating Inhibitor Binding to the Endonuclease Domain of Influenza pH1N1 (2009) Polymerase, *PLoS pathogens 8*, e1002831.

[25] Noble, E., Cox, A., Deval, J., and Kim, B. (2012) Endonuclease substrate selectivity characterized with full-length PA of influenza A virus polymerase, *Virology 433*, 27-34.

[26] Bauman, J. D., Patel, D., Baker, S. F., Vijayan, R. S., Xiang, A., Parhi, A. K., Martinez-Sobrido, L., LaVoie, E. J., Das, K., and Arnold, E. (2013) Crystallographic fragment screening and structure-based optimization yields a new class of influenza endonuclease inhibitors, *ACS Chem Biol 8*, 2501-2508.

[27] UCSF Resource for Biocomputing, V. a. I. UCSF Chimera--an Extensible Molecular Modeling System, (EF, P., TD, G., CC, H., GS, C., DM, G., EC, M., and TE, F., Eds.), University of California.

[28] Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004) UCSF Chimera—A visualization system for exploratory research and analysis, *Journal of Computational Chemistry 25*, 1605-1612.

[29] Ponder, J. W., and Case, D. A. (2003) Force Fields for Protein Simulations, In *Advances in Protein Chemistry*, pp 27-85, Academic Press.

[30] Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983) Comparison of simple potential functions for simulating liquid water, *The Journal of chemical physics 79*, 926.

[31] Sun, H., Li, Y., Tian, S., Xu, L., and Hou, T. (2014) Assessing the performance of MM/PBSA and MM/GBSA methods. 4. Accuracies of MM/PBSA and MM/GBSA methodologies evaluated by various simulation protocols using PDBbind data set, *Physical Chemistry Chemical Physics 16*, 16719-16729.

[32] Greenidge, P. A., Kramer, C., Mozziconacci, J. C., and Wolf, R. M. (2013) MM/GBSA binding energy prediction on the PDBbind data set: successes, failures, and directions for further improvement, *Journal of chemical information and modeling 53*, 201-209.

[33] Godschalk, F., Genheden, S., Soderhjelm, P., and Ryde, U. (2013) Comparison of MM/GBSA calculations based on explicit and implicit solvent simulations, *Physical chemistry chemical physics : PCCP 15*, 7731-7739.

[34] Mitomo, D., Fukunishi, Y., Higo, J., and Nakamura, H. (2009) Calculation of protein-ligand binding free energy using smooth reaction path generation (SRPG) method; A comparison of the explicit water model, GB/SA model and docking score function, *Genome Information 23*, 85-97.

[35] Hou, T., Wang, J., Li, Y., and Wang, W. (2011) Assessing the performance of the MM/PBSA and MM/GBSA methods. 1. The accuracy of binding free energy calculations based on molecular dynamics simulations, *Journal of chemical information and modeling 51*, 69-82.

[36] Shlens, J. (2005) A tutorial on Principal Component Analysis, pp 1-13, University of California, San Diego.

[37] Humphrey, W., Dalke, A., and Schulten, K. (1996) VMD: visual molecular dynamics, *J Mol Graph 14*, 33-38, 27-38.

[38] Bakan, A., Meireles, L. M., and Bahar, I. (2011) ProDy: Protein Dynamics Inferred from Theory and Experiments, *Bioinformatics 27*, 1575-1577.

[39] Salentin, S., Schreiber, S., Haupt, V. J., Adasme, M. F., and Schroeder, M. (2015) PLIP: fully automated protein–ligand interaction profiler, *Nucleic Acids Research*.

Chapter 8

8. General conclusions

8.1 Conclusion

The study had three distinct aims:

4. To determine, differentiate and distinguish the mechanism of resistance of neuraminidase in influenza A; 2009-pandemic H1N1 (2009-*p*H1N1) and H5N1 subtypes in the presence of mutation/s H274Y and I222K.

5. To propose potential neuraminidase scaffolds for new and innovative prophylactic chemotherapies *via* the aid of *in silico* technique ligand-based virtual screening (LB-VS) and structure-based virtual screening (SB-VS).

