SIZE DOES NOT MATTER: A MOLECULAR INSIGHT INTO THE BIOLOGICAL ACTIVITY OF CHEMICAL FRAGMENTS UTILIZING COMPUTATIONAL APPROACHES

GERALDENE MUNSAMY

216039011

A thesis submitted to the College of Health Sciences, University of KwaZulu-Natal Westville, in fulfilment of the requirements of the degree of Master of Medical Sciences

Supervisor

Prof. Mahmoud Soliman

Durban

2016

SIZE DOES NOT MATTER: A MOLECULAR INSIGHT INTO CHEMICAL FRAGMENTS UTILIZING COMPUTATIONAL APPPROACHES

2016

GERALDENE MUNSAMY

216039011

A thesis submitted to the College of Health Sciences, University of KwaZulu-Natal, Westville, in fulfilment of the requirements of the degree of Master of Medical Sciences

This is to certify that the contents of this thesis are the original research work of Miss **Geraldene Munsamy**

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

Supervisor:

Signed:…… Date:……...

ABSTRACT

Insight into the functional and physiological state of a drug target is of essential importance in the drug discovery process, with the lack of emerging (3D) drug targets we propose the integration of homology modeling which may aid in the accurate yet efficient construction of 3D protein structures. In this study we present the applications of homology modeling in drug discovery, a conclusive route map and detailed technical guideline that can be utilised to obtain the most accurate model. Even with the presence of available drug targets and substantial advancements being made in the field of drug discovery, the prevalence of incurable diseases still remains at an all-time high. In this study we explore the biological activity of chemically derived fragments from natural products utilising a range of computational approaches and implement its use in a new route towards innovative drug discovery. A potential avenue referred to as the *reduce to maximum concept* recently proposed by organic chemists, entails reducing the size of a chemical compound to obtain a structural analogs with retained or enhanced biological activity, better synthetic approachability and reduced toxicity. Displaying that size may not in fact matter.

Molecular dynamic simulations along with toxicity profiling were comparatively performed, on natural compound Anguinomycin D and its derived analog SB 640 each in complex with the CRM1 protein which plays an avid role in cancer pathogenesis. Each system was post-dynamically studied to comprehend structural dynamics adopted by the parent compound to that exhibited by the analog. Although being reduced by 60% the analog SB 640 displayed an overall exhibition of attractive pharmacophore properties which include minimal reduction in binding affinity, enhanced synthetic approachability and reduced toxicity in comparison to the parent compound. Potent inhibitor of CRM1, Leptomycin B (LMB) displayed substantial inhibition of the CRM1 export protein by binding to four of the PKIαNES residues (ϕ0, ϕ1, ϕ2, ϕ3, and ϕ4) present within the hydrophobic binding groove of CRM1. Although being drastically reduced in size and lacking the presence of the polyketide chain present in the parent compound Anguinomycin D and LMB the analog SB 640 displaced three of these essential NES residues. The potential therapeutic activity of the structural analog remains undeniable, however the application of this approach in drug design still remains ambiguous as to which chemical fragments must be retained or truncated to ensure retention or enhanced pharmacophore properties. In this study we aimed to implement the use of thermodynamic calculations, which was accomplished by incorporating a MM/GBSA per-residue energy contribution footprint from molecular dynamics simulation. The proposed approach was generated for each system. Anguinomycin D and analog SB 640 each in complex with CRM1 protein, each system formed interactions with the conserved active site residues Leu 536, Thr 575, Val 576 and Lys 579. These residues were highlighted as the most energetically favourable amino acid residues contributing substantially to the total binding free energy. Thus implying a conserved selectivity and binding mode adopted by both compounds despite the omission of the prominent polyketide chain in the analog SB 640, present in the parent compound. A

strategic computational approach presented in this study could serve as a beneficial tool to enhance novel drug discovery. This entire work provides an invaluable contribution to the understanding of the phenomena underlying the reduction in the size of a chemical compound to obtain the most beneficial pharmacokinetic properties and could largely contribute to the design of potent analog inhibitors for a range of drug targets implicated in the orchestration of diseases.

DECLARATION 1 – PLAGIARISM

I, **Geraldene Munsamy**, declare that:

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

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

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

This thesis does not contain other persons' writing, unless specifically acknowledged as being sourced from other researchers. Where other resources have been quoted, then: their words have been re-written but the general information attributed to them has been referenced.

Where their exact words have been used, then their writing has been placed in italics and inside quotation marks and referenced.

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.

Signed G Munsamy……

DECLARATION 2- PUBLICATIONS

1. Homology Modeling in drug discovery: An update of the last decade (Accepted Letters in Drug Design and Discovery)

Contribution:

Geraldene Munsamy: Author- contributed to the project by performing all literature reviews, as well as manuscript preparation and writing.

Prof. Mahmoud Soliman: Supervisor

2. Size does not matter: A molecular insight into chemical fragments using thermodynamic calculations (Submitted to Journal of Theoretical Biology)

Contribution:

Geraldene Munsamy: Author- contributed to the project by performing all literature reviews, experimental work, and data analysis, interpretation of the results as well as manuscript preparation and writing.

Dr. Ndumiso N. Mhlongo: Provided assistance with technical support

Dr. Ampie Johannes Niehaus: Co-Supervisor

Prof.Mahmoud Soliman: Supervisor

RESEARCH OUTPUT

PUBLICATIONS

1. Homology Modeling in drug discovery: An update of the last decade (Manuscript accepted by Letters in Drug Design and Discovery)

2. Size does not matter: A molecular insight into chemical fragments using thermodynamic calculations (Manuscript submitted under review)

ACKNOWLEDGEMENTS

Philippians 4:13 "I am able to do all things through Christ who strengthens me".

I extend my gratitude to the following parties:

My supervisor Prof. Mahmoud Soliman for his constant support, guidance and invaluable knowledge

My colleagues, family and friends for their support.

The Centre for High Performance Computing (CHPC) for technical support and computational resources

The National Research Foundation of South Africa and College of Health Sciences, UKZN for financial support.

My Co-supervisor and post-doctoral colleagues for their editorial support.

My parents, Mr Ricky and Mrs Jean Munsamy, and my siblings for their unconditional love, inspiration, understanding and support during my studies and in all my life's ventures

Lastly I would like to thank Mr Ronald Budhram for his support, optimism and motivation during my MSc.

TABLE OF FIGURES

TABLE OF TABLES

LIST OF ABBREVIATIONS

3D	Three Dimensional
ADME	Absorption, Distribution, Metabolism, Excretion
APC	Adenomatous Polyposis Coli protein
Cl	Chlorine
CRM1	Chromosome Region Maintenance 1
Cys	Cysteine
δ	Delta
FEP	Free Energy Perturbation
GAFF	General AMBER Force Field
GNP	Phosphoaminophosphoric acid-guanylate ester
HIV 1-REV	Human Immunodeficiency virus type 1 regulatory protein
HTS	High-throughput Screening
IARC	International Agency for Research on Cancer
Ile	Isoleucine
Leu	Leucine
LMB	Leptomycin B
LR	Linear Response
Lys	Lysine
MD	Molecular Dynamics
Mg	Magnesium
MM	Molecular Mechanics
MM/GB-SA	Molecular Mechanic/Generalized Born Surface Area
MM/PB-SA	Molecular Mechanics/Poisson-Boltzmann Surface Area
MMV	Molecular Modeling Viewer
MOA	Mode of Action
NES	Nuclear Export Signal
NFAT	Nuclear Factor of Activated T cells
NLS	Nuclear Localization Signal
NMR	Nuclear Magnetic Resonance

LIST OF ABBREVIATIONS

NP	Natural Products
NPC	Nuclear Pore Complex
PASS	Prediction of Activity Spectra for Substances
PDB	Protein Data Bank
PES	Potential Energy Surface
Phe	Phenylalanine
PKI	Protein Kinase Inhibitor
PME	Particle mesh Ewald Method
Rb	Retinoblastoma protein
RCSB	Research Collaboratory for Structural Bioinformatics
RESP	Restrained Electrostatic Potential
Rg	Radius of Gyration
RMSD	Root Mean Square Deviation
RMSF	Root Mean Square Fluctuation
RNA	Ribonucleic acid
SASA	Solvent Accessible Surface Area
SNUPN	Snurportin-1
Thr	Threonine
TI	Thermodynamic Integration
Val	Valine
XPO-1	Exportin-1
α	Alpha
β	Beta

TABLE OF CONTENT

CHAPTER 1

1.1 Background and Rationale behind this study

With high mortality rates, increased drug resistance, lack of effectivity of therapeutic drugs, reduced bioavailability and limited availability of resources, a more rational approach to drug design and development is required. The identification of prospective bioactive molecules that limit the development of a disease condition, remains a crucial component in understanding the molecular mechanism of pathogenesis displayed by a drug target. Based on a molecular perspective, these requirements translate into a scenario where a drug molecule interacts with one or more target proteins that are directly implicated in the pathophysiology of a disease and may act as inhibitors, agonists or modulators[1]. A full characterization of these interactions on a molecular and structural level may add considerable knowledge that can be utilised in drug design processes[2]. Employing the insight extracted directly from the 3D structural details of a drug target may constitute an accurate account of the drug-protein interaction, providing an acute perspective on the mode of action (MOA) of the drug molecule which may further advance the process of drug design[3]. Utilization of accurate 3D protein structures may ensure drug-target specificity and effectivity[4], which is an eminent requirement in all fields of drug design. Therefore not only the identification of potential drug candidates but the availability of accurate drug protein targets remains a high priority in structure-based drug design protocols. The availability of drug targets, may ensure the development of current therapeutic with enhanced drug target specificity and effectivity, potentially reducing developed drug resistance, lack of effectivity and potency. There are many fields that take these requirements into consideration in the process of drug design and development, however computational chemistry is highlighted as one of the most prominent fields that exploits the use of various tools that may aid in the discovery of not only accurate drug targets but enables the identification of optimal drug candidates[5]. Homology modeling is identified as a potential computational approach, utilised to construct a protein 3D structure, by employing the use of a protein sequence of a known 3D model as a template. The template sequence of one or more homologous proteins with a known structure, is often targeted to construct the most accurate model of the target protein[6]. This study provides a route map that can be used for the construction of the most accurate 3D protein structure, accompanied by detailed technical guidelines and the formidable applications of homology modeling in drug design over the last decade. Once the most accurate model of a protein target, has been established it then becomes possible to identify potential drug candidates that may elucidate avid inhibitory activity against the target protein[7].

In the midst of substantial investments being made in drug discovery, there is a limited number of new chemical entities being introduced to the drug discovery market[2,8]. The optimization of current therapeutic drugs forms an attractive alternative towards bridging the gap in drug design and

development[9].A relatively new approach introduced by organic chemists known as the "*reduce to maximum concept*"[10] entails reducing the size of a chemical compound, by the removal of unwanted fragments resulting in retention or improvement of biological parameters such as potency and selectivity[10]. The concept of reducing the size of compound whilst maintaining maximum bioactivity presents itself as an ideal platform for next generation drug discovery and design[11]. From experimental studies, the application of this concept is associated with enhanced synthetic approachability, gain of time and resources, reduced toxicity, improved bioavailability and sustained selectivity[12].Over the last 10 decades, natural products have made a resurgence in the field of drug discovery as they harvest beneficial pharmacokinetic properties evidently observed as nature's secrets are enshrined in natural products[13,14]. Natural products (NP) are considered to be the ideal target candidates in the application of the *reduce to maximum concept* as they possess chemical fragments that exhibit a vast range of beneficial therapeutic properties[15]. The prime use of statins derived from the natural product lovastatin has led to the development of treatment against cardiovascular diseases, which further exemplifies the essential role that NPs have played in drug discovery. Organic chemists were able to harvest the pivotal chemical scaffold of 3,5-dihydroxypentoate derived from lovastatin, which led to the introduction of more potent and therapeutic effective derivatives, with an influential impact on the current state of society and medicine today[16,17].

Substantial contributions have been made by synthetic protocols towards the discovery of new innovative drugs predominantly in chemical synthesis, however even efforts made by chemical synthesis cannot be the sole solution to the pending disease crisis that looms thus efforts from other fields of drug discovery is encouraged[18]. Computational chemistry is highlighted as the one of the most propagated avenues currently integrated in the drug design and discovery process. As the continuous implementation of computational methods over the many years in drug discovery has led to a substantial reduction in time and resource requirements, which are crucial in chemical synthesis and biological protocols[19]. The use of computational approaches may ultimately alleviate the need for pre-clinical and clinical studies and its associated costs. Harvesting the essential knowledge extracted from NPs in combination with the technical versatility of computational methods may enable the integration of natural products derivatives as displayed in antibody–drug conjugates, stereo-chemical complex fragments, or initiate scaffold repurposing, in next generation drug discovery[20].

It is proposed the incorporation of the *reduce to maximum concept* may enable the synthesis of potential therapeutic drugs with enhanced potency, and a substantial reduction in the total number of synthetic steps[21]. However, it still remains ambiguous as to which fragments of the chemical compounds are essential for activity and which fragments must be truncated. Thus quantitative protocols are required to gain insight into the analogy of which fragments of the chemical compound are essential for activity and which fragments can be omitted[22].In recent years, the use of computational approaches projected

in modern drug discovery processes has transpired into a close equivalent to experimental studies, the insight extracted from computational tools has provided a better understanding in complex biological phenomena[23]. The use of computational approaches may enable a deeper understanding of the chemical compounds structural features and interactions within the active site[24], which is imperative as it may lead to the generation of molecules that emulate the transition state[25,26]. Molecular modeling is a robust computational tool used to study the conformational dynamics of biological systems. In this work, molecular dynamic simulations and in-depth post dynamics analysis were performed for the natural parent compound Anguinomycin D and its derived analog SB 640 in complex with CRM1.Chromosome Region Maintenance protein 1 (CRM1) plays an essential role in all eukaryote organisms as it mediates the transport of cargoes that contain the nuclear export signal (NES) from the nucleus to the cytoplasm[27]. However, the upregulation of this transportation process is primarily associated with common hallmarks for a vast spectrum of cancers. The elevation of the CRM1 protein is associated with a range of cancers such as pancreatic[28], kidney[29] and ovarian[30] along with osteosarcoma[31]. Current nuclear export inhibitors of the CRM1 are associated with elevated levels of toxicity, as displayed by bacterial inhibitor Leptomycin B[32]. Anguinomycin D shares close structural features to Leptomycin B, thus displays inhibitory activity against the CRM1 protein. From experimental studies, it was observed structural analog of Anguinomycin D, analog SB 640 which is reduced by 60% in its structure maintained the inhibitory activity and overall retention of bioactivity[33] displayed by the parent compound Anguinomycin D.

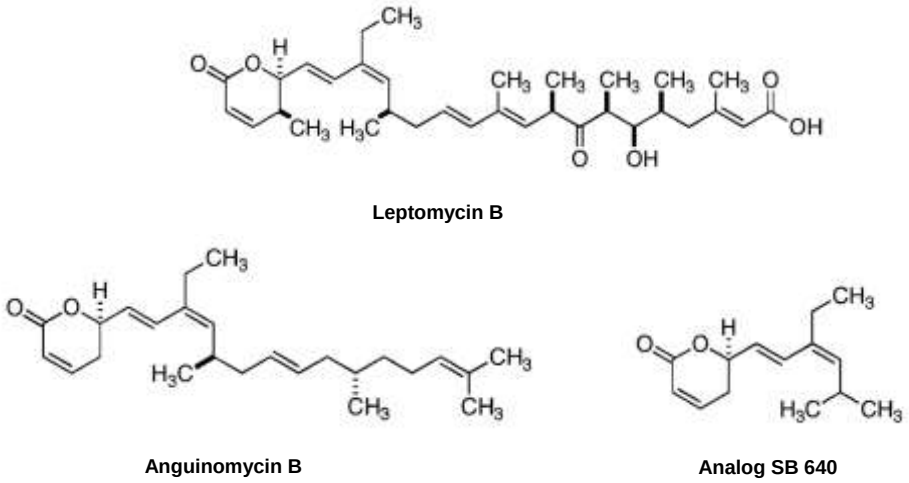


Figure 1.1 Displays the structure of natural products Leptomycin B, which shares close structural similarity to Anguinomycin D and its derived analog SB 640

This study aimed to validate the "*reduce to maximum concept"* and provide further insight in the application of a range of computational approaches that may aid in the implementation of this concept in the design of potential drugs. Per-residue energy decomposition analysis has become an effective approach to drug discovery. Its use in drug design enables the identification of the most highly contributing amino acid residues to the total binding affinity[25], which can be correlated to determine which fragments of the chemical compound are imperative for the molecular and physiochemical properties exhibited by a compound. Per-residue energy decomposition footprints were performed in this study to provide a guideline in distinguishing which fragments of a chemical compound must be retained from the large parent structural and chemical composition to elucidate biological activity and which fragments of a chemical compound can be omitted as its presence may assert toxicity, resulting in desired pharmacophore properties exhibited by the truncated analog.

1.2. Novelty and significance of this study

With the lack of accurate 3D protein models currently available the ability to establish the molecular mechanism of drug interaction between a potential drug candidate and protein target involved in the infestation of a disease condition still remains ambiguous, this study elucidates the use of computational tool of homology modeling which provides a conclusive route map that can be utilized to generate the most accurate 3D protein model. Once the identification of a target protein is established the next step in effective drug design is to generate effective yet potent drug targets. NPs remain as the sole source and the best inspiration for the drug-discovery process. This study aims to reveals the impact of the integration of experimental and theoretical approaches as a new form of rational drug discovery. Thus we implement the use of computational approaches to gain insight into a relatively new approach introduced by organic chemists referred to as the "*reduce to maximum concept"*, which may lead to next generation drug discovery.

The "*reduce to maximum concept"* entails the extraction of the most essential fragments of a chemical compound derived from natural products to obtain a smaller structural compound with retained or enhanced bioactivity, better synthetic approachability and potential reduced toxicity. This approach presents itself as an ideal platform for innovative drug discovery[10]. This study forms as a rationale to understand the mechanism of binding, elucidated by the structural analog SB 640 (of Anguinomycin D) although being drastically reduced in its size, and in doing so may validate the *reduce to maximum concept* that can be implemented in all fields of drug discovery by implementing screening protocols stemming from the use of computational approaches such as binding interaction calculations. Even after conclusive experimental studies conducted the level of interaction portrayed by the reduced chemical compound, elucidating which structural features are essential for retention or enhanced biological activity and which fragments may be omitted remains ambiguous. Utilizing the knowledge extracted

from this study may provide a deeper understanding underlying the molecular and structural interaction of the reduced analog and the target protein. The inception of this approach, integrated with the application of computational tools may form a formidable impact in current drug development, cascading the introduction to next generation drug discovery. This study aimed at utilizing computational approaches to provide a guideline to decipher which chemical fragments of a compound are essential for activity and which fragments can be omitted based on thermodynamic calculations. Elucidating that the size of chemical compound may in fact not matter.

Medicinal chemists and other pharmaceutics have resorted to molecular modelling drug design techniques as its application provides insight underlying the mechanism, dynamics as well as the energetics of proteins and protein-drug complexes[34,35]. Molecular dynamics (MD) is a common tactic presently employed to overcome drug resistance and ultimately adopt a protocol that can be administered in all fields of drug design. There is no previous theoretical study that unveiled the precise molecular level understanding underlying the reduction in the size of a chemical compound and its mechanism of inhibitory activity correlating to its reduced size. Computational chemistry integrates itself as the most suitable avenue to gain insight into the structural dynamics. This is a first attempt at implementing the use of computational approaches such as molecular modeling and per-residue energy decomposition analysis to gain insight into the reduction in the size of a compound to further obtain the most optimal and desired biological parameters. In this study the computational approach was employed to confirm and validate the experimental outcomes and to understand the changes in drug binding landscape. We believe this study may provide immeasurable insight into a new approach that can be implemented in drug design and development.

1.3. Aim and Objectives

This study has three major aims:

1. To investigate the use of the computational approach of homology modeling, highlighting its active role in the drug discovery process:

To achieve this, the following objectives were outlined:

1.1 The protocol of homology modeling is determined, along with technical guidelines that may result in the most accurate models being generated

1.2 The application of homology modeling in drug design and discovery was investigated over the last decade.

2. Validation of the "reduce to maximum" concept based on the magnitude of ligand binding affinity of the parent compound Anguinomycin D in comparison to the analog SB 640 as well as per-residue energy decomposition analysis

To achieve this, the following objectives were outlined:

2.1 Molecular dynamic simulations were performed along with post MD analysis of the parent compound Anguinomycin D and analog SB 640, to analyse the mechanism of binding exhibited by both compounds

2.2 To estimate the ligand binding free energies of Anguinomycin D and its derived analog SB 640 in complex with the CRM1 protein to determine if the reduction in the size of chemical compound alters the total binding free energy contribution.

2.3 Per-residue energy decomposition analysis were performed to determine which fragments of the chemical compound is essential for the biological activity exhibited and which features of the chemical compound can be omitted.

3. To determine the toxicity profile of the parent compound Anguinomycin D and the analog SB 64

To achieve this, the following objectives were outlined:

3.1 The PASS online prediction test were performed for Anguinomycin D and analog SB 640

1.4. Overview of this work

The format of this thesis is by publication with Chapter4-5 containing copies of papers that have been accepted and/or submitted to journals. This thesis is divided into six chapters, with this one included:

Chapter 2: This chapter begins with highlighting the formidable application of homology modeling, in the construction of accurate 3D protein targets that can be used in effective drug design. It further highlights the "*reduce to maximum concept"* and its potential application in the drug discovery process. It inaugurates the use natural products as a major source for novel compounds, focusing on chemical fragments that can be isolated from natural products. It highlights factors associated with the prevalence of many diseases, highlighting cancer as one of the highest contributors to the high mortality observed. It further focuses on the CRM1 protein which is associated with the pathogenesis of a range of cancers, highlighting the first known inhibitor of CRM1, Leptomycin B and its discontinuous use in clinical trials, attributed to its associated high levels of toxicity associated with its use, it also highlights the inhibitory activity of natural product Anguinomycin D and its drastically reduced structural analog SB 640 against the CRM1 protein.

Chapter 3: This is a brief introduction to computational chemistry, various molecular modeling and molecular simulation protocols and their applications. Some computational methods have been theoretically explained, followed by a range of computational tools applied in the study of Anguinomycin D and its derived analog SB 640 in complex with the CRM1 protein, with the fundamental focus on molecular dynamics simulations, binding free energy calculations and homology modelling.

Chapter 4 (Manuscript Accepted)

This chapter presents a review on homology modeling titled "Homology modeling in drug discovery: An update on the last decade. It highlights the use of homology modeling in the drug discovery process, a technical guideline to construct the more accurate model as well as the application of homology in drug discovery over the last decade.

Chapter 5 (Submitted for publication)

This chapter presents results from the study titled "Size does not matter: A molecular insight into the biological activity of chemical fragments using thermodynamic calculations". It presents data extracted from molecular dynamics simulations performed on Anguinomycin D and analog SB 640, as well thermodynamic calculations, which elucidated a minimal reduction in the binding free energy of the analog in comparison to the Anguinomycin D, enhanced synthetic approachability and an overall reduction in toxicity.

Chapter 6

This chapter expounds the overall concluding remarks of the entire thesis and future plans and recommendations

References

(1) König, J., Müller, F., and Fromm, M. F. (2013) Transporters and Drug-Drug Interactions : Important Determinants of Drug Disposition and Effects. *Pharmacol. Rev. 65*, 944–966.

(2) Bathula, S. R., Akondi, S. M., Mainkar, P. S., and Chandrasekhar, S. (2015) "Pruning of biomolecules and natural products (PBNP)": an innovative paradigm in drug discovery. *Org. Biomol. Chem. 13*, 6432–6448.

(3) Kreiger, E., Nabuurs, S. B., and Vriend, G. (1963) Homology modeling. *Struct. Bioinforma.*

(4) Kalyaanamoorthy, S., and Chen, Y. P. (2011) Structure-based drug design to augment hit discovery. *Drug Discov. Today 16*, 831–839.

(5) Swinney, D. C., and Anthony, J. (2011) How were new medicines discovered ? *Nat. Rev. 10*, 507–519.

(6) Antolovic, R. (2007) Homology Modeling of Human Fyn Kinase Structure: Discovery of Rosmarinic Acid as a New Fyn Kinase Inhibitor and in Silico Study of Its Possible Binding Modes. *J. Med. Chem. 50*, 1090–1100.

(7) Kirchmair, J., Williamson, M. J., Tyzack, J. D., Tan, L., Bond, P. J., Bender, A., and Glen, R. C. (2012) Computational Prediction of Metabolism: Sites, Products, SAR, P450 Enzyme Dynamics, and Mechanisms. *J. Chem. Inf. Model. 52*, 617–648.

(8) Bernhoft, A. (2010) Bioactive compounds in plants – benefits and risks for man and animals, in *The Norwegian Academy of Science and Letters*, pp 13–14.

(9) Atanasov, A. G., Waltenberger, B., Pferschy-wenzig, E., Linder, T., Wawrosch, C., Uhrin, P., Temml, V., Wang, L., Schwaiger, S., Heiss, E. H., Rollinger, J. M., Schuster, D., Breuss, J. M., Bochkov, V., Mihovilovic, M. D., Kopp, B., Bauer, R., Dirsch, V. M., and Stuppner, H. (2015) Discovery and resupply of pharmacologically active plant-derived natural products : A review. *Biotechnol. Adv. 33*, 1582–1614.

(10) Crane, E. A., and Gademann, K. (2016) Capturing Biological Activity in Natural Product Fragments by Chemical Synthesis. *Angew. Chemie Int. Ed. 55*, 2–23.

(11) Gademann, K. (2015) Copy, Edit, and Paste: Natural Product Approaches to Biomaterials and Neuroengineering. *Acc. Chem. Res. 48*, 731–739.

(12) Bonazzi, S., Eidam, O., Gu, S., Wach, J., Zemp, I., Kutay, U., and Gademann, K. (2010) Anguinomycins and Derivatives : Total Syntheses , Modeling , and Biological Evaluation of the Inhibition of Nucleocytoplasmic Transport. *J. Antimicrob. Chemother. 132*, 1432–1442.

(13) Pathania, S., Ramakrishnan, S. M., and Bagler, G. (2015) Original article Phytochemica : a platform to explore phytochemicals of medicinal plants. *Database Sci. Journals 2015*, 1–8.

(14) Lahlou, M. (2013) The Success of Natural Products in Drug Discovery. *Sci. Res. 2013*, 17–31.

(15) Hong, J. (2014) Natural Product Synthesis at the Interface of Chemistry and Biology. *Chem. A Eur. J. Concept* 10204–10212.

(16) Istvan, E. S., and Deisenhofer, J. (2012) Structural Mechanism for Statin Inhibition of HMG-CoA Reductase. *Am. Assoc. Pharm. Sci. 1160*.

(17) Jain, K. S., Kathiravan, M. K., Somani, S., and Shishoo, C. J. (2007) The biology and chemistry of hyperlipidemia. *Bioorg. Med. Chem. 15*, 4674–4699.

(18) Harvey, L. A., Edrada-Ebel, R., and Quinn, J. R. (2011) The re-emergence of natural products for drug discovery in the genomics era. *Biotechnol. Adv.* 1–52.

(19) Honarparvar, B., Govender, T., Maguire, G. E. M., Soliman, M. E. S., and Kruger, H. G. (2014) Integrated Approach to Structure-Based Enzymatic Drug Design : Molecular Modeling , Spectroscopy , and Experimental Bioactivity. *Chem. Rev. 114*, 493–537.

(20) Harvey, A. L. (2008) Natural products in drug discovery. *Drug Discov. Today 13*, 894–901.

(21) Gu, J., Gui, Y., Chen, L., Yuan, G., Lu, H., and Xu, X. (2013) Use of Natural Products as Chemical Library for Drug Discovery and Network Pharmacology. *Peer-reviewed open accesses J. 8*, 1–10.

(22) Cragg, G. M., and Newman, D. J. (2014) Natural Products: A Continuing Source of Novel Drug Leads. *Biochim. biophys acta J. 1830*, 3670–3695.

(23) Nagle, G. D., and Zhou, Y.-D. (2010) Natural Product-Derived Small Molecule Activators of Hypoxia-inducible Factor (HIF-1). *NIH Public Access 12*, 2673–2688.

(24) Oikawa, M., and Yonetani, Y. (2010) Molecular Dynamics Free Energy Calculations to Assess the Possibility of Water Existence in Protein Nonpolar Cavities. *Biophysj 98*, 2974–2983.

(25) Onufriev, A., Bashford, D., and Case, A. D. (2016) Modification of the Generalized Born Model Suitable for Macromolecules. *J. Phys. Chem. 2000*, 3712–3720.

(26) Wang, J., and Deng, Y. (2006) Absolute Binding Free Energy Calculations Using Molecular Dynamics Simulations with Restraining Potentials *91*, 2798–2814.

(27) Solá, R. J., and Griebenow, K. (2011) Glycosylation of Therapeutic Proteins: An Effective Strategy to Optimize Efficacy. *Biodrugs 24*, 9–21.

(28) Zhou, F., Qiu, W., and Yao, R. (2013) CRM1 is a novel independent prognostic factor for the poor prognosis of gastric carcinomas. *Med. Oncol. 30726*, 1–7.

(29) Dickmanns, A., Monecke, T., and Ficner, R. (2015) Structural Basis of Targeting the Exportin CRM1 in Cancer. *Cells 4*, 538–568.

(30) Faustino, R. S., Nelson, T. J., and Terzic, A. (2007) Nuclear Transport : Target for Therapy. *Clin. Pharmacol. Ther. 81*, 880–886.

(31) Yedavalli, V. R. K. (2013) CRM1 Organization of CRM1 Protein Cargo Transport Functions of CRM1 Rev Is an Export Substrate of CRM1. *AIDS*. Springer Science.

(32) Turner, G. J., Dawson, J., Cubitt, C., Baz, R., and Sullivan, M. D. (2015) Inhibition of CRM1-dependent nuclear export sensitizes malignant cells to cytotoxic and targeted agents. *Semin. Cancer Biol. 0*, 62–73.

(33) Hayakawa, Y., and Sohda, K. (1995) Anguinomycins C and D, NewAntitumor Antibiotics with Selective Cytotoxicity against Transformed Cells. *J. Antibiot. (Tokyo). 48*, 954–961.

(34) Lioa, C., Sitzmann, M., and Pugliese, A. (2012) Software and resources for computational medicinal chemistry. *Med. Chem. Res. 3*, 1057–1085.

(35) Lionta, E., Spyrou, G., Vassilatis, D. K., and Cournia, Z. (2014) Structure-Based Virtual Screening for Drug Discovery : Principles , Applications and Recent Advances. *Curr. Top. Med. Chem. 14*, 1923–193

CHAPTER 2

2.0 Literature Review

2.1. Introduction

The prevalence of incurable diseases has exponentially increased over many years, in spite of the substantial investments in the field of drug design and discovery[1]. The identification of innovative drugs emerging from chemical and pharmaceutical industries remains minimal, with accompanying attrition rates experiencing an all-time low[2]. In order to fully characterise the effectivity of a drug, the mode or mechanism of action must be established thus drug target selection prevails as a decisive factor in drug discovery research and productivity[3]. Since the early inception of the genomics era in the 1990s, the sole focus of drug discovery has been focused on the identification of potential drug targets[4], ideally proteins that play a crucial in disease pathogenesis. Modification of biological activity of the target protein form the rational basis for the discovery of new drug candidates[5]. A target-centric approach imparts a specific biological hypothesis, based on the structural, and molecular composition of the target protein, which can be utilised as a starting point for the identification of potential therapeutics[6].

In order to gain precise insight into the target protein mechanism and structural dynamics of the model of the 3D target protein must be established. Homology modeling is identified as an attractive approach branching from the field of computational chemistry. Homology modeling also known as comparative modeling can be utilised for the construction of an accurate 3D target model, in a relatively short period of time based on the 3D model from a template protein sequence with a known structure[7]. Homology modeling enables the refinement of novel therapeutic drug targets which may aid significantly in drug design and development. Once the target protein is available, the definitive mechanism by which a potential drug candidate elucidates its inhibitory activity can be proposed, on the basis of its interaction with the drug target and its conformational fit within the active site[8].

An accurate model of the 3D drug target may enable a better understanding of the role it undertakes during the molecular and structural transition leading to the disease state[9]. Utilizing the crucial knowledge extracted from the three-dimensional structure of the biological target, may assist in proficient yet cost-efficient lead discovery and optimization[5]. However the lack of new chemical entities towards effective drug discovery depicts how difficult it is to randomly generate potent and selective compounds[2].The chemical versatility and pharmacophore properties displayed by NPs has been a key factor in the resurgence of their use in drug research and development.

2.2. Biological activity of Natural products in lead drug discovery

Natural products (NP) encompass many centuries of un-relinquished potential, the chemical diversity confined within NP[10,2], continue to illustrate its importance in modern drug discovery efforts. There has been an evident resurgence of NPs over the last decade, harvesting the knowledge encompassed in these rich commodities results in elevated biological activity, high specificity and reduced toxicity. chemically derived fragments with high specificity and beneficial biological activity[11].The unique yet divergent structural composition of NPs, have certainly broadened the current database of existing organic molecules. As the integration of NPs in drug discovery has led to the introduction of novel chemical fragments[12], with contrasting molecular skeletons and distinctive biological and functional features that would otherwise be overlooked, with the application of synthetic protocols[13].

2.3. Resurgence of natural products in drug discovery

In recent times, there has been a relapse in the search towards alternative drug discovery methods, limiting the discovery of prospective drug candidates[14], required by essential therapeutic areas such as immunosuppression[15], anti-cancer drugs[16], anti-inflammatory and treatment against viral diseases[17]. The therapeutic potential disclosed by NPs reaffirms their promising use in the field of drug design and development[18]. These fields include genetics and enzymology which have gained immeasurable insight and thrived, by assimilating the process of biosynthesis of NPs[19]. NPs have also substantially improved industrial sectors such as the pharmaceutical and agriculture industries with the provision of considerable insight into the identification of useful molecules and lead compounds[1].

The contribution of NPs towards the drug discovery substantially supersedes the contributions made by combinatorial chemistry which prompts the development of synthesized compounds[20]. Observations have shown that between the years 1981-2010, there has been a minimal contribution of 36% of newly discovered chemical entities, that have been synthetically derived with no influence from the structural and chemical composition of NPs[21]. Although there has been a long standing dependence on combinatorial chemistry by drug developing companies, minimal progress in drug development has been achieved over the many years[22,23].

The intrinsic chemical and structural composition of NPs display a large-scale of diversity in comparison to synthetic compounds, consisting of an array of biologically active compounds playing an active role of a protagonist in innovative drug discovery[24]. The use of NPs is illuminated in cancer research as they display interesting lead structures with promising chemical entities[25]. The versatile chemical composition of NPs, make up the fundamental scaffolds and are considered privileged structures, that have been a product of evolution[26]. The utilization of the chemical fragments harboured by NPs has had an immense influence on all spheres of science and has advertently inaugurated the use of these chemically derived fragments in the recent process of drug design and discovery[27].

2.4. Pitfalls in the use of natural products as potential drug leads

There are various liable reasons for the scepticism associated with the current integration of NPs in drug discovery, as the time and protocol required for NPs isolation is a rather consuming and an elaborate affair. The ability to differentiate existing chemical compounds from NP, from those that may be newly discovered to avoid strenuous synthetic protocols associated with de-replication still remains a plaguing challenge[28]. Although NP display beneficial and pharmacological properties, NPs may produce an array of toxins, utilised by plants as a mechanism of defence that may not be embraced in the field of drug discovery due to its association with adverse side effects and lethal dosage of toxicity[29]. Some major setbacks associated with the use NPs as active pharmaceutical ingredients, include limited oral bioavailability and low stability under physiological conditions. The use of NP also lacks adequate synthetic approachability, as NP tend to lack stringent criteria of "drug likeness" following Lipinski's "rule of five[30]" and related ADME/pharmacokinetic[31] criteria (ADME=absorption, distribution, metabolism, excretion. Lipinski's rule of five was first introduced to provide guidelines that synthetic chemists could adhere to, to establish better biophysical properties and to ensure optimal bio-active drug candidates[32]. In the late 1990s the phasing out of NP in drug development was prompted by the application of high-throughput screening (HTS)[33] of small molecules. However its use was short lived due to the downward trend of potential drug candidates in drug design. Therefore highlighting the need for improvised approaches to accelerate the drug discovery process.

2.5. Harvesting chemical fragments from natural sources in drug discovery

Prompted by the relentless search for authentic novel lead compounds that could be integrated in the development of potential therapeutic drugs, led to resurgence of NPs. At the inception of research into natural products, the utilisation of chemically derived fragments harvested from these rich resources, rapidly lead to the emergence of innovative strategies towards drug discovery and later, towards the replication of these essential compounds[34].. The chemical fragments isolated from natural products make them a viable option for lead structure in drugs as they display complex chemical diversity[35], a vast range of biochemical specificity and beneficial bioactive properties which include a broader spectrum of molecular property distribution, such as a contrasting molecular mass. The prominent structural features of NP, include the eminent presence of chiral centres, with enhanced steric complexity. The high presence of oxygen atoms in comparison to nitrogen, sulphur and halogen containing groups is also observed in NP, accompanied by generation of resulting OH bond formation. From a statistical analytical perspective NP display, a lower ratio of aromatic rings to the sum of the heavy atoms, whilst the number of hydrogen-bonds donors that are solvated to those that are acceptors displays higher favourability, these properties exhibited by NP differ quiet to distinctively from counterparts such synthetic drugs and combinatorial libraries[13]. With the majority of the rings present in the structure of NP, selectively or completely saturated proves advantageous as NP display greater

interactivity with an array of proteins and enzymes[36]. Current therapeutic drugs implement the use of less than one-fifth of the ring systems present in natural products. The use of chemical fragments derived from NP in modern pharmacotherapy remains imperative, as clear indications of its importance is displayed in its use in a range of an anti-cancer agents, which include paclitaxel and its derivatives from yew (*Taxus*) species, brentuximab vedotin, vincristine and vinblastine from Madagascar periwinkle (*Catharanthus roseus* (L.) G. Don), and *camptothecin;* and a range of derived analogs. Recently, other classes of compounds such as the epothilones have shown promise in the battle against cancer and one of its derivatives[37].

However in recent years, despite the obvious importance, there has been a residual decline in the isolation of these chemical fragments isolated from NP by pharmaceutical industries. Which can be attributed to the relentless process of isolation and limited chemical tractability reduced number of newly discovered therapeutic drugs reaching the market, decreasing substantially from that which was anticipated[38]. Other identified obstacles included low stability under physiological conditions and limited oral bio-availability[12].

Therefore a new alternative approach for drug discovery is required. Introducing an evolving approach encouraging the modification of chemical compounds from natural resources by reducing the size of the compound whilst retaining or even enhancing the biological activity exhibited elucidating that size does matter[39]. This unique approach provides an avenue to tailor natural products through synthesis of a smaller compounds with essential parts, providing an extraordinary approach to unlock the full potential of natural products, producing optimal results with a total reduction in time and cost.

2.6. The "*Reduce to Maximum concept*"

The *reduce to maximum concept* entails the reduction in the size of a chemical compound, whilst still retaining or enhancing bioactivity, ensuring better synthetic approachability perpetuating the reproduction of essential chemical compounds derived from NP, that would otherwise be inaccessible from natural resources. It was first proposed by Crane and collegues[34] the approach enables the versatile use of these chemical compounds derived from NP and offers a unique platform for the use of NP as a major source for innovative drug discovery[25].The retention or enhanced activity exhibited by a chemical compound that has been drastically reduced in size can be attributed to the compound now being able to eloquently position itself within the active size eliminating factors such as, steric hindrance, bulky side chains, producing optimal torsional flexibility[40]. The "*reduce to maximum concept"* introduces the generation of functionally optimized analogues. Their integrated use has led to innovative drug discovery in various fields, from biofilm prevention, drug resistance alleviation and neuroengineering[41]. A key example would be the use of Eribulin mesylate, derived analog of marine sponge *Halichondria okadai* Kadota[42] the natural compound Halichondrin B displayed in figure 2.1. The synthetic analog

displayed synergistic activity as the parent compound, whilst still superseding the desired pharmacokinetic properties of the parent compound comprising of a 35% reduction in molecular weight.

Halichondrin B **Eribulin**

Figure 2.1.Halichondrin B isolated from marine sponge *Halichondria okadai Kadota* and its structural analog Eribulin[42].

The lack of availability, strenuous synthetic approachability and dose-limiting toxicities associated with the natural product Bryostatin isolated from *Endobugula sertula*, led to the development of picolog a structurally simplified analog of Bryostatin. Although being reduced by just 20% from the parent compound, the structural analog picolog superseded the biological activity of Bryostatin used in the treatment of cardiovascular disease, stroke, pain, or cognitive dysfunction[43].

This approach enables the retention of suitable effects in minimized, bioactive scaffolds which can be attributed to "fragment likeness" perpetuating a high-degree of similarity[29]. This *reduce to maximum concept* provides an avenue to overcome the limitations associated with the use of parent natural product structural complexity such as lack of accessibility or viable synthetic approachability, poor synthetic specificity and low producing yields. In addition this approach may overcome barriers in drug design such as drug resistance high levels of toxicity, reduced potency and effectivity, limited bioavailability and adverse side effects, providing a viable yet rational protocol that can be followed for next generation drug discovery. Enabling the utilization of the most essential fragments of a chemical compound as not all constituents of the natural products can be easily synthesized due to high structural complexity possessed by NP as well as the high costs associated with its synthesis on an industrial scale. Exploitation of the chemical and structural databases which are comprised of a range of diverse yet unique chemo types derived from NPs, may facilitate the introduction of new versatile chemical entities.

Amongst the stumbling block encountered with regard to drug design and discovery, drug resistance is one of highest contributing factors[44]. Administration of pharmaceutical drugs for the treatment of a range of diseases, may often lead to the tolerance of these disease to the administration of therapeutic treatment, resulting in drug resistance. There is a large degree of complexity, escalating from *in vivo*

drug activation, which exhibits an intricate mechanism in which particular substances interact with different proteins. The interactions that occur between the target protein or enzymes and the therapeutic drugs, result in the modification, degradation and optimally the activation of the drug to perform its biological activity. Many anticancer drugs must undergo metabolic activation in order to acquire clinical efficacy[45]. However decreased drug activation may stem from cancer cells as they gradually gain resistance to treatment. The distinct prevalence of drug resistance highlighted the dire need to facilitate further research for a more permanent solution to the pending life threatening disease. The alarming rate of resistance exhibited by a range of cancers to once effective traditional therapies[46] has heightened the search for potential drug candidates. The battle against cancer has been one that has been on-going for many decades, in more recent times the treatment of various cancers are often associated with adverse side effects that may contribute to the reduction in effective therapeutic treatment. Further prevalence of this disease may be attributed to a range of contributing factors such as high levels of toxicity and continuous developed resistance[47].Thus the search of alternative treatment is a focal point in alleviating the burden inflected by this dreaded disease.

2.7. The role of Nucleocytoplasmic transport in cancer pathogenesis

Studies conducted by the International Agency for Research on Cancer (IARC), estimated 14.1 million new cases of cancer observed in the year 2012 globally, of these new cases 8 million occurred in developing countries, which constitute 82% of the world's population[48] as observed in figure 2.2. The dire effects of this pernicious disease is predicted to further impend a devastating burden on a global spectrum if it remains incurable, the number of newly emerging cases is expected to rise to 21.7 by the year 2030, with 13 million cancer deaths arising from the expansion and aging of the population[49].

The burden impended by cancer is one that is being experienced globally, there are wide range of resources dedicated towards the research of cancer and the identification of potential onco-targets. It has been observed, that the upregulation of nucleocytoplasmic transport is predominantly associated with the elevated expression of CRM1 protein. The elevated expression of the CRM1 protein correlates with the high incidence of cancers such as ovarian[50], pancreatic[51,52], kidney[53] and cervical cancers[54] as well as gastric carcinoma[55], glioma[56], osteomsarcoma[57], leukemia[58,59]. In addition to mantle cell lymphoma[60], multiple myeloma[61] and melanoma[62,63] and a range of others. In 2016 it has surfaced that majority of these cancers are some of the most prevalent cases of cancer. Nucleocytoplasmic transport upregulation is also associated with drug resistance and stands out as a poor prognosis factor in many malignancies[64].

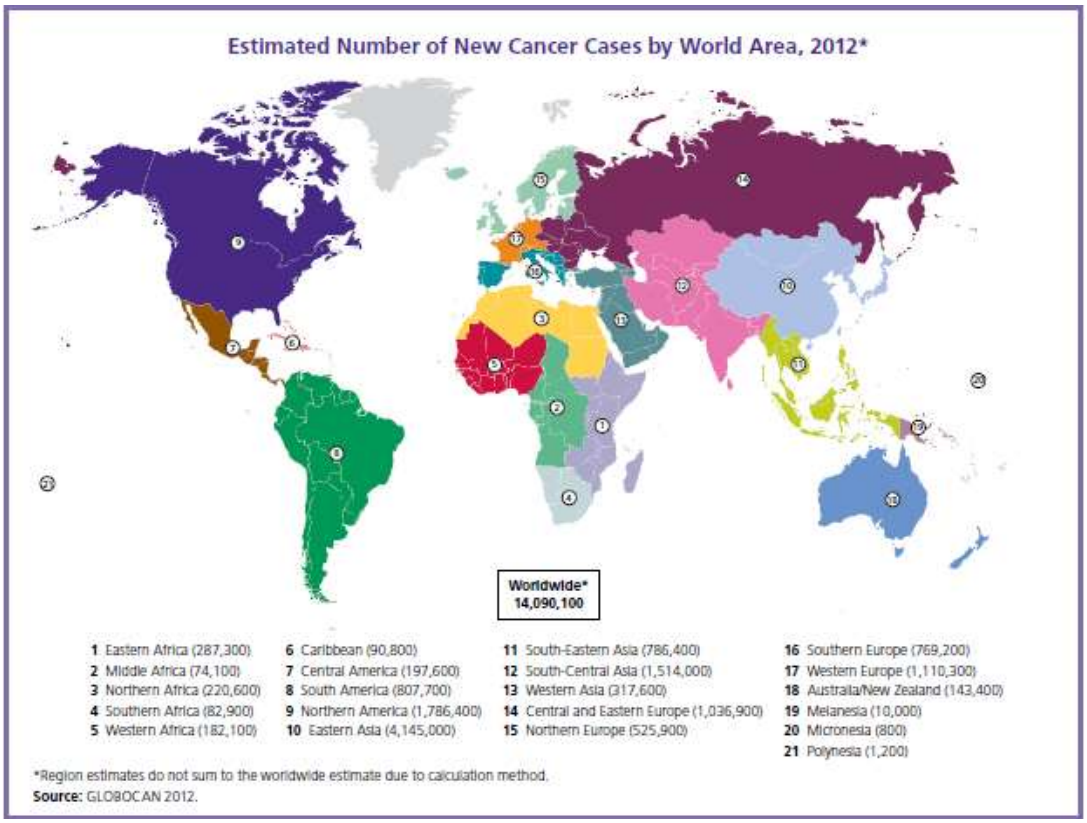


Figure 2.2. Estimate number of newly emerging cases of cancer cases in 21 world areas, 2012[65].

The presence of the cell nucleus distinguishes eukaryotic organisms from prokaryotes, it is enclosed by the nuclear envelope and isolated from the rest of the cell[66]. The nuclear pore complex (NPC) is the prominent portal mediating the transfer of proteins or nucleic acid between the nucleus and the cytosol. Relatively smaller molecules are freely permeable across the NPC, however to travel through these pores, larger molecules such as proteins exceeding a mass of 30KDa are adversely restricted by the NPC and are reliant on the reinforcement of transport receptors[67]. The transport receptors are regarded as an alternative mode of transport for these larger molecules or even smaller molecules to be transported efficiently in and out of the nucleus. This presence of soluble transport factors called karyopherin proteins comprise of these transport receptors[68].

There are a wide range of karyopherin proteins encoded by the human genome utilised for transporting a set of cargoes (protein or RNA) that comprise of specific sequences/motifs, also known as nuclear localization signal (NLS) or Nuclear export signal (NES) or both[69]. The karyopherin proteins can be distinguished as importins or exportins depending on their direction of transport in or out of the nucleus. Karyopherin proteins that can mediate both the import and export of molecules are referred to as

transportins[70]. The Karyopherin protein mediate their mechanism of interaction by directly binding to exposed NLS or NES, and then regulate the movement of cargo to the cytoplasm or nucleus[71]. Dynamic nuclear–cytoplasmic Shuttling of specific proteins out of the nucleus is essential for the regulation of the cell cycle and proliferation of both normal and malignant tissues. Therefore a dysregulation of this fundamental process may alter a wide radius of cellular processes such as gene expression, signal transduction, immune response, and cell differentiation, fundamentally leading to the prognosis of cancer[72]. The crucial role of the CRM1 protein is highlight in nucleocytoplasmic transport, in addition to signal transduction, immune response as well as cell differentiation[74] (Figure 2.3)

2.8. The role of CRM1: Nuclear export factor

Chromosome Region Maintenance 1 protein (CRM1), also known as exportin-1 (XPO1), belongs to the karyopherin β family[75]. The exportin CRM1 is a 120 KDa protein and is the most prominent nuclear export receptor in the cell.The CRM1 protein mediates the export of cellular proteins and is associated with a leucine-rich nuclear export signal (NES, a short 8–15 amino acid hydrophobic motif)[59] out of the nucleus. The location of gene coding for CRM1 is situated on chromosome 2p15 and is widely preserved in a range of species such candida, yeast, drosophila, xenopus, and mammals. The evolutionary conserved nature of CRM1 highlights its central role for NES-dependent nuclear export of protein complexes[76]. For a cargo to exit the nucleus, the CRM1 must recognise the nuclear export signal (NES), thereafter initiating the cooperative formation of a tight trimeric complex CRM1/NES-*cargo*/RanGT. The complex translocates together into the cytoplasm, where RanGTP is hydrolyzed to RanGDP by RanGAP[42] This decreases the affinity between NES and exportin, causing dissociation of cargoes. Bidirectional karyopherins bind to NLS cargoes in the cytoplasm and bind to NES cargoes when exiting nucleus, with similar cargo association/dissociation mechanism to importins and exportins discussed above[50].

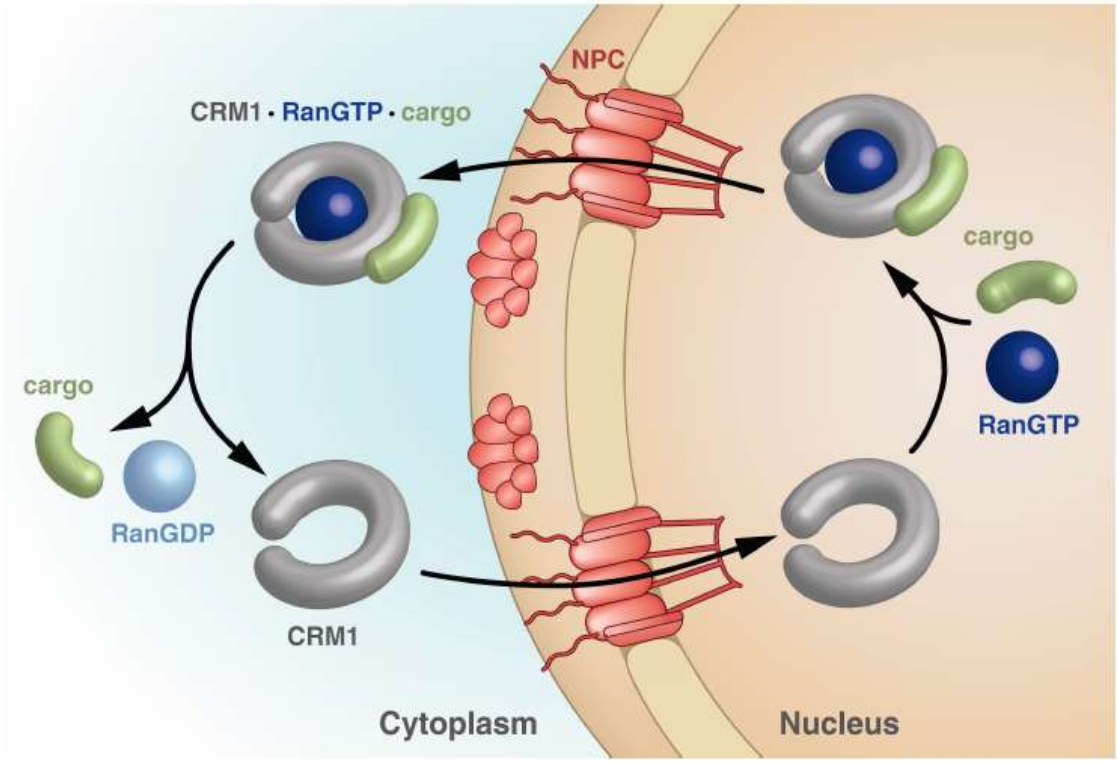


Figure 2.3. Schematic drawing representing the mechanism of nucleocytoplasmic transport. The steps highlighted are of the varying stages involved in the transport between the nucleus and the cytoplasm[66].

CRM1 perpetuates the cellular localization of an expanding range of diverse-functioning protein cargos, inclusive of many tumour suppressor, cell cycle proteins, and viral proteins. Examples of nuclear effectors which are exported into the cytoplasm in cancer include the drug targets topoisomerase IIα[77] and BCR-ABL[78] and tumour suppressor proteins such as Rb[79], APC[80], p53[81], p21[82] and p27[83,74] like APC (adenomatous polyposis coli protein), NFAT (nuclear factor of activated T-cells), β-catenin or Survivin, Rb (retinoblastoma protein), p53 and Bcr-Abl mislocalize in different cancer cells (Figure 2.4). Another protein that is dependent on the CRM1 protein for mediated transport is the human immunodeficiency virus type 1 (HIV-1) regulatory protein Rev; the NES of Rev is recognized by the exportin 1 and transported out of the nucleus. HIV-1 Rev protein moderates the regulation of the HIV-1 mRNA which initiates the export of un-spliced and partially spliced mRNA[25]. The diagnosis of many cancers as well as viral and inflammatory diseases can be directing related to the aberrant mislocalization of cellular CRM1 cargoes, which may interrupt normal cellular functioning systems[84]. Studies reveal restricting the level on interaction occurring between the CRM1 and proteins containing the NES may provide potential therapeutic benefits that can be utilised in the development of more permanent treatment against cancer and other disease conditions[85]. This can be attributed to cancer cells utilizing the nuclear-cytoplasmic transport system to enable the stimulation of tumour growth and simultaneously evade apoptotic processes[47]

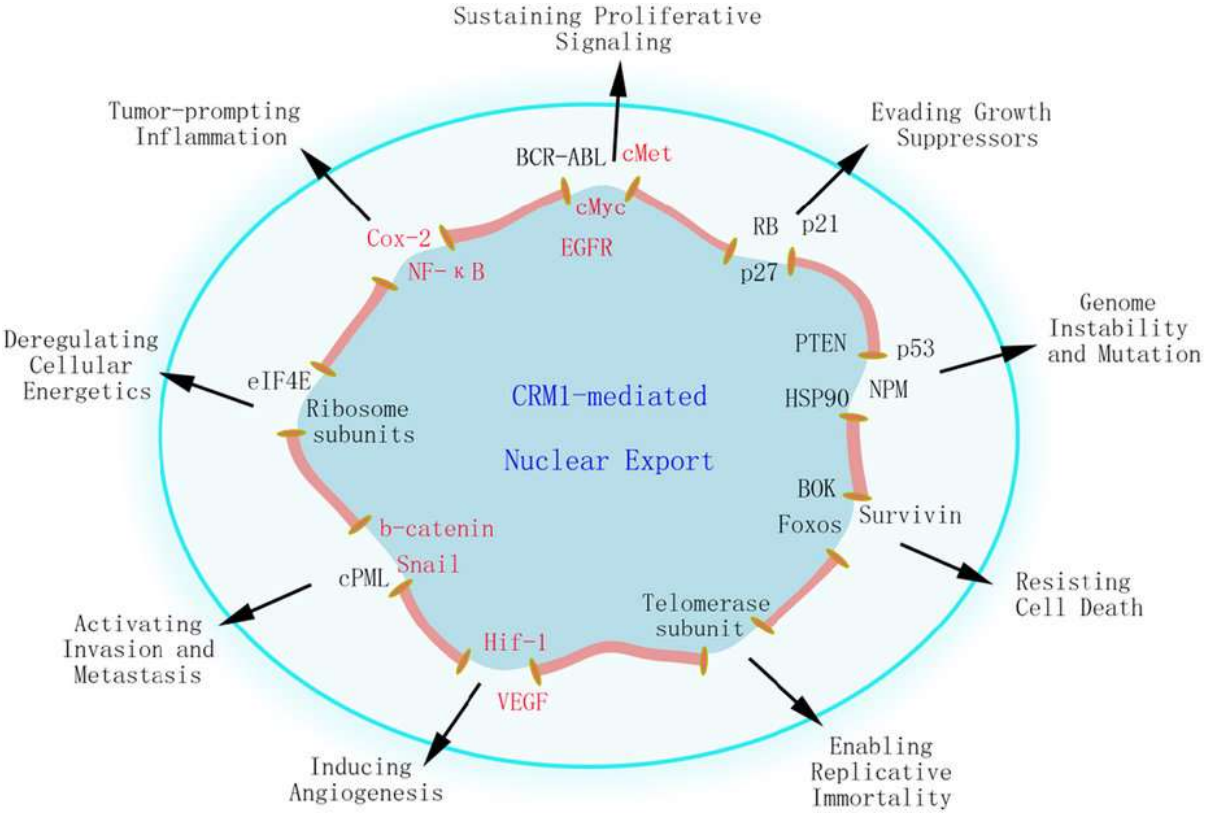


Figure 2.4. CRM1-mediated nuclear export highlighting the most prominent cancer hallmarks. Proteins in black display direct cargoes of CRM1. Proteins in red, are shown to be suppressed by nuclear export inhibition through a diverse mechanism[86].

2.9. Nuclear export inhibitors

Cognate peptide segments present in protein cargoes to which CRM1 binds are referred to as the fundamental nuclear export signals or NESs (commonly known as leucine-rich NESs). The amino acids residues that contribute to the NES peptides are regularly spaced conserved hydrophobic residues which form a groove situated on the outer/convex surface of the ring-shaped CRM1[87]. The NES peptides display an elevated affinity range for the CRM1 protein and are highly diverse in its composition. NES peptides are usually 8–15 amino acids long with regularly spaced conserved hydrophobic residues as depicted in Figure 2.5[88]. Data extracted from Sequence, peptide-library, and bioinformatic analyses best concluded the description for NESs is a set of six consensus sequences, which differ in the spacing's between four key hydrophobic residues ϕ1, ϕ2, ϕ3, and ϕ4[82].There are three diverse accessible structures for NESs which include cargos extracted from protein kinase A inhibitor (PKI), Snurportin-1 (SNUPN), and the HIV1-Rev protein, all of which are bound to CRM1.

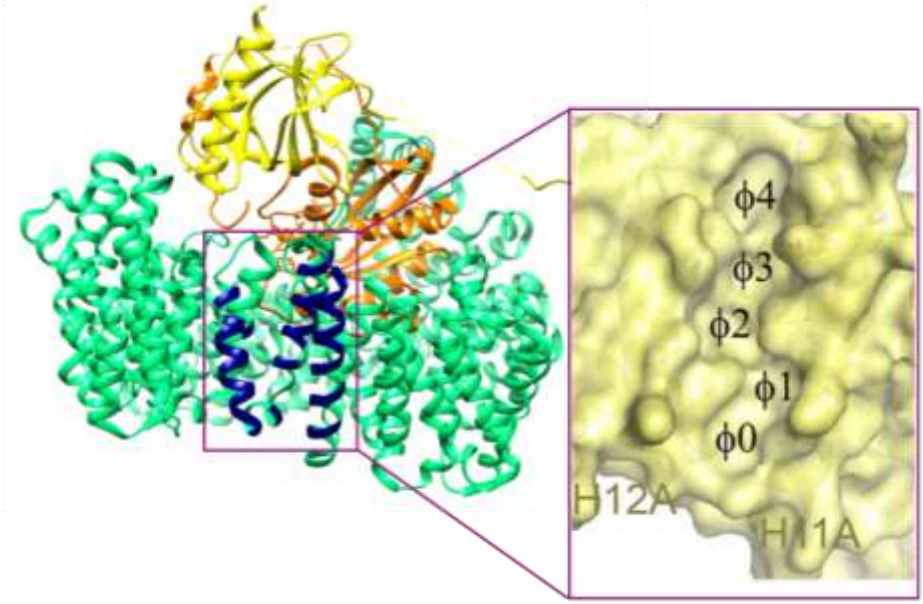


Figure 2.5. Representative image of the NES binding groove present on the CRM1 protein displaying the essential consensus sequence comprising of hydrophobic residues[62].

2.10. First generation inhibitor- Leptomycin B

The development of anticancer agents has always been steadily reliant upon the chemical constituents harboured by NPs as 60% of anti-tumoural compounds currently on the drug market used for the treatment of cancer has originated directly from natural sources. After the first classification as antifungal compounds, the potent anti-tumoural activity of Leptomycin was elucidated[89,90]. The first members of the Leptomycin family identified were Leptomycin A and B (LMB) which were first isolated from a strain of *Streptomyces*[91]. Biological studies further demonstrated the cogent inhibitory activity of the CRM1-dependent protein export from the cell nucleus. Figure 2.6 displays the structure of LMB, and its bound conformation to the hydrophobic NES binding groove of CRM1. LMB occupies the same space as four of five hydrophobic PKIαNES residues (ϕ0, ϕ1, ϕ2, ϕ3, and ϕ4). Displaying its mechanism of inhibition as it displaces the essential NES peptides required for binding to the CRM1 export protein, thus elucidating its broad spectrum of inhibition of nuclear export. LMB covalently modifies chromosomal region maintenance 1 (CRM1; exportin 1) at the nucleophilic sulfhydryl group of a reactive Cysteine residue by utilizing its α β-unsaturated δ-lactone, thus restricting export of protein cargoes that are dependent on this cleft by inhibiting the formation of the ternary CRM1/cargo substrate/Ran complex, or the binary complex CRM1/cargo substrate in the absence of Ran[76]

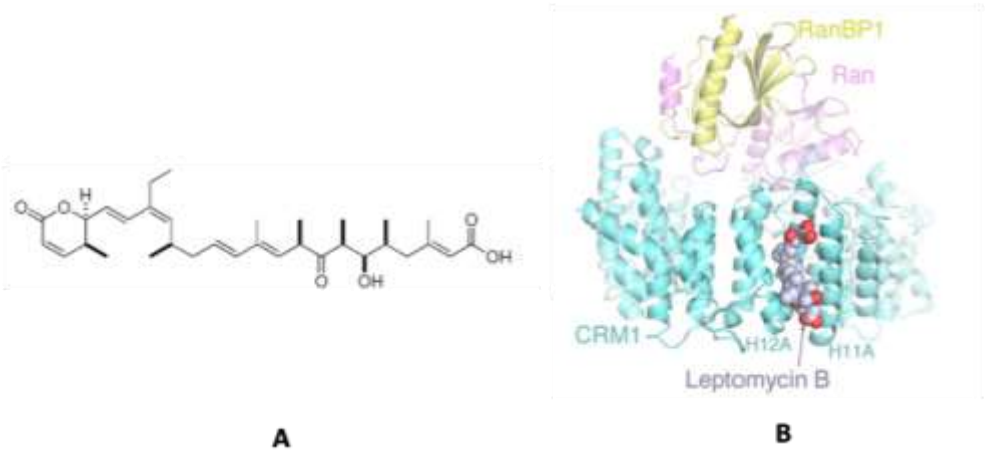


Figure 2.6. (A) Illustrates the structural composition of anti-fungal compound Leptomycin B, (B) displays the bound conformation of LMB to the hydrophobic NES binding groove of CRM1[68].

The strong dosage-limiting toxicities associated with the administration of LMB lead to minimal clinical benefits , resulting in profound anorexia and malaise, which are potentially off-target effects as LMB treated cells permanently retarded nuclear export posing not only lethal for cancer cells, but also for normal cells. Although dose-limiting toxicity prevented further development of Leptomycin B as an antitumor agent[92], Gademann and co-workers[41] saw the un-touched potential locked within the Anguinomycin core and went on to synthesize Anguinoymcin D, a close structural relative and derived analogues for further studies with the aim of investigating if the polyketide side chain mimics[93] the hydrophobic leucine-rich nuclear export signal of the cargo protein and is necessary for activity[41]. Anguinomycins A-D were reported to induce apoptosis in pRB-inactivated tumour cells, mediating inhibitory activity of immortalized cells, simultaneously inducing restraint of growth in normal cells, therefore limiting the total inhibition of the CRM1. The CRM1 protein plays an essential role in the nucleocytoplasmic transport, thus its presence remains eminent within normal cells. The selectivity displayed by Anguinomycin regarding their mode of action still remains unknown[94]. It was first hypothesized by Bonazzi and colleagues to completely omit the polyketide chain present in Anguinomycin D to derive the simplified structural analog SB 640. Experimental testing conducted on the parent compound Anguinomycin D and analog SB 640 displayed substantially inhibitory activity depicted by the analog SB 640 which contains a truncated polyketide chain displayed in Figure 2.7. Displaying a minimal reduction in the inhibitory activity as compared to the parent compound illustrating a retention of biological activity[95].

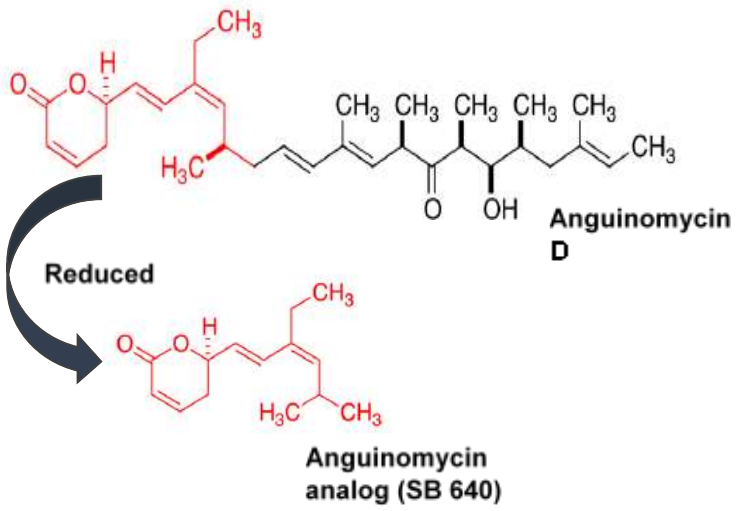


Figure 2.7. Structural comparison of Anguinomycin D isolated from Streptomyces sp. and its simplified structural analog Anguinomycin D analogue (were constructed using Chemdraw Ultra[96])

This is not the only experimental study revealing the potent activity of the derived analogue; another avid example is demonstrated by militarinone D, truncated to yield structurally simplified analogues with improved activity[46,41]. This hypothesis has[97] led to the prediction that truncated chemical structures might retain biological activity, as long as some of the key features were still maintained within the chemical fragment scaffold.

References

(1) Butler, M. S. (2004) The role of natural product chemistry in drug discovery. *J. Nat. Prod. 67*, 2141–2153.

(2) Cragg, G. M., and Newman, D. J. (2014) Natural Products: A Continuing Source of Novel Drug Leads. *Biochim. biophys acta J. 1830*, 3670–3695.

(3) Swinney, D. C., and Anthony, J. (2011) How were new medicines discovered ? *Nat. Rev. 10*, 507–519.

(4) Kirchmair, J., Williamson, M. J., Tyzack, J. D., Tan, L., Bond, P. J., Bender, A., and Glen, R. C. (2012) Computational Prediction of Metabolism: Sites, Products, SAR, P450 Enzyme Dynamics, and Mechanisms. *J. Chem. Inf. Model. 52*, 617–648.

(5) Lionta, E., Spyrou, G., Vassilatis, D. K., and Cournia, Z. (2014) Structure-Based Virtual Screening for Drug Discovery : Principles , Applications and Recent Advances. *Curr. Top. Med. Chem. 14*, 1923–1938.

(6) Lioa, C., Sitzmann, M., and Pugliese, A. (2012) Software and resources for computational medicinal chemistry. *Med. Chem. Res. 3*, 1057–1085.

(7) Browne, J. W., North, A. C. ., and Phillips, D. . (1969) A Possible Three-dimensional Structure of Bovine a-Lactalbumin based on that of Hen's Egg-White Lysozyme. *J. Mol. Biol. 42*, 65–86.

(8) Klein, J., Heal, J. R., Hamilton, W. D. O., Boussemghoune, T., Tange, T. Ø., Delegrange, F., Jaeschke, G., Hall, N., Street, T., and Cb, C. (2014) Yeast Synthetic Biology Platform Generates Novel Chemical Structures as Sca ff olds for Drug Discovery. *ACS Synth. Biol. 3*, 314–323.

(9) Madhusudhan, M. S., and Marti-Renom, M. A. (2011) Comparative Protein Structure Modeling, in *The Proteomics Protocol Handbook*, pp 831–837. Human Press Inc, Tokotowa, NJ.

(10) Xu, A. J., and Dias, D. (2013) Reinventing medicinal chemistry. *Pharmacol. Biotechnol. Life Sci. 9*.

(11) Kingston, D. G., and Newman, D. J. (2008) Natural products as drug leads: an old process or the new hope for drug discovery? *J. Med. Chem. 51*, 2589–2599.

(12) Albericio, F., and Kruger, G. (2011) Therapeutic peptides. *Futur. Med. Chem. 12*, 1527–1531.

(13) Koehn, F. E., and Carter, G. T. (2005) The evolving role of natural products in drug discovery. *Nat. Rev. 4*, 206–220.

(14) Lahlou, M. (2013) The Success of Natural Products in Drug Discovery. *Pharmacol. Pharm. 4*, 17–31.

(15) Amirghofran, Z. (2012) Herbal medicines for immunosuppression. *Iran. J. Allergy, Asthma Immunol. 11*, 111–119.

(16) Rajesh, E., Sankari, L. S., Malathi, L., and Krupaa, J. R. (2015) Naturally occurring products in cancer therapy. *J. Pharm. Bioallied Sci. 7*, S181-3.

(17) Ji, H.-F., Li, X.-J., and Zhang, H.-Y. (2009) Natural products and drug discovery. *Eur. Mol. Biol. Orgnanization. 10*, 194–200.

(18) Beutler, J. a. (2009) Natural Products as a Foundation for Drug Discovery. *Curr. Protoc. Pharmacol. 46*, 1–21.

(19) Atanasov, A. G., Waltenberger, B., Pferschy-wenzig, E., Linder, T., Wawrosch, C., Uhrin, P., Temml, V., Wang, L., Schwaiger, S., Heiss, E. H., Rollinger, J. M., Schuster, D., Breuss, J. M., Bochkov, V., Mihovilovic, M. D., Kopp, B., Bauer, R., Dirsch, V. M., and Stuppner, H. (2015) Discovery and resupply of pharmacologically active plant-derived natural products : A review. *Biotechnol. Adv. 33*, 1582–1614.

(20) Bathula, S. R., Akondi, S. M., Mainkar, P. S., and Chandrasekhar, S. (2015) Pruning of biomolecules and natural products (PBNP)": an innovative paradigm in drug discovery. *Org. Biomol. Chem. 13*, 6432–6448.

(21) Newman, D. J., and Cragg, G. M. (2016) Natural Products as Sources of New Drugs from 1981 to 2014. *J. Nat. Prod. 79*, 629–661.

(22) Sliwoski, G., Kothiwale, S., Meiler, J., and Lowe, E. W. (2014) Computational Methods in Drug Discovery. *Pharmacol. Rev.* 334–395.

(23) Cragg, G. M., and Newman, D. J. (2013) Natural products: A continuing source of novel drug leads. *Biochim. Biophys. Acta - Gen. Subj. 1830*, 3670–3695.

(24) Dias, D. A., Urban, S., and Roessner, U. (2012) A Historical Overview of Natural Products in Drug Discovery. *Metabolites 2*, 303–336.

(25) Bonazzi, S. (2009) Total Syntheses of Anguinomycins C & D , Synthetic Studies on Sporolides and Preparation of Eudistomin Derivatives : Biological Evaluation Against Cancer and Malaria. Swiss Federal Institue of Technology ETH Zurich.

(26) Kellenberger, E., Hofmann, A., and Quinn, R. J. (2011) Similar interactions of natural products with biosynthetic enzymes and therapeutic targets could explain why nature produces such a large proportion of existing drugs. *Nat. Prod. Rep. 40*, 3341–3350.

(27) David, B., Wolfender, J.-L., and Dias, D. A. (2014) The pharmaceutical industry and natural products: historical status and new trends. *Phytochem. Rev. 14*, 299–315.

(28) Spandl, R. J., Bender, A., and Spring, D. R. (2008) Diversity-oriented synthesis; a spectrum of approaches and results. *Org. Biomol. Chem. 6*, 1149–1158.

(29) Hong, J. (2014) Natural Product Synthesis at the Interface of Chemistry and Biology. *Chem. A Eur. J. Concept* 10204–10212.

(30) Franc, I., Lipinski, A., and Feeney, P. J. (1997) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings *23*.

(31) Lipinski, C. A., Lombardo, F., Dominy, B. W., and Feeney, P. J. (2001) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. drug Deliv. 46*, 3–26.

(32) Ou-Yang, S.-S., Lu, J.-Y., Kong, X.-Q., Liang, Z.-J., Luo, C., and Jiang, H. (2012) Computational drug discovery. *Acta Pharmacol. Sin. 33*, 1131–40.

(33) Szymański, P., Markowicz, M., and Mikiciuk-Olasik, E. (2012) Adaptation of high-throughput screening in drug discovery-toxicological screening tests. *Int. J. Mol. Sci. 13*, 427–452.

(34) Crane, E. A., and Gademann, K. (2016) Capturing Biological Activity in Natural Product Fragments by Chemical Synthesis. *Angew. Chem. Rev. Int. Ed. 55*, 2–23.

(35) Dixit, S., and Ali, H. (2010) Antioxidant Potential Some Medicinal Plants of Central India. *J. Cancer Ther. 2010*, 87–90.

(36) Kalyaanamoorthy, S., and Chen, Y. P. (2011) Structure-based drug design to augment hit discovery. *Drug Discov. Today 16*, 831–839.

(37) Bao, R., and Chan. (2011) Novel compounds in the treatment of lung cancer: current and developing therapeutic agents. *J. Exp. Pharmacol. 3*, 21–34.

(38) Rodrigues, T., Reker, D., Schneider, P., and Schneider, G. (2016) Counting on natural products for drug design. *Nat. Publ. Gr.* 1–10.

(39) Harvey, L. A., Edrada-Ebel, R., and Quinn, J. R. (2011) The re-emergence of natural products for drug discovery in the genomics era. *Biotechnol. Adv.* 1–52.

(40) Prota, A. E., Bargsten, K., Zuwerra, D., Field, J. J., and Diaz, J. F. (2013) Molecular Mechanism of Action of Microtubule-Stabilizing Anticancer Agents. *Science (80-.). 339*, 587–591.

(41) Gademann, K. (2015) Copy, Edit, and Paste: Natural Product Approaches to Biomaterials and Neuroengineering. *Acc. Chem. Res. 48*, 731–739.

(42) Dabydeen, D. A., Burnett, J. C., Bai, R., Verdier-pinard, P., Hickford, S. J. H., Pettit, G. R., Blunt, J. W., Munro, M. H. G., Gussio, R., and Hamel, E. (2006) Comparison of the Activities of the Truncated

Halichondrin B Analog NSC 707389 (E7389) with Those of the Parent Compound and a Proposed Binding Site on Tubulin. *Mol. Pharmacol. 70*, 1866–1875.

(43) Crane, E. A., and Gademann, K. (2016) Capturing Biological Activity in Natural Product Fragments by Chemical Synthesis. *Angew. Chemie Int. Ed. 55*, 2–23.

(44) Keller, T. H., Pichota, A., and Yin, Z. (2006) A practical view of " druggability ." *Curr. Opin. Chem. Biol. 10*, 10;357-361.

(45) Morse, S. S. (1995) Factors in the Emergence of Infectious Diseases. *Emerg. Infect. Dis. 1*, 7–15.

(46) Srivastava, V., Negi, A. S., Kumar, J. K., Gupta, M. M., and Khanuja, S. P. S. (2005) Plant-based anticancer molecules : A chemical and biological profile of some important leads. *Biorgan. Med. Chem. 13*, 5892–5908.

(47) Bozic, I., Allen, B., and Nowak, M. A. (2012) Dynamics of targeted cancer therapy. *Trends Mol. Med. 18*, 311–316.

(48) Torre, L., Siegel, R., and Jemal, A. (2012) Global Cancer. *Am. Cancer Soc.* 1–13.

(49) Goldstein, I., Madar, S., and Rotter, V. Cancer research , a field on the verge of a paradigm shift ? *Trends Mol. Med. 18*, 299–303.

(50) Noske, A., Weichert, W., Niesporek, S., Buckendahl, A., Sehouli, J., Dietel, M., and Denkert, C. (2008) Expression of the Nuclear Export Protein Chromosomal Region Maintenance / Exportin 1 / Xpo1 Is a Prognostic Factor in Human Ovarian Cancer. *Wiley Intersci. 112*.

(51) Huang, W., and Qiu, W. (2009) Prognostic value of CRM1in pancreas cancer. *Clin. Investments Med. 32*, 315–321.

(52) Shacham, S., and Mohammad, R. M. (2014) Selective Inhibitors of Nuclear Export Block Pancreatic Cancer Cell Proliferation and Reduce Tumor Growth in Mice. *Gastroenterology 144*, 447–456.

(53) Inoue, H., Kauffman, M., Shacham, S., Landesman, Y., Yang, J., Evans, C. P., and Weiss, R. H. (2013) CRM1 Blockade by Selective Inhibitors of Nuclear Export Attenuates Kidney Cancer Growth. *J. Urol. 189*, 2317–2326.

(54) Watt, P. J. Van Der, Maske, C. P., Hendricks, D. T., Parker, M. I., Denny, L., Govender, D., Birrer, M. J., and Leaner, V. D. (2009) The Karyopherin proteins , Crm1 and Karyopherin b 1 , are overexpressed in cervical cancer and are critical for cancer cell survival and proliferation. *Int. J. Cancer 1840*, 1829–1840.

(55) Zhou, F., Qiu, W., and Yao, R. (2013) CRM1 is a novel independent prognostic factor for the poor

prognosis of gastric carcinomas. *Med. Oncol. 30726*, 1–7.

(56) Shen, A., Ph, D., Wang, Y., and Zou, L. (2009) Expression of CRM1 in human gliomas and its significance in p27 expression and clinical prognosis. *Neurosurgery 65*, 153–160.

(57) Yao, Y., Dong, Y., Lin, F., Zhao, H. U. I., and Shen, Z. A. N. (2009) The expression of CRM1 is associated with prognosis in human osteosarcoma. *Oncol. Rep. 21*, 229–235.

(58) Lapalombella, R., Sun, Q., Williams, K., Tangeman, L., Jha, S., Zhong, Y., Goettl, V., Mahoney, E., Berglund, C., Gupta, S., Farmer, A., Mani, R., Johnson, J., Lucas, D., Mo, X., Daelemans, D., Sandanayaka, V., Shechter, S., Shacham, S., Kauffman, M., Chook, Y. M., and Byrd, J. C. (2012) Selective inhibitors of nuclear export (SINE) show that CRM1 / XPO1 is a target in chronic lymphocytic leukemia. *BLOOD J. Hematol. 6*, 1–23.

(59) Kojima, K., Kornblau, S. M., Ruvolo, V., Dilip, A., Duvvuri, S., Davis, R. E., Zhang, M., Wang, Z., Coombes, K. R., Zhang, N., Qiu, Y. H., Burks, J. K., Kantarjian, H., Shacham, S., Kauffman, M., and Andreeff, M. (2016) Prognostic impact and targeting of CRM1 in acute myeloid leukemia. *BLOOD J. Hematol. 121*, 4166–4175.

(60) Zhang, K., Wang, M., Tamayo, A. T., Shacham, S., Kauffman, M., Lee, J., Zhang, L., Ou, Z., Li, C., Sun, L., Ford, R. J., and Pham, L. V. (2013) Novel selective inhibitors of nuclear export CRM1 antagonists for therapy in mantle cell lymphoma. *Exp. Hematol. 41*, 67–78.e4.

(61) Schmidt, H. B., and Görlich, D. (2016) Transport Selectivity of Nuclear Pores, Phase Separation, and Membraneless Organelles. *Trends Biochem. Sci. 41*, 46–61.

(62) Pathria, G., Wagner, C., and Wagner, S. N. (2012) Inhibition of CRM1-Mediated Nucleocytoplasmic Transport : Triggering Human Melanoma Cell Apoptosis by Perturbing Multiple Cellular Pathways. *J. Invest. Dermatol. 132*, 2780–2790.

(63) Dickmanns, A., Monecke, T., and Ficner, R. (2015) Structural Basis of Targeting the Exportin CRM1 in Cancer. *Cells 4*, 538–568.

(64) Hanahan, D., Weinberg, R. A., and Francisco, S. (2000) The Hallmarks of Cancer. *Cell 100*, 57–70.

(65) Society, A. C. (2016) Cancer Facts & Figures. *AtlantaAmerican Cancer Soc.*

(66) Yee, H., Fung, J., Fu, S., Brautigam, C. A., and Chook, Y. M. (2015) Structural determinants of nuclear export signal orientation in binding to exportin. *eLIFE Biochem. Biophys. Struct. Biol. 6*, 1–19.

(67) Xu, D., Marquis, K., Pei, J., Fu, S., Cag, T., Chook, Y. M., and Grishin, N. V. (2015) Sequence analysis LocNES : a computational tool for locating classical NESs in CRM1 cargo proteins. *Bioinformatics 4*, 1–9.

(68) Sun, Q., Carrasco, Y. P., Hu, Y., Guo, X., Mirzaei, H., and Macmillan, J. (2013) Nuclear export inhibition through covalent conjugation and hydrolysis of Leptomycin B by CRM1. *Proceeding Natl. Acad. Sci. 110*, 1303–1308.

(69) Xu, D., Grishin, N. V, Min, Y., and Hetzer, M. (2012) NESdb : a database of NES-containing CRM1 cargoes. *Mol. Biol. Cell 23*, 3673–3676.

(70) Faustino, R. S., Nelson, T. J., and Terzic, A. (2007) Nuclear Transport : Target for Therapy. *Clin. Pharmacol. Ther. 81*, 880–886.

(71) Fornerod, M., Ohno, M., Yoshida, M., and Mattaj, I. W. (1997) CRM1 Is an Export Receptor for Leucine-Rich. *Cell 90*, 1051–1060.

(72) Monecke, T., Güttler, T., Neumann, P., Dickmanns, A., Görlich, D., and Ficner, R. (2007) Crystal Structure of the Nuclear Export Receptor CRM1 in Complex with Snurportin1 and RanGTP. *Science (80-.). 69*, 1–22.

(73) Xu, D., Marquis, K., Pei, J., Fu, S., Cag, T., Chook, Y. M., and Grishin, N. V. (2015) Sequence analysis LocNES : a computational tool for locating classical NESs in CRM1 cargo proteins. *Oxford Univ. Press 31*, 1357–1365.

(74) Gravina, G., Senapedis, W., McCauley, D., Baloglu, E., Shacham, S., and Festuccia, C. (2014) Nucleo-cytoplasmic transport as a therapeutic target of cancer. *J. Hematol. Oncol. 7*, 85.

(75) Hurtley, S. M. (1996) Nucleocytoplasmic Transport. *Science (80-.). 271*, 1477.

(76) Nguyen, K. T., Holloway, M. P., and Altura, R. A. (2012) The CRM1 nuclear export protein in normal development and disease. *Int. J. Biochem. 3*, 137–151.

(77) Mirski, S. E. L., Sparks, K. E., Friedrich, B., Köhler, M., Mo, Y. Y., Beck, W. T., and Cole, S. P. C. (2007) Topoisomerase II binds importin ?? isoforms and exportin/CRM1 but does not shuttle between the nucleus and cytoplasm in proliferating cells. *Exp. Cell Res. 313*, 627–637.

(78) Abraham, S. A., and Holyoake, T. L. (2013) Redirecting traffic using the XPO1 police. *Blood 122*, 2926–2928.

(79) Ohtani, N., Brennan, P., Gaubatz, S., Sanij, E., Hertzog, P., Wolvetang, E., Ghysdael, J., Rowe, M., and Hara, E. (2003) Epstein-Barr virus LMP1 blocks p16INK4a-RB pathway by promoting nuclear export of E2F4/5. *J. Cell Biol. 162*, 173–183.

(80) Henderson, B. R. (2000) Nuclear-cytoplasmic shuttling of APC regulates beta-catenin subcellular localization and turnover. *Nat. Cell Biol. 2*, 653–660.

(81) Santiago, A., Li, D., Zhao, L. Y., Godsey, A., and Liao, D. (2013) p53 SUMOylation promotes its

nuclear export by facilitating its release from the nuclear export receptor CRM1. *Mol. Biol. Cell 24*, 2739–52.

(82) Alt, J. R., Gladden, A. B., and Diehl, J. A. (2002) p21(Cip1) Promotes cyclin D1 nuclear accumulation via direct inhibition of nuclear export. *J. Biol. Chem. 277*, 8517–23.

(83) Wang, Y., Wang, Y., Xiang, J., Ji, F., Deng, Y., Tang, C., Yang, S., Xi, Q., Liu, R., and Di, W. (2014) Knockdown of CRM1 inhibits the nuclear export of p27 Kip1 phosphorylated at serine 10 and plays a role in the pathogenesis of epithelial ovarian cancer. *Cancer Lett. 343*, 6–13.

(84) Weinberg, R. A. (2014) Essay Coming Full Circle — From Endless Complexity to Simplicity and Back Again. *Cell 157*, 267–271.

(85) Yedavalli, V. R. K. (2013) CRM1 Organization of CRM1 Protein Cargo Transport Functions of CRM1 Rev Is an Export Substrate of CRM1. *AIDS*. Springer Science.

(86) Sun, Q., Chen, X., Zhou, Q., Burstein, E., Yang, S., and Jia, D. (2016) Inhibiting cancer cell hallmark features through nuclear export inhibition. *Signal Transduct. Target. Ther. 1*, 1041–1050.

(87) Dong, X., Biswas, A., Suel, K. E., Jackson, L. K., Martinez, R., Gu, H., and Chook, Y. M. (2012) Structural basis for leucine-rich nuclear export signal recognition by CRM1. *Nature 458*, 1136–1141.

(88) Dobbelstein, M., and Moll, U. (2014) Targeting tumour-supportive cellular machineries in anticancer drug development. *Nat. Rev. 13*, 179–196.

(89) Cook, A., Bono, F., Jinek, M., and Conti, E. (2007) Structural biology of nucleocytoplasmic transport. *Annu. Rev. Biochem. 76*, 647–671.

(90) Dechristopher, B. A., Fan, A. C., Felsher, D. W., and Wender, P. A. (2012) " Picolog ," a Synthetically-Available Bryostatin Analog , Inhibits Growth of MYC-Induced Lymphoma In Vivo. *Oncotarget 3*, 58–66.

(91) Jang, B. C., Muñoz-Najar, U., Paik, J. H., Claffey, K., Yoshida, M., and Hla, T. (2003) Leptomycin B, an inhibitor of the nuclear export receptor CRM1, inhibits COX-2 expression. *J. Biol. Chem. 278*, 2773–2776.

(92) Adachi, M. (2004) Leptomycin B. Tokyo.

(93) Nobuaki, K., Nobuaki, M., Hiroshi, T., Daisuke, F., Erwin, P. S., Wolff, B., Yoshida, M., and Horinouchi, S. (1999) Leptomycin B inactivates CRM1 ¯exportin 1 by covalent modification at a cysteine residue in the. *Proceeding Natl. Acad. Sci. 96*, 9112–9117.

(94) Bonazzi, S., Eidam, O., Gu, S., Wach, J., Zemp, I., Kutay, U., and Gademann, K. (2010) Anguinomycins and Derivatives : Total Syntheses , Modeling , and Biological Evaluation of the

Inhibition of Nucleocytoplasmic Transport. *J. Antimicrob. Chemother*. *132*, 1432–1442.

(95) Bonazzi, S., Eidam, O., Güttinger, S., Wach, J. Y., Zemp, I., Kutay, U., and Gademann, K. (2010) Anguinomycins and derivatives: Total syntheses, modeling, and biological evaluation of the inhibition of nucleocytoplasmic transport. *J. Am. Chem. Soc. 132*, 1432–1442.

(96) Cousins, Kimberley R. (California State University, S. B. (2005) ChemDraw Ultra 9.0. CambridgeSoft, 100 CambridgePark Drive, Cambridge, MA 02140. *J. Am. Chem. Soc*. *127*, 4115–4116.

(97) Yao, Xiao Z, Lui R, Li T, Jiao W, and Y. C. (2013) Aldehyde or a , b -Dibromoaldehyde with Isatins : An Efficient Synthesis of. *Chem. A Eur. J. 19*, 456–45

CHAPTER 3

3.0. Introduction to computational chemistry

3.1. Introduction

Computational chemistry is rapidly emerging as a subfield of theoretical chemistry, where the principal emphasis is focused on finding solutions to chemical related problems by employing the use of computers. Computational chemistry is not directly involved in developing theoretical methods, but rather in obtaining results relevant to chemical problems[1]. There is nevertheless a strong interplay between traditional theoretical and computational chemistry. The integration of computational chemistry may enable newly emerging as well as long plaguing problems, to be studied in the search for effective solutions. Data extracted from computational integrated calculations may reveal limitations and possible suggestions towards suitable improvisations that can be implemented in modern drug discovery[2]. Depending on the accuracy required, and the nature of the system at hand, it is now possible to obtain useful information for systems containing up to several thousand particles. The only limitation associated with computational chemistry may be the selection of an appropriate theory or method for a given problem, and the ability to evaluate the quality of the obtained results[3]. There are two standard methods emanating from computational chemistry, the first being to study the chemistry of molecules at an electronic level, known as quantum mechanics, the second being molecular dynamics, which neglects explicit electron treatment and focuses on classical laws of physics. This chapter outlines the range of computational and theoretical tools applied in this study.

3.2. Quantum mechanics

In 1900, German theoretical physicist, Max Plank, inadvertently gave rise to the field of quantum mechanics, as he discovered that energy is discharged in small packets (called quanta) and emitted in wavelengths[4,5,] Quantum mechanics is an essential constituent of computational chemistry, enabling the prediction of observable chemical properties. The fundamental law of quantum mechanics aims to illustrate that microscopic systems can be described by wave functions that capsulate and characterize all physical properties of a system [6]. Quantum mechanics is the branch of mechanics that focuses on the mathematical analysis of the motion and interaction of subatomic particles[7], principally dealing with the influence of electromagnetic forces on the movement of electrons[8]. To perceive the electronic behaviour in molecules and consequently of the structures and reaction of molecules, knowledge of quantum mechanics, particularly the Schrödinger equation[9]. By solving the Schrodinger equation in quantum chemistry, properties of a system in terms of a wave function can then be extracted. Quantum

mechanics can be integrated to provide a better understanding and predict large-scale phenomena, initiating fundamental calculation of electronic structure and interactions[1]

3.3. The Schrödinger equation

The Schrödinger equation is regarded as the fundamental core of physics, as it entails a descriptive analysis of quantum mechanical behaviour within a system[10]. It was initially introduced by Austrian physicist Erwin Schrödinger in 1926[10]. In mathematical physics, the Schrödinger's equation undertakes the same role as the Hamilton's laws of motion as one of the basic equations in non-relativistic quantum mechanics and non-relativistic classical mechanics respectively[11,12].

There are two types of Schrödinger's equation: the first is the time-dependent Schrödinger's equation, this being the most applied equation in computational chemistry[13], and defines the Hamiltonian operator as the accumulated value of the potential and kinetic energy. The second type is the time-independent Schrödinger's equation[14]. The simplest form of Schrodinger equation is presented as follows[15]:

$$\mathbf{H}\psi = \mathbf{E}\psi \qquad \textbf{Eq.1}$$

Where H denotes the molecular Hamiltonian, ψ a wave function that expounds the probability of the electron and nuclear within disclosed locations, and E depicts the energy of the system (*3, 5*). The molecular Hamiltonian is the sum of the kinetic (T) and potential (V) energy, which can be denoted as[16]:

$$\mathbf{H} = \mathbf{T} + \mathbf{V} \qquad \textbf{Eq. 2}$$

Particles are referred to as point masses, under the assumption that relativistic effects are not considered. The sum of kinetic and potential energy operators make up the composition of the Hamiltonian, which can then be presented in detail as:

$$\mathbf{H} = -\frac{h^2}{8\pi^2}\Sigma_i \frac{1}{mj}\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}\right) + \Sigma_i\Sigma_{<j}\left(\frac{e_i e_i}{r_{ij}}\right) \qquad \textbf{Eq.3}$$

Where *A* and *B* represents the nuclei: *i* and *j* are the electrons, M_A depicts the mass of nucleus *A*, m_e the mass of an electron, R_{AB} the distance between nuclei *A* and *B*, r_{ij} the distance between electrons *i* and *j*, Z_A the charge of nucleus *A*, r_{Ai} the distance between nucleus *A* and electron *i*[17]. The 1st term in Eq. 3 is the operator of kinetic energy of electrons, the 2nd term is the kinetic energy of the nuclei operator, the potential energy of electron-nuclei attractions operator is presented by term 3, the 4th term is the operator for potential energy of electron-electron repulsions, and the last term is the operator for potential energy of nuclei-nuclei repulsions[18]. However, the Schrodinger equation integrates a range of equations, and hence cannot be resolved for a molecular system other than H_2, therefore the implementation of the Born-Oppenheimer approximation is considered an applicable solution[5].

3.4. Born-Oppenheimer approximation

The Born-Oppenheimer approximation is considered an imperative commodity in finding a solution to the Schrödinger equation, where the coupling between the nuclei and electronic *motion* is often omitted[19]. This enables the nuclear parameters to be taken into consideration when solving the electronic part and for the resulting *potential energy surface* (PES) to be integrated in finding a solution for motion[20]. The Born-Oppenheimer approximation enables the use of the Schrodinger equation for a specific molecular system to be distinguished into two equations, namely the electron and nuclear equations from which the total energy of the system can be established[13]. The energy of a molecule is a function of the electron coordinates, but depends on the parameters of the nuclear coordinates, which define the molecular geometry. As nuclei are fixed, so the nuclear kinetic energy operator is neglected, as observed in Eq.4, permitting the statically distribution of electron input within a molecule to be determined [21,14].

$$\mathbf{T_{elec}} = -\frac{h^2}{8\pi^2 m}\ \Sigma_i^{electrons}\ \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}\right) \qquad \textbf{Eq.4}$$

Presented below is the Schrödinger equation for fixed nuclei electrons[17] which is incorporated in range of computational chemistry software:

$$\mathbf{H^{elec}}\ \boldsymbol{\varphi}^{elec}\ \mathbf{(r, R) = E^{eff}(R)}\ \boldsymbol{\varphi}^{elec}\mathbf{(r,R)} \qquad \textbf{Eq.5}$$

3.5. Potential energy surface

The potential energy surface (PES) is a mathematical function that correlates the energy of a molecule as a function of its geometry[21], thereby enabling a deeper visual insight of structural characterization, as derived from the latent relationship of potential energy versus a molecules geometry[22]. PES is utilized to decipher energy minima, as well as the states of transition that occur within chemical reactions, with respect to the position of the nuclei[22]. The Born-Oppenheimer approximation is invoked in molecular systems to generate the PES[23]. The concept of potential energy surface arises from variations in the mass and magnitude of the velocity between electrons and nuclei, a phenomenon defined by the Born-Oppenheimer approximation[24]. The phenomenon of the Born-Oppenheimer approximation stipulates the instantaneous variation in the position of the electrons with regard to the nuclei displacement, thereby permitting the depiction of potential energy surface, as the potential of atoms motion within a molecule or atoms in collision with each other is often referred to as the adiabatic motion[25,22,26].

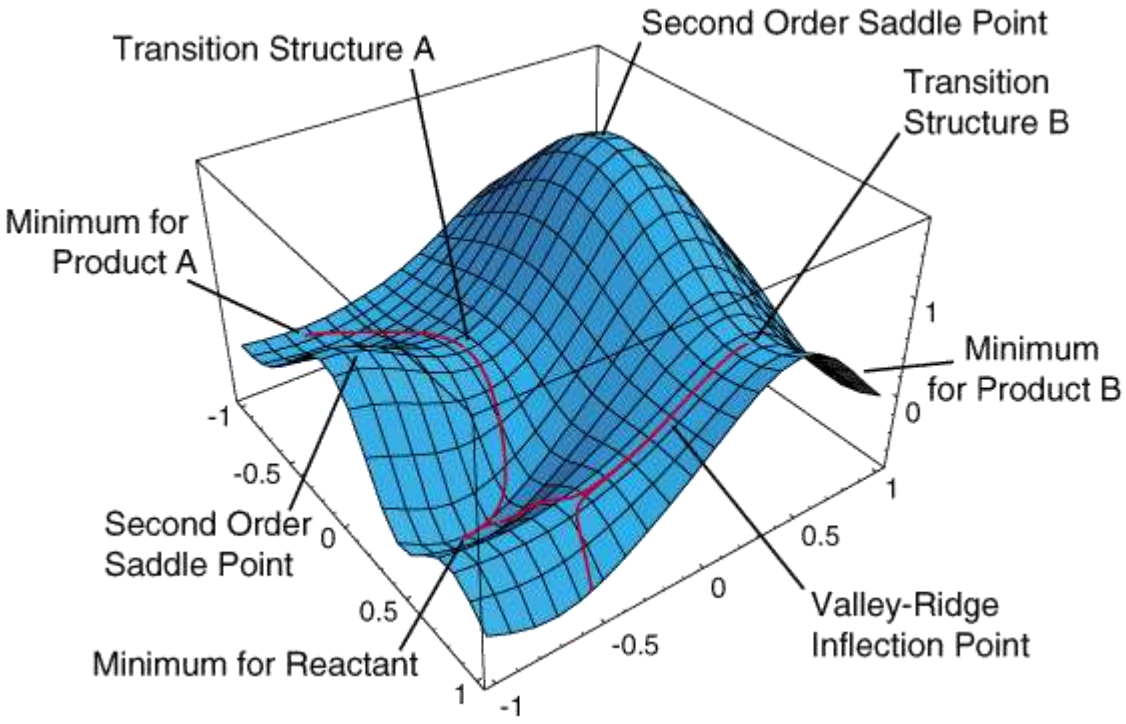


Figure 3.1. A Graphical representation of a two-dimensional potential energy surface[27].

The potential energy surface displays regions of unfavourable electronic interactions as high potential energy, as they depict molecular conformations or high-energy nuclear arrangements[27], with the regions of low energy being used as an indicator to display nuclear interactions presented as low energy molecular conformations and accompanied by favourable electronic interactions[28]. Reactants and products are commonly rendered as stable structures, resulting in relatively low potential energy, which is often associated with the minima on the PES of a reaction coordinate. Figure 3.1 indicates that the minima represents the compounds structure during a state of equilibrium, whereas the first order saddle point represents the transition state or activated complex. The saddle point is referred to as an indicative factor of particular modifications in nuclear arrangements, which may result in decreased potential energy, while the others lead to higher energy.

3.6. Molecular mechanics

Molecular mechanics (MM) is an approach that permits the parameterization of a simple algebraic expression for the total energy of a compound, disregarding the computation of wave function or the sum of the electron density[29,30]. MM is based on the classical laws of physics, which is able to compute a molecule's geometrical and transitional state of equilibrium, as well as its relative energies present among conformers[15]. In this method, there is an assumption that the electrons will attain their optimal position once the position of the nuclei is defined, and are hence not considered explicitly. The Born-

Oppenheimer approximation forms the basis of the assumption, as it depicts nuclei as heavier than electrons, thus making their motion negligibly compared to that of electrons[21].

MM is commonly applicable to larger systems, such as those in pharmaceutical and biological fields of research as they may exceed the use of more computer-intensive molecular orbital protocols[31]. The application to large systems can be attributed to suitable approximations (and its associated fast speed) and ability to determine the molecular conformation, or the atomic arrangement of a molecule, on the basis of its structural characterization and relative potential energies[32]. MM is an empirical method, and is reliant on force field parameters that comprise of a range of parameters extracted from experimental data. One of the simplest illustrations of MM is that of a-ball-and-spring model of atoms and molecules, with classical forces being present between them[8]. The total energy of a molecule, as defined by molecular mechanics, comprises of a range of contributing interactions that includes bond length and angles, torsions and non-bonded interactions, such as van der Waals and electrostatic contribution, as presented in the equation 6[33]:

$$\boldsymbol{E_{tot} = E_{str} + E_{bend} + E_{tors} + E_{vdw} + E_{elec}} \qquad \textbf{Eq. 6}$$

Where E_{tot} is the total energy, E_{str} is defined as the bond-stretching energy, and the energy contribution from angle-bending is denoted by E_{bend}. The torsional energy contribution is represented by the term E_{tors}, and the terms E_{vdw} and E_{elec} represent van der Waals and electrostatic energy contributions respectively. This equation in complex, with the parameters required to describe the characteristics of various molecules being referred to as the force field.

3.7. Force fields

A force field encompasses a set of functions and constants, defined as parameters, which can be applied to correlate the energy of the system in accordance with its particles[34]. The parameters aim to define the reference points and force constants, providing a description of PES for various types of molecular systems with contrasting degrees of freedom that result from the inclusion of attractive or repulsive interactions between atoms. Force fields such as AMBER[35], CHARMM[36], NAMD[37] and GROMOS[38] display popularity, and are commonly used to set parameters that can be applied to the simulation of biomolecules. The parameters incorporated in force fields are generated from the derivation of the ab initio method, or the semi-empirical quantum mechanical calculations. They can also be generated from experimental data, such as X-ray and electron diffraction, NMR and neutron spectroscopy[39,34]. The forces acting within a molecule vary with regard to each system, with the administration of each force field needing to be adjusted accordingly. The different force fields are associated with a range of strengths and weaknesses, relative to the data applied, allowing a particular problem to be dealt with.

The moderate low computational cost and accurate prediction protocols incorporated in the use of force fields highlights their use as an attractive option is molecular dynamics simulations and molecular mechanic calculations[40].There are a wide range of force fields that can be applied to a system, the force field of choice must be selected cautiously as each force field contains specific parameter that are designed for certain families of molecules[41]. Therefore it is crucial to choose the correct force field as it may increase the accuracy and applicability of theoretical structure function of your system of study. Force fields can only crudely approximate electrons interaction and hence cannot be used for bonding breaking nor formation calculations[42]. The AMBER force field[43] was employed in this study, by applying the General AMBER Force Field (GAFF) parameters accompanied by the standard AMBER[43] force field for the protein being introduced.

3.8. Molecular dynamics

Numerous studies have incorporated a vast range of molecular systems that include organic molecules in solution and biological macromolecules[44]. Molecular dynamics[45,46] and Monte Carlo[47] simulations are two techniques that provide useful insights into the structural, thermodynamic, and, for molecular dynamics, dynamical properties of systems in the condensed phase[48]. Molecular dynamics (MD) is a computational technique used for simulating intricate molecular systems at an atomic level, as well as for computing the motions of individual molecules[49]. The application of molecular dynamics is a sought after computational tool[50], as it enables the fluctuations that may occur in the motion of the system to be evaluated over a set period of time. MD enables the kinetic and thermodynamic properties of the molecular system to be determined[48]. The low-energy deformation states derived from the inherent use of MD simulations can be incorporated to inspect the conformational space present in a large restricted system[51,52]. If the initial geometry of a system is derived from experimental data stemming from X-ray or NMR structures, then MD techniques can be employed for sampling the conformational space[53]. In order to set a MD system for simulation, the force, in combination with the energy of all particles within the system, must be calculated[54]. MD integrates the use of Newton's equation of motion for atoms on an energy surface[55].

$$F_i = m_i \frac{d^2 r_i(t)}{dt^2} \qquad \textbf{Eq. 7}$$

3.9. Molecular docking

Molecular docking is an essential component of computational chemistry, which involves determining the most optimal position of two molecules with respect to each other. Molecular docking is utilized in structure-based drug design, and is often highlighted as one of the main contributing factors to the problems arising in global optimization[56]. The dynamic level of interaction displayed between ligand molecules and their receptors is often dependent upon the molecular recognition of the lock and key

mechanism[57]. Docking is often referred to as the positioning of a small molecule, such as an inhibitor or drug candidate, often referred to as a ligand, into the active site of macromolecules of known structural conformation[44]. These macromolecules include proteins, such as nucleic acid, receptor or enzyme. Docking is performed to generate the optimal conformational pose, as well as to determine the crude binding affinity between the protein-ligand interaction[58]. Being able to predict the binding of small molecules to target proteins plays a crucial role in structure-based drug design, as it enables the screening of virtual libraries[59] of "drug like" molecules, thereby assisting in next generation drug development[60]. The ligand-receptor binding energy is calculated as follow[61]:

$$E_{binding} = E_{target} + E_{ligand} + E_{target-ligand} \qquad \textbf{Eq. 8}$$

Numerous molecular docking programs are used for academic and commercial purposes[62], such as Dock[63], AutoDock GOLD[64], FlexX[65], -+ GLIDE[66], ICM[62] , Surflex and others. While each program displays sufficient suitability for precise docking, the docking program Autodock generates two-orders of magnitude in comparison to other programs, while maintaining substantial accuracy in its binding mode prediction[67]. The docking method used in this research study is the advanced version of AutoDock, AutoDock Vina[68]. The application of molecular docking in this study enabled the confirmation of the most optimal binding pose of both compounds Anguinomyicn D and derived analog SB 640 in the NES binding groove of the CRM1 protein[69] (detailed discussion in chapter 5).The binding affinity generated by many docking software's such as Autodock[63], neglect the presence of protons of the enzyme and inhibitors, thus the scores generated are often regarded as unreliable. MD takes into consideration protons and the solvent often water molecules. Thus MD results generates a more accurate binding affinity score as opposed to docking protocols[42].

3.10. Thermodynamic calculations

The embodiment of thermodynamics in computational chemistry enables a deeper understanding of chemical reactions, providing a platform to calculate molecular properties and its derived entities, and predicts the chemical reactivity. The essential role of thermodynamic calculations is highlighted, due to its current contribution in the field of quantum mechanics, allowing the optimization of geometry and calibration. Thermodynamic calculations aid in distinguishing the energy surface associated with a particular chemical reaction. The use of thermodynamics can retroactively justify minimization of energy, and its interconnection with energy surface may therefore provide ample knowledge based on the transition structure and reaction pathways[70]. Thermodynamic calculations assist in determining the binding free energy as an endpoint calculation, which provides indispensable information about the binding interaction that occurs between the ligand-protein complexes.

3.10.1 Binding free-energy calculations

Binding free energy calculations are eminent in the ligand-receptor complex formation[71,72], and have aided substantially in studies that include computational chemistry, thereby providing in-depth knowledge about drug design, protein structure determination[73] and protein-protein complexes[74,75]. Two popular methods used to estimate the free binding energy with success of small ligands to biological macromolecules are the Molecular Mechanics/ Poisson-Boltzmann Surface Area (MMPB-SA) and Molecular Mechanics/Generalized Born Surface Area (MMGB-SA) approaches[70,76]. MM/GB-SA and MM/PB-SA rely on molecular simulations of the ligand-protein complex to compute rigorous statistical-mechanical binding free energy within a specified force field[77]. Both approaches display favourable use, which can be attributed to their modular nature and lack of calculations that stem from training set integrating continuum solvation models merging with molecular mechanics calculations[78] . Each approach displays avid accuracy and computational effort between the empirical scoring and stringent alchemical perturbation methods, and can be compared in order to reproduce and rationalize experimental data[79]. The MM/PB-SA and MM/GB-SA methods are utilised to determine free energy decomposition, which can be meticulously ranged into various groups, depending on the groups of atoms or types of interactions from which they originated [74,80,]

The MM-PBSA employs a more rigorous algorithm than the MM/GB-SA, and simultaneously substitutes the MM/GB-SA model of electrostatics in water [81,82]. However, with regard to calculations incorporating protein-drug interaction ,including carbohydrates[83] and nucleic acids [84], the MM-GBSA method is favoured over the MM-PBSA [85]. The use of binding free-energy calculations can also be utilized to enhance the results of virtual screening and the docking of therapeutics drugs[58]. The binding free energy between the ligand and receptor highlighting the MM/GB-SA is given by[86]:

$$G_{BIND} = G_{COMPLEX} - G_{RECEPTOR} - G_{LIGAND} \quad \textbf{Eq. 9}$$

$$\Delta G_{BIND} = \Delta E_{MM} + \Delta G_{GBSA} - T\Delta S \quad \textbf{Eq.10}$$

$$\Delta E_{MM} = \Delta E_{INT} + \Delta E_{VDW} + \Delta E_{EEL} \quad \textbf{Eq. 11}$$

$$\Delta G_{GBSA} = \Delta G_{EGB} + \Delta G_{ESURF} \quad \textbf{Eq.12}$$

where ΔE_{MM} is the molecular mechanics energy of the system in a vacuum, ΔG_{GBSA} denotes the solvation free energy, $T\Delta S$ is the entropy, the sum of the bonded internal energy (ΔE_{Int}) is represented by ΔE_{MM}, non-bonded van der Waals (ΔE_{VDW}), electrostatic (ΔE_{EEL}) and ΔG_{GBSA} consists of polar contributions accounted for by the generalised born model (ΔG_{EGB}) and non-polar contributions

(ΔG_{ESURF})[87,88]. An additional preference for applying the MM/GB-SA method is due to its potential to aid in analysing per residue energy decomposition. [89,90]. The dynamic analysis of binding affinity aids in determining the approximate inhibitory activity of each inhibitor[91,92]. During the MM-GBSA binding energy calculation, the correct binding conformation of each ligand can be determined prior to the binding energy estimation.[93.] Herein, ligand-protein binding free energies were predicted using a MM/GB-SA approach.

3.11. PASS toxicity prediction

High levels of toxicity and adverse side effects displayed by current therapeutic still remains a challenge in the field of drug discovery. Therefore, alternative methods are required to improve toxicity prediction and safety assessment of potential drug candidates, prior to be being administered in clinical trials. Computational methods aim to complement *in vitro* and *in vivo* toxicity tests to potentially alleviate the burden associated with animal testing ultimately diminishing high associated expenses and time constraints. Incorporating the use of computational tools may permit the assessment of the toxicity profile of a chemical compound prior to synthesis.

The *in silico* tool referred to as PASS (Prediction of Activity Spectra for Substances)[94] is an integral resource that can be utilised to predict a spectrum of biological characteristics of a chemical fragment. PASS is an online based *in silico* prediction tool, it initiates the prediction of the biological activities of selected chemical compounds, their mechanisms of action and related levels of toxicity interacting with a single or a range of biological targets within an organism[94]. The proposed *in silico* prediction tool establishes the activity of a chemical compound based on the chemical constituents of the chemical fragment which extends further to reveal novel biological activities of selected phytochemical leads, their bioactive constituents and related side-effects[95].

The current version of PASS encompasses the prediction capacity of over 3750 pharmacological effects, biochemical mechanisms of action, specific toxicities levels and metabolic terms on the basis of structural formulae of drug-like substances with average accuracy >95%. Results obtained from the use of this *in silico* tool has been validated with the application of *in-vitro* as well as *in-vivo* assays, the results obtained from these studies correlate to those predicted by PASS online software[94]. The PASS online prediction was performed for parent compound Anguinomycin D and analog SB 640 (results are presented in chapter 5)

3.12. Homology modeling

In the midst of newly developing diseases, the crystal structure of a target protein is of utmost importance in the field of drug discovery and development[96]. While various techniques can be utilised

to generate the structure of macromolecule, which include the use of as X-ray crystallography, NMR spectroscopy and electron microscopy, they are often associated with multiple pitfalls. Homology modeling is gradually emerging as the sought after tool to be used to construct 3D macromolecule structure, as it is done with ease and accuracy in comparison to the other techniques[97]. Homology modeling enables the construction of a proteins structure by using its sequence as a reference template, of which the X-ray crystal structure is known. An accurate homology model depends on the existence, detection and quality of the known template structure from which it will be modelled. Although high-resolution structures are optimally extracted through X-ray crystallography, this approach is associated with a high cost increment, considerable experimental time and many trial runs[98]. Furthermore, some biologically important macromolecules lack X-ray crystal structures or high resolution 3D- structural properties, with reference to their protein sequence, the implementation of homology modeling resolves this challenge. An essential aim of drug discovery is to contrive bioactive molecules that are intended to target the disease condition, with minimal side effects, and hence are beneficial to the patient. The 3D structure of the target protein is a fundamental factor to obtaining full characterisation and exhibition of the mechanism of interaction on a structural and molecular level, elucidating the mode of action (MOA) of a drug molecule, and hence, greatly facilitate drug design. This study provides a precise protocol that can be followed for homology modeling to generate a protein model of high accuracy that can aid in developing therapeutic drugs. The role and application of homology modeling in next generation drug discovery is presented in Chapter 4.

References

(1) Breslow, R., and Tirrel, V. M. (2003) Beyond the Molecular Frontier: Challenges for Chemistry and Chemical Engineering. National Academy of Sciences, Yale.

(2) Ekins, S., Mestres, J., and Testa, B. (2007) In silico pharmacology for drug discovery : methods for virtual ligand screening and profiling. *Br. J. Pharmacol. 152*, 9–20.

(3) Grant, B. J., Gorfe, A. A., and Mccammon, J. A. (2009) Ras Conformational Switching : Simulating Nucleotide- Dependent Conformational Transitions with Accelerated Molecular Dynamics. *PLOS Comput. Biol. 5*, 1–10.

(4) Atkins, P., and Friedman, R. (2005) Molecular Quantum Mechanics Fourth Edition. Oxford University Press, New York.

(5) Jensen, F. (2007) Introduction to Computational chemistry. *Introd. to Comput. Chem.* John Wiley and Sons Ltd, West Sussex.

(6) Sauer, J., and Sierka, M. (2000) Combining Quantum Mechanics and Interatomic Potential Functions in Ab Initio Studies of Extended Systems. *J. Comput. Chem. 21*, 1470–1493.

(7) Amara, P., and Field, M. J. (2009) Combined Quantum Mechanical and Molecular Mechanical Potentials. *Oxford Univ. Press 2*, 1–9.

(8) Bornemann, F. A., Nettesheim, P., Schu, C., and Introduction, I. (1996) Quantum-classical molecular dynamics as an approximation to full quantum dynamics ¨. *J. Chem. Phys. 105*, 1074–1083.

(9) Tannor, J. D. (2010) introduction to quantum mechanics : a time dependent perspective. John Wiley and Sons, Houston.

(10) Schrodinger, E. (1926) An undulatory theory of the molecules of atoms and molecules. *Phys. Rev. 22*.

(11) Car, R. (2002) Introduction to Density-Functional Theory and ab-Initio Molecular Dynamics. *Quantum Struct. Act. relations 21*, 97–104.

(12) Tomasi, J., Mennucci, B., and Cammi, R. (2005) Quantum Mechanical Continuum Solvation Models. *Chem. Rev. 105*, 2999–3093.

(13) Shih, H.-W. (2009) Fundamentals of Computational Chemistry. Oxford University Press, West Sussex.

(14) Atkins, P., and Paula, de J. (2006) ATKINS physical chemistry. W.H Freeman and Company, New York.

(15) Lewars, E. (2004) Computational Chemistry: An introduction to the theory applications of molecular and quantum mechanics. Kluwers Academic Publishers, Moscow.

(16) Quinn, R., and Schimdt, B. (2003) The theory behind Computational Chemistry. *PLOS Comput. Biol. 4*, 34–68.

(17) Schroeder, D. V. (2016) The Schrodinger Equations, in *Techniques in Computational Chemistry*, pp 1–5.

(18) Kruglyak, A. Y., and Whitman, D. . (1963) Calculations of Quantum chemistry. Russia.

(19) Bechstedt, F. (2015) Born-Oppenheimer Approximation, in *Many-Body Approach to Electronic Excitations*, pp 3–12.

(20) Bechstedt, F. (2015) Born-Oppenheimer Approximation. *Springer 5*, 3–12.

(21) Cramer, C. J. (2004) Essentials of computational chemistry second edition: Theories and models. John Wiley and Sons Ltd, West Sussex.

(22) Elsawy, K. M., Hodgson, M. K., and Caves, L. S. D. (2005) The physical determinants of the DNA conformational landscape : an analysis of the potential energy surface of single-strand dinucleotides in the conformational space of duplex DNA. *Nucleic Acids Res. 33*, 5749–5762.

(23) Jaquet, R. (2002) Introduction to potential energy surfaces and graphical interpretation.

(24) Lewars, E. G. (2011) The Concept of the Potential Energy Surface. *Comput. Chem. 6*, 9–13.

(25) Hratchian, P. H., and Schlegel, H. B. (2005) Finding minima, transition states, and following reaction pathways on ab intio potential energy surfaces, in *Theory and Applications of Computationl Chemsitry: The first forty years*, pp 195–249.

(26) Truhlar, D. G. (2001) Potential Energy Surfaces. Academic Press, New York.

(27) Sherrill, C. D. (2009) Potential Energy Surfaces. Georgia.

(28) Long, J. (2010) Use of potential energy surfaces (PES) in spectroscopy and reaction dynamics. University of Iceland.

(29) Lin, H., and Truhlar, D. G. (2007) QM / MM : what have we learned , where are we , and where do we go from here ? *Theor. Chem. Acc. 117*, 185–199.

(30) Burke, K., and History, I. I. A. B. (2012) Perspective on density functional theory. *J. Chem. Phys. 150901*, 1–9.

(31) Dejaegere, A., Lafont, V., Schaefer, M., and Stote, R. H. (2007) Protein – Protein Recognition and Interaction Hot Spots in an Antigen – Antibody Complex : Free Energy Decomposition Identifies “‘

Efficient Amino Acids .'" *PROTEINS Struct. Funct. Genet. 434*, 418–434.

(32) Kwan, E., Wiley, S. M., and Wiley, C. J. (2010) Introduction to Computational Chemistry. London.

(33) Vanommeslaeghe, K., Guvench, O., and Mackarellel, D. A. (2015) Molecular Mechanics. *Curr. Protoc. Comput. drug Des. 20*, 3281–3292.

(34) González, M. A. (2011) Force fields and molecular dynamics simulations. *EDP Sci. 12*, 169–200.

(35) Case, D. A., Cheatham, T. E., Darden, T. O. M., Gohlke, H., Luo, R. A. Y., Merz, K. M., Onufriev, A., Simmerling, C., Wang, B., and Woods, R. J. (2005) The Amber Biomolecular Simulation Programs. *J. Comput. Chem. 26*, 1668–1688.

(36) Brooks, B. R., Iii, C. L. B., Mackerell, A. D., Nilsson, L., Petrella, R. J., Roux, B., Won, Y., Archontis, G., Bartels, C., Boresch, S., Caflisch, A., Caves, L., Cui, Q., Dinner, A. R., and Feig, M. (2009) CHARMM : The Biomolecular Simulation Program. *J. Comput. Chem. 30*.

(37) Phillips, J. C., Braun, R., Wang, W. E. I., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R. D., and Poincare, H. (2005) Scalable Molecular Dynamics with NAMD. *J. Comput. Chem. 26*, 1781–1802.

(38) Bakowies, D., Baron, R., Christen, M., Hu, P. H., Geerke, D. P., Heinz, T. I. M. N., Kastenholz, M. A., and Gunsteren, W. F. V. A. N. (2005) The GROMOS Software for Biomolecular Simulation :GROMOSO5. *J. Comput. Chem. 26*.

(39) Abildgaard, J. (1998) Quantum Chemical Models in Molecular Structure Elucidation. W.H Freeman and Company, Roskilde.

(40) Cusack, P. (2002) Molecular Force Fields. New Jersey.

(41) Mackerell, A. D. (2010) Protein Force Fields. Baltimore, USA.

(42) Honarparvar, B., Govender, T., Maguire, G. E. M., Soliman, M. E. S., and Kruger, H. G. (2014) Integrated Approach to Structure-Based Enzymatic Drug Design : Molecular Modeling , Spectroscopy , and Experimental Bioactivity. *Chem. Rev. 114*, 493–537.

(43) Pearlman, D. A., Case, D. A., Caldwell, J. W., Ross, W. S., Cheatham, T. E., Debolt, S., Ferguson, D., Seibel, G., and Kollman, P. (1995) AMBER , a package of computer programs for applying molecular mechanics , normal mode analysis , molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. *Comput. Phys. Commun. 91*, 1–41.

(44) Ramanchandran, K. I., Deepa, G., and Namboori, K. (2008) Computational chemistry and molecular modeling. Springer-Verlag Berlin Heidelberg, Coimbatore.

(45) Petruk, A. A., Defelipe, L. A., Marti, M. A., and Turjanski, A. G. (2013) Molecular Dynamics

Simulations Provide Atomistic Insight into Hydrogen Exchange Mass Spectrometry Experiments. *J. Chem. Theory Comput. 9*, 658–669.

(46) Nad, S. (2007) Molecular dynamics simulations and their application to four-stranded DNA. *Methods 43*, 278–290.

(47) Neyts, E. C., Bogaerts, A., Carlo, M., Molecular, K., Monte, Á., and Long, C. Á. (2013) Combining molecular dynamics with Monte Carlo simulations : implementations and applications. *Theor. Chem. Acc. 132*, 1–12.

(48) Petrenko, R. (2010) Molecular Dynamics. *InEncyclopedia Life Sci.*

(49) Amadei, A., Linssen, A. B. M., and Berendsen, H. J. C. (1993) Essential Dynamics of Proteins. *PROTEINS Struct. Funct. Genet. 425*, 412–425.

(50) Mhlongo, N. N., and Soliman, M. E. S. (2015) RSC Advances Single H5N1 in fl uenza A neuraminidase mutation develops resistance to oseltamivir due to distorted conformational and drug binding landscape : multiple molecular dynamics analyses †. *RSC Adv. 5*, 10849–10861.

(51) Hehre, W. J. (2003) A Guide to Molecular Mechanics and Quantum Chemical Calculations. Wavefunction, Inc, Irvine.

(52) Road, H. (1976) Theoretical Studies of Enzymic Reactions. *J. Mol. Biol. 103*, 227–249.

(53) Dewar, M. J. S., Zoebisch, E. G., Healy, E. F., and Stewart, J. J. P. (1993) A quantum mechanical molecular model. *J. Am. Chem. Soc. 49*, 5003.

(54) Rogers, D. W. (2003) Computational Chemistry Using the PC Third Edition. John Wiley and Sons, Inc, New Jersey.

(55) Adcock, S. A., and Mccammon, J. A. (2006) Molecular Dynamics : Survey of Methods for Simulating the Activity of Proteins. *Chem. Rev. 106*, 1589–1615.

(56) Eweas, A. F., Maghrabi, I. A., and Namarneh, A. I. (2014) Advances in molecular modeling and docking as a tool for modern drug discovery. *Der Pharmceutical Chem. 6*, 211–228.

(57) Du, X., Li, Y., Xia, Y., Ai, S., Liang, J., Sang, P., and Ji, X. (2016) Insights into Protein – Ligand Interactions : Mechanisms , Models , and Methods. *Int. J. Mol. Sci. 17*, 1–34.

(58) Aalten, D. M. F. Van, Findlay, J. B. C., and Amadei, A. (1995) Essential dynamics of the cellular retinol-binding protein evidence for ligand-induced conformational changes. *Protein Eng. 8*, 1129–1135.

(59) Talele, T. T., Khedkar, S. A., and Rigby, A. C. (2010) Successful Applications of Computer Aided Drug Discovery : Moving Drugs from Concept to the Clinic. *Curr. Top. Med. Chem. 10*, 127–141.

(60) Young, D. C. (2001) CHEMISTRY COMPUTATIONAL CHEMISTRY A Practical Guide for Applying Techniques to Real-World Problems. John Wiley and Sons, Inc, Brisbane.

(61) Nervall, M. (2007) Binding Free Energy Calculations on Ligand-Receptor Complexes Applied to Malarial Protease Inhibitors. Uppsala Universitet.

(62) Cross, J. B., Thompson, D. C., Rai, B. K., Baber, J. C., and Fan, K. Y. (2009) Comparison of Several Molecular Docking Programs : Pose Prediction and Virtual Screening Accuracy. *J. Chem. Inf. Model. 49*, 1455–1474.

(63) Ewing, T. J. A., Makino, S., Skillman, A. G., and Kuntz, I. D. (2001) DOCK 4 . 0 : Search strategies for automated molecular docking of flexible molecule databases. *J. Comput. Aided. Mol. Des. 15*, 411–428.

(64) Verdonk, M. L., Cole, J. C., Hartshorn, M. J., Murray, C. W., and Taylor, R. D. (2003) Improved Protein – Ligand Docking Using GOLD. *PROTEINS Struct. Funct. Genet. 623*, 609–623.

(65) Joseph-mccarthy, D., and Alvarez, J. C. (2003) Automated Generation of MCSS-Derived Pharmacophoric DOCK Site Points for Searching Multiconformation Databases. *PROTEINS Struct. Funct. Genet. 202*, 189–202.

(66) Friesner, R. A., Banks, J. L., Murphy, R. B., Halgren, T. A., Klicic, J. J., Mainz, D. T., Repasky, M. P., Knoll, E. H., Shelley, M., Perry, J. K., Shaw, D. E., Francis, P., and Shenkin, P. S. (2004) Glide : A New Approach for Rapid , Accurate Docking and Scoring . 1 . Method and Assessment of Docking Accuracy. *J. Med. Chem. 47*, 1739–1749.

(67) Vajda, S., and Kozakov, D. (2009) Convergence and combination of methods in protein – protein docking. *Curr. Opin. Struct. Biol. 19*, 164–170.

(68) Trott Oleg, O. J. A. (2011) AutoDock Vina: improving the speed and accuracy of docking with a new scoring function,efficint optimization and multithreading. *NIH Public Access 31*, 455–461.

(69) Crane, E. A., and Gademann, K. (2016) Capturing Biological Activity in Natural Product Fragments by Chemical Synthesis. *Angew. Chemie Int. Ed. 55*, 2–23.

(70) Wang, J., Hou, T., and Xu, X. (2006) Recent Advances in Free Energy Calculations with a Combination of Molecular Mechanics and Continuum Models. *Curr. Protoc. Comput. drug Des. 7*, 287–306.

(71) Kamerlin, C. L. S., Harancyzk, M., and Warshel, A. (2009) Progress in Ab Initio QM / MM Free-Energy Simulations of Electrostatic Energies in Proteins : Accelerated QM / MM Studies of p K a , Redox Reactions and Solvation Free. *J. Phys. Chem. 113*, 1253–1272.

(72) Kollman, P. (1993) Free Energy Calculations: Applications to Chemical and Biochemical

Phenomena. *Chem. Rev. 93*, 2395–2417.

(73) Homeyer, N., and Gohlke, H. (2012) Free Energy Calculations by the Molecular Mechanics Poisson À Boltzmann Surface Area Method. *Mol. Inf. 31*, 114–122.

(74) Gohlke, H., Kiel, C., and Case, D. A. (2003) Insights into Protein – Protein Binding by Binding Free Energy Calculation and Free Energy Decomposition for the Ras – Raf and Ras – RalGDS Complexes. *J. Mol. Biol. 2836*, 891–913.

(75) Hou, T., Chen, K., Mclaughlin, W. A., Lu, B., and Wang, W. (2006) Computational Analysis and Prediction of the Binding Motif and Protein Interacting Partners of the Abl SH3 Domain. *PLOS Comput. Biol. 2*.

(76) Genheden, S., and Ryde, U. (2016) Expert Opinion on Drug Discovery The MM / PBSA and MM / GBSA methods to estimate ligand-binding affinities The MM / PBSA and MM / GBSA methods to estimate ligand-binding affinities. *Expert Opin. Drug Discov. 441*.

(77) Wang, J., Hou, T., Li, Y., and Wang, W. (2012) Assessing the performance of the MM/PBSA and MM/GBSA medthods: I. The accuracy of binding free energy calculations based on molecular dynamics simulations. *J. Chem. Inf. Model. 51*, 69–82.

(78) 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. *J. Chem. Inf. Model. 51*, 69–82.

(79) Zoete, V., Meuwly, M., and Karplus, M. (2005) Study of the Insulin Dimerization : Binding Free Energy Calculations and Per-Residue Free Energy Decomposition. *PROTEINS Struct. Funct. Genet. 93*, 79–93.

(80) Hou, Tingjun, Li Nan, W. W. (2013) Characterization of Domain–Peptide Interaction Interface: Prediction of SH3 Domain-Mediated Protein–Protein Interaction Network in Yeast by Generic Structure-Based Models. *NIH Public Access 11*, 2982–2995.

(81) Case, D. A. (2001) Theory and Applications of the Generalized Born Solvation Model in 275–291.

(82) Onufriev, A., Bashford, D., and Case, A. D. (2016) Modification of the Generalized Born Model Suitable for Macromolecules. *J. Phys. Chem. 2000*, 3712–3720.

(83) Feig, M., and Iii, C. L. B. (2002) Evaluating CASP4 Predictions With Physical Energy Functions *245*, 232–245.

(84) Wang, J., and Deng, Y. (2006) Absolute Binding Free Energy Calculations Using Molecular Dynamics Simulations with Restraining Potentials *91*, 2798–2814.

(85) Sun, H., Li, Y., Tian, S., and Hou, T. (2014) and MM / GBSA methodologies evaluated by various simulation protocols using PDBbind data set † 16719–16729.

(86) Decherchi, S., Masetti, M., Vyalov, I., and Rocchia, W. (2015) Implicit solvent methods for free energy estimation. *Eur. J. Med. Chem. 16*, 27–42.

(87) Genheden, S., and Ryde, U. (2015) The MM / PBSA and MM / GBSA methods to estimate ligand-binding affinities. *Expert Opin. Drug Discov. 10*, 449–461.

(88) Cappel, D., Hall, M. L., Lenselink, E. B., Qi, J., Bradner, J. E., Sherman, W., Cappel, D., Hall, M. L., Lenselink, E. B., Beuming, T., Qi, J., and Bradner, J. (2016) Relative Binding Free Energy Calculations Applied to Protein Homology Models. *J. Chem. Inf. Model. 5*, 1–15.

(89) Huang, Y. M., Chen, W., Potter, M. J., and Chang, C. A. (2012) Insights from Free-Energy Calculations : Protein Conformational Equilibrium , Driving Forces , and Ligand-Binding Modes *103*, 342–351.

(90) Genheden, S., and Ryde, U. (2016) Expert Opinion on Drug Discovery The MM / PBSA and MM / GBSA methods to estimate ligand-binding affinities The MM / PBSA and MM / GBSA methods to estimate ligand-binding affinities. *Expert Opin. Drug Discov. 441*, 449–461.

(91) Chang, C. A., Chen, W., and Gilson, M. K. (2007) Ligand configurational entropy and protein binding. *J. Proceeding Natl. Acad. Scienes 104*, 1534–1539.

(92) Wright, D. W., Hall, B. A., Kenway, O. A., Jha, S., and Coveney, P. V. (2014) Computing Clinically Relevant Binding Free Energies of HIV - 1 Protease Inhibitors. *Jounral Chem. theroy Comput. 10*, 1228–1241.

(93) Srivastava, M., Singh, H., and Naik, K. P. (2010) Molecular modeling evaluation of the antimalarial activity of artemisinin analogues : molecular docking and rescoring using Prime / MM-GBSA Approach. *Curr. Res. J. Biol. Sci.* 83–102.

(94) Kumar, R., and Damanpreet, G. (2011) CHEMISTRY PASS-assisted exploration of new therapeutic potential of natural products. *Acc. Chem. Res. 20*, 1509–1514.

(95) Gu, J., Gui, Y., Chen, L., Yuan, G., Lu, H., and Xu, X. (2013) Use of Natural Products as Chemical Library for Drug Discovery and Network Pharmacology. *Peer-reviewed open accesses J. 8*, 1–10.

(96) Ashburn, T. T., and Thor, K. B. (2004) DRUG REPOSITIONING : IDENTIFYING AND DEVELOPING NEW USES FOR EXISTING DRUGS. *Nat. Rev. 3*, 673–683.

(97) Forrest, L. R., Tang, C. L., and Honig, B. (2006) On the Accuracy of Homology Modeling and Sequence Alignment Methods Applied to Membrane Proteins. *Biophys. J. 91*, 508–517.

(98) Forrest, L. R., Tang, C. L., and Honig, B. (2006) On the Accuracy of Homology Modeling and Sequence Alignment Methods Applied to Membrane Proteins. *Biophys. J. 91*, 508–517

CHAPTER 4

Homology Modelling in Drug Discovery- An update on the last decade

Geraldene Munsamy[a] and Mahmoud E. S. Soliman[a*]

[a] Molecular Modelling and Drug Design Research Group, 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 031260 8048, Fax: +27 031 260 7872

Webpage: http://soliman.ukzn.ac.za/

4.1. Abstract

The continual evolution of the world is chaperoned by the emergence of new diseases, which presents itself at every turn therefore determining the 3D structure of a protein drug target becomes a crucial factor in the process of drug discovery and design. The 3D structure of a protein plays a critical role when establishing the functional domain of a protein, enabling the structural dynamic interactions with specified ligands and proteins to be studied and understood on a molecular level. The essential role of Homology modeling also regarded as comparative modeling, is described in this review as its use enables the provision of low-resolution structures that may aid molecular biologists, and pharmaceutical scientists with considerable insight regarding the spatial conformation of important residues within the protein structure providing a template for the design of new innovative drugs. This reviews provides a conclusive route map of the process of homology modeling that can be followed, accompanied by technical guidelines and tools that can be utilised. This review highlights the features of each tool enabling the construction of the most accurate model that may aid in next generation drug discovery.

4.2. Introduction

There has always been a steadfast focus on the field of drug design and development, focusing on structural biological studies of protein-drug interactions[1,2]. Understanding the underlying mechanism of protein-drug interaction is pivotal, as it provides insight into the structural features that are prominent for ligand affinity[3], drug specificity and optimization during these interactions[4,5] to ultimately ensure optimal effectivity of the therapeutic drug. The continual use of homology modeling for the generation of three-dimensional (3D) protein structures[6] has molded the way to develop docking and structure-based virtual screening protocols[7]. These 3D generated models are being sought after to gain intricate information behind the mechanism of interaction between drug-protein complexes such as binding mode analysis[8]. In addition, they provide indispensable insight into a protein's 3D structural and mechanistic molecular functions. In the last decade, homology modeling has transitioned into the most popular *in silico* tool for generation of three-dimensional (3D) structures of molecular targets [9,10]. Homology modeling also known as comparative modeling uses homologous sequences with known 3D structures for the modeling and prediction of the structure of a target sequence[11,12].

Due to the combined efforts of experimental structural biologists and genomics, an increasing fraction of protein families has at least one member with a known experimental structure present in the Protein Data Bank (PDB)[13,14]. This is accompanied by the development of sensitive and precise HMM profile methods[15], prompting researchers to grasp the availability of sequence information to aid in the detection of remote template relationships[16]. The recent advances in homology modeling, distinctly in identification and alignment of reference sequences in relation to the template structures, distant homologues, modelling of loops and side chains, 3D model optimisation and validation contributes substantially to the consistent prediction of a protein's structure which were not possible even several years ago[10, 17,18]. The core essence underlying the use of homology modelling to predict the 3D structure

of a protein relies on the use of the target sequence to be used in comparison to other known structures (the templates), this notion implies that similar sequences (evolutionary related) display similarity in their structures[10].

A crucial role of homology modeling amongst many others has been the generation of 3D molecular targets of cancer[19] in a bid to develop and discover effective therapeutics drugs to attenuate this malignant disease[20].Recent studies, highlight the utilization of homology modeling for the construction of the 3D structure of RNA-dependent RNA-polymerase of the Ebola virus[21] elucidating the instrumental role of homology modeling in drug discovery[22]. Another key example of the development of new drug targets constructed utilizing homology is the NS5 protein of the Zika virus which has been identified as a global health threat by the World Health Organization (WHO)[23]. Current technologies such as X-ray difractometry[24] and Nuclear Magnetic resonance (NMR) are utilized for experimental illustration of the 3D structure of a protein [25], however these tools are associated with a range of limitations such as an increase in cost and time consumption[26, 27]. Despite prominent advances in X-ray crystallography, NMR spectroscopy, and a wide range of other structural identification tools, the lack of newly emerging 3D protein structures, still remains a challenge in drug discovery. The potential use of homology modeling for the generation of these missing protein 3D structures is an attractive option as its use is often associated with low-cost and time effective protocol[28].

This review provides intricate information of some of the most popular homology modeling tools and software's that have been utilized for over a decade. It also highlights the essential role homology modeling has played in the drug discovery research along with some of the challenges that maybe associated with its application. In addition this review provides technical guidelines that can be incorporated to achieve the most accurate model. This review also provides a clear outline of the protocol of homology modeling and steps that can be followed, presented in Figure 4.1.

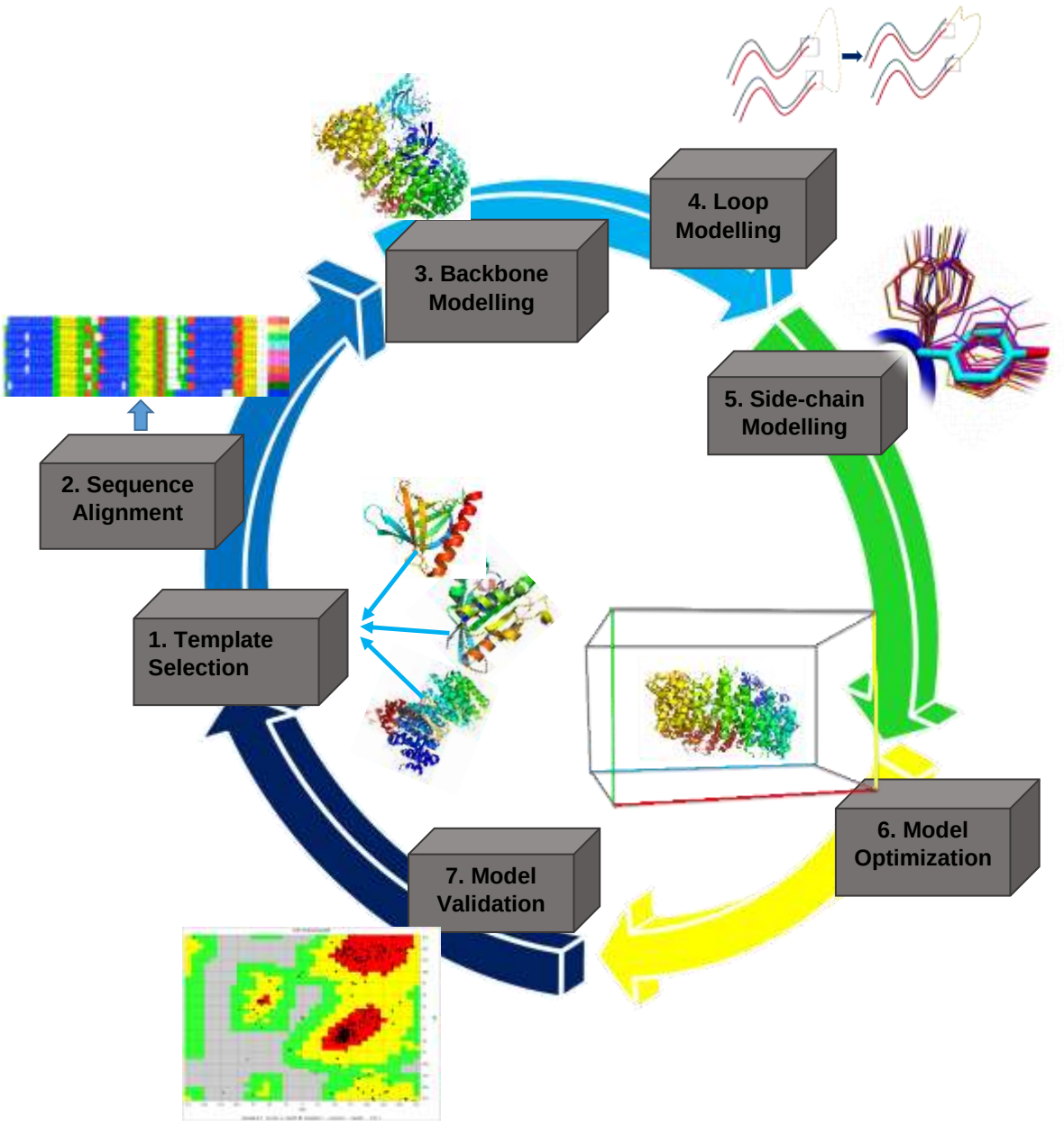


Figure 4. 1. General Protocol of 3D Protein Model generation by Homology Modeling

4.3. Protocol of Homology Modeling

There is a standard protocol of homology modeling which is adopted, with minor variations occurring in the validation and optimization steps of homology modeling. Homology modeling procedure consists of the following steps:

4.3.1. Template recognition – Comparative modeling is initiated by the identification of at least a single protein[29], with a known 3D experimental structure, that is in relation to the target sequence

serving as the template sequence[30,31]. A protein's 3D structure may be largely determined by its amino acid sequence as proteins from the same homologous line display greater conservation in their 3D structure in comparison to their amino-acid sequences[32, 33, 34,35,36]. Elucidating similarity between two proteins which may be detectable at sequence level, and may therefore imply shared structural similarity among these proteins [37,38].

The target sequence is rendered as the query sequence in the search of template sequences against databases such as PDB (Protein Data Bank)[39] for known protein structures using the target sequence as a query. A search is conducted as each structure in the database is compared to that of the target sequence. If the target sequence is not known, popular algorithm tools such as BLAST (Basic Local Alignment Search Tool) which incorporate the pairwise sequence-sequence comparison[40], structural

protein sequence databases such as GenBank[41] NCBI[36], protein database FastA[42], or protein identification resource UniProt[43] can be utilized to search for structures in the data banks of amino acid sequences that could be compared to the target sequence. The accuracy and validity of the generated 3D structure is largely dependent upon the sequence similarity of the template sequences in relation to the target sequence [44,45]. Sequential identities >25% suggest that the template and target have similar 3D structures and, therefore, the template is appropriate for modeling[46,10,34]. However, sequence similarity greater than 25% is the minimum requirement for generating useful and accurate models[47,48].

Once suitable templates have been identified, a comprehensive literature search must be undertaken on the chosen template sequences, analysing all the entities of the sequence inclusive of its biological role and relation to the target sequence[49,50]. Apart from choosing a template sequence with a high sequence similarity, there are various factors that must also be taken into consideration when selecting an appropriate reference template. These factors include selecting templates that are derived from similar phylogenetic tree to the target sequence[51,52] which may aid in the correct selection of template sequences which are in close relation to the target sequence resulting in the most accurate 3D model being generated [53] .

Other factors include assessment of environmental[34] parameters such as pH[54,55] solvent [44], and identified ligands [10,18] of the template sequences in comparison to the target sequence. This must be accomplished when selecting the most appropriate templates as it ensures the most optimal conditions are adhered for the generation of an accurate target model [48]. The E-value (Expected value) [56] is also referred to when selecting the most appropriate templates to be used, as the closer the E-value is to zero the greater similarity amongst the templates ensuring overall accuracy. Template selection may also be based on multiple alignment as the use of multiple template sequences may increase the accuracy of the 3D protein model[34]. The refinement of the experimental structure is another essential factor when selecting a template, such as the resolution and the R-factor of a crystallographic structure[57], inclusive of the

number of restraints per residue for an NMR structure[58] as it is indicative of the template sequence reliability and precision.

4.3.2. Template Alignment – Once the most suitable reference templates have been identified, alignment is performed which integrates a specialized method to align the template sequences to the target sequence[55]. Local alignment is referred to as the alignment of the query sequence to the substring of the target sequence[59], whereas global alignment involves the use of substrings of both sequences during alignment[60]. Local alignment is performed to aid in the detection of possible templates [61], whereas global alignment is used for model construction[62].

It has been observed that for closely related protein sequences with a greater percentage similarity than 25% there is pronounced accuracy observed during alignment and thus renders it appropriate for modeling[63,64,65,66]. There is an increased observation of misplaced gaps, representing insertions or deletions which cause residues to be misplaced in the template ultimately leading to alignment errors, commonly occurring in models that are generated from sequences with a low sequence similarity. However alignment errors are the main cause of deviations in comparative modeling even when the correct template is chosen[67]. Elucidating the essential role of sequence alignment as a crucial step in homology modeling and thus the quality of sequences to be aligned is of utmost importance and is an indicative factor of accuracy of the generated 3D model[68]. Multiple alignments can be used when more than one template is available. This approach proves advantageous as it increases the spectrum from which the target 3D protein can be modelled[69,31] providing more options to model bad-aligned regions and affords a model reflecting the mean values among all templates[31]. Careful inspection and adjustment on automatic alignment may improve the quality of the modeling[51,70]. There are a wide range of programs that are utilized for sequence alignment such as PSI-BLAST[40] which aligns the target sequence to a sequence profile constructed from multiple alignment of members derived from a protein family. Further advancement of this class of methods is to align two sequence profiles[37] such as the FFAS[71,72] and SALIGN[73,45]. Alignment accuracy improves as one progresses from one generation of the profile methods to another [74,55].

5.3.3. Model building - Proceeding the target-template alignment, the next step in homology modeling is model building. Table 4.1 displays useful tools and services that can be used for model building, some programs presented are also able to generate loop and side chain regions of the 3D model. There is a versatile range of methods that can be utilized to build a protein model one these methods is the rigid-body assemble method[75]. This method dissects various conserved regions of the protein such as the loop regions which anchors the proteins and the side chains which decorate the backbone of the protein model. This method is based on natural dissection enabling the construction of a protein model, by assembling the framework of a small number of rigid bodies, extracted from the aligned template protein structures[76,1], [36,77].

Table 4.1. Popular Tools or Services used for Homology Modeling

Homology modeling Tools or Services	**Description**	Links
CPHmodels	Is a web-server that is able to predict a 3D protein structure from a single template homology model This server employs a novel remote homology-modeling algorithm incorporated in the hybrid scoring functions of the CPHmodels-2.0 [78,79]	http://www.cbs.dtu.dk/services/CPHmodels/
MODELLER	Is a comparative modeling tool that constructs 3D protein structures based on spatial restraints. MODELLER is able to generate a 3D protein model from the provision of an aligned sequence with known related structures[44,80]	http://www.salilab.org/modeller/
SWISS-MODEL	Is a fully automated protein structure homology-modeling server. User-friendly web interfaces. The SWISS-MODEL template library provides annotation of quaternary structure and essential ligands in combination with co-factors to assist in the construction of precise structural models, inclusive of oligomeric structure. This server makes considerable use of model quality estimation for selection of the most suitable templates and provides approximation of the expected accuracy of the resulting models. SWISS-MODEL only requires an amino acid sequence input[81, 82, 83, 84]	http://swissmodel.expasy.org/
Phyre2	This remote homology modeling utilizes a range of detection tools to build 3D models. Special features include prediction of ligand binding as well as analysis of variants among the amino acid sequence of a protein sequence [85].	http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index
Pcons.net	PconsM, an automated protocol that uses single templates to build protein models, whereas PconsM utilizes multiple templates to generate a protein model. PconsM is implemented as a separate extension to this pipeline that is run when the internal and external predictions by Pcons.net are completed, and updated as soon as there are new alignments available[86].	http://pcons.net/
HHpred	HHpred generates 3D models based on pair wise comparison of profile hidden Markov models (HMMs) from a single or multiple query sequence. HHpred displays high sensitivity when generating homology models and searches a range of alignment databases such as Pfam or SMART(different from other software's that use Uniprot) [87].	http://toolkit.tuebingen.mpg.de/hhpred
LOMETS	Local Meta-Threading-Server) is an on-line web service that incorporates spatial restraints to construct 3D protein structures. This server generates the model via collection of high-scoring target-to-template alignments from 8 locally-installed threading programs (FUGUE[88] HHsearch, MUSTER, PPA, PROSPECT2, SAM-T02, SPARKS, SP3).	http://zhanglab.ccmb.med.umich.edu/LOMETS/
Robetta	Develops both *ab initio* and comparative models of protein domains. By using the ROSETTA fragment insertion method[89].	http://www.robetta.org/

	This server is able to construct full-length with the use of a domain prediction method [90,90]	
Chunk-TASSER	A protein structure prediction method that integrates threading templates from SP3 and ab into folded chunk structures, and displays great accuracy when modeling difficult targets[91,92]	http://cssb.biology.gatech.edu/skolnick/webservice/chunk-TASSER/index.html
PSiFR	Provides a vast range of tools for protein tertiary structure prediction and structure and sequence-based function annotation. Incorporates a range of protein structure prediction methods such as TASSER, TASSER-Lite and METATASSER	http://psifr.cssb.biology.gatech.edu/
Protein Model Portal (PMP)	The PMP displays a unique interface highlighting the structural features of a protein. PMP offers interactive services for model building and quality assessment[81,93]	http://www.proteinmodelportal.org/
ProModel	ProModel generates 3D protein structures either from a reference template or a defined template, allowing comprehensive analysis of the target protein such as the active site and channels. ProModel possesses a built-in roamer library able to generate side-chain conformation[38]	http://www.vlifesciences.com/products/VLifeMDS/Protein_Modeller.php
SCWRL4	SCWRL is ranked amongst the best homology modelling tools due to its speed, accuracy, and ease-of-use. It uses a backbone-dependent rotamer library. Due to its speed and accuracy it is an attractive tool specializing in generation of side chain conformations [94]	http://dunbrack.fccc.edu/scwrl4/index.php
IntFOLD	The IntFOLD server is a novel independent server and provides a unified interface for Tertiary structure prediction/3D modeling, 3D model quality assessment, Intrinsic disorder prediction, Domain prediction, Prediction of protein-ligand binding residue [95].	http://www.reading.ac.uk/bioinf/IntFOLD/
PSIPRED	The PSIPRED is a simple and versatile model construction tool, incorporating two feed-forward