6. To identify the binding mode of the polymerase endonuclease of H1N1 and isolate potential drug therapies to selectively inhibit enzyme activity, using LB-VS and SB-VS.

Overall, each aim was achieved. Molecular dynamics proved an invaluable tool, specifically the multiple MD simulations. The enriched MD simulation offered improved resolution of the protein conformation and three-dimensional structure.

The comparative study of the wild type and mutant species of neuraminidase of the 2009-*p*H1N1 and the highly pathogenic H5N1 strains, described a change in the active site of the enzyme. It was found that a single point mutation of H274Y compared to the point mutation I222K influenced the structure of the active site to a greater extent by lowering the binding affinity of the drug, oseltamivir. The binding affinity of the drug was further reduced in the presence of the double mutation, *i.e.* H274Y-I222K. However, the 2009 pandemic swine flu-H1N1 exhibited greater susceptibility to oseltamivir than the highly pathogenic avian flu-H5N1. This urged a closer inspection of the different neuraminidase binding landscapes. Evidence suggested a structural collapse of the active site of the wild type influenza A virus in the mutant species. The contributions to the binding free energy term, the solvent accessible surface area analyses and radius of gyration analyses proposed a loss in hydrophobicity in the systems. However, the volume ratio of the active site and overall compactness of the ligand-protein complex of H5N1 mutants and mutants of H1N1 when compared to their wild type structures, was significantly altered. Furthermore, when comparing the H5N1 systems to those

of the H1N1, the H5N1 ligand-protein complexes were found to exhibit a greater folding capacity which ultimately led to decreased surface solvent exposure.

The difference in the binding site of the influenza viruses offers an explanation into the discrepancy in the mechanism and extent of resistance in the H1N1 and H5N1 systems. However, this only reiterates the need for new and innovative anti-viral agents. Using ligand-based and structure-based pharmacophore techniques, in the construction of three-dimensional models a set of 10 discriminant compound scaffolds was isolated. These were preferentially screened for their ability to bind to the transmutated active site of the H1N1 mutant species. The *in silico* investigation of the H1N1 species containing the mutations H274Y and/or I222K, revealed pertinent information about the pharmacophoric features of potential inhibitors. With the binding site volume being larger in the $H274Y_{H1N1}$ and $I222K_{H1N1}$ systems than in the WT_{H1N1}, it could accommodate:

1) a larger profile inhibitor;
2) an inhibitor with increased hydrophilic property
3) structurally the inhibitor should have a longer hydrophobic 'tail' to be accommodate in the binding region which held the bulky group of oseltamivir.

The binding mode study of endonuclease revealed that co-ordination of two divalent metal centers to a potential inhibitor, is required for the selective impair enzyme activity. The binding of the inhibitor in the active site of endonuclease is indirect. The interaction profile of the ligand-to-protein comprises:

1) salt bridges;
2) water bridges;
3) metal chelation; and
4) electrostatic interactions predominantly with Arg, Asp and Glu residues.

The designed drug should exhibit hydrophilic properties and functional moieties capable of ionising at a pH range of $5-7$ whilst retaining its stability. The binding affinity of the inhibitor to the imperative active site residues and metal ions should improve with the drop in pH.

The therapeutic targets neuraminidase and endonuclease, are exceptional focal points in the effort to treat influenza A viral infections. The evolutionary capability of neuraminidase remains a conundrum, which requires constant vigilance and monitoring to prevent future pandemics. Endonuclease inhibition is an ideal target being responsible for initiating the 'cap-

snatching' mechanism in protein synthesis. Furthermore, the enzyme shares a conserved protein sequence identity in the different types and subtypes of influenza viruses. Thus potential inhibitors would have a broad spectrum anti-viral activity.

8.2 Future work

The World Health Organisation has been on high alert since the 2006 outbreak of the highly transmissible and infectious H5N1, avian flu. The virus stoked the fears of the world as natural evolution of the strain overcame cross-species barriers, infecting both birds and humans alike with detrimental consequences. The realisation of the human population being vulnerable to such infections heightened the urgency for new chemoprophylaxis [1].

In order for this pre-emptive effort to be successful, researchers must be able to predict possible antigenic variations of influenza strains. The drawback of this technique lies in the fact that not all highly resistant, highly transmissible and highly virulent strains of the influenza virus have been expressed. Drug design relevant to these potential strains seems a never ending game of "cat and mouse". This study has accentuated the reliability, usefulness and dynamic capacity that computational techniques lend to the pharmaceutical, medicinal, biological and biochemical industries.

Molecular dynamics simulations, binding free energy calculations and other post-dynamic analyses have demonstrated their ability to provide insurmountable evidence in explaining the resistance mechanisms of complex systems [2]. Three-dimensional detail of protein structures, such as protein folding can be elucidated. Further to this, the structural and conformational effects mutations have on protein complexes can also be determined [3]. The potential success, effectiveness and design of drugs based on prerequisite structure-based or ligand-based pharmacophore anomalies, can timeously and easily be interpreted. Going forward, computational studies would be an indispensable tool in reducing resources by:

1) understanding the mode of action and resistance mechanism of critical enzyme targets;
2) acting as a preface to screen the design and efficacy of drugs prior to synthesis;
3) providing an atomistic perspective of the structural and conformational detail of proteins bound or unbound; and
4) being used as a reliable and accurate comparative tool.

With regard to the systems evaluated in this thesis, a purely computational perspective was used. Further investigation into the synthesis of the scaffold of compounds isolated from the *in silico* studies of potential drug candidates of neuraminidase can be performed. Enhancement and pharmacophore expansion of the scaffolds into a family of compounds may lead to the discovery of more potent chemotherapeutic agents of influenza.

Endonuclease inhibitors, as mentioned earlier have exhibited duality in their therapeutic activity as anti-cancer agents. Comparative research in:

1) understanding the mode of action; and
2) structural similarities, of endonuclease and the therapeutic targets of cancer would offer exceptional insight into the behaviour of the proteins and provide possible pharmacophoric features in the design and development of new drug candidates [4].

Further research should be invested in targeting hemagglutinin and the M2 ion channels. The computational software tool, Discovery Studio maybe used to model the glycoprotein structure and proton channel [5]. In doing so, the mode of action of each protein may be described and a possible mechanism of inhibition may be suggested. The software is further capable of designing and screening potential drug candidates which would adhere to the pharmacophore outlined from determining the mechanism of inhibition. ESI/Q-tof hybrid techniques can be employed to test the design of the M2 ion channel inhibitors. This is achieved by constructing a bilipid layer resembling the endosome, with an induced potential difference across the bilayer to mimic the M2 ion channel [6].

8.3 References

[1] Organisation, W. H. (2015) Antimicrobial resistance, World Health Organisation.

[2] Karplus, M., and McCammon, J. A. (2002) Molecular dynamics simulations of biomolecules, *Nature 9*, 646.

[3] Ooms, F. (2000) Molecular modeling and computer aided drug design. Examples of their applications in medicinal chemistry, *Current medicinal chemistry 7*, 141-158.

[4] Kaufmann, S. H. (1989) Induction of Endonucleolytic DNA Cleavage in Human Acute Myelogenous Leukemia Cells by Etoposide, Camptothecin, and Other Cytotoxic Anticancer Drugs: A Cautionary Note, *Cancer Research 49*, 5870-5878.

[5] BIOVIA, D. S. (2015) Discovery Studio Modeling Environment, 4.5 ed., Dassault Systemes, San Diego.

[6] Demmers, J. A. A., Haverkamp, J., Heck, A. J. R., Koeppe, R. E., and Killian, J. A. (2000) Electrospray ionization mass spectrometry as a tool to analyze hydrogen/deuterium exchange kinetics of transmembrane peptides in lipid bilayers, *Proceedings of the National Academy of Sciences 97*, 3189-3194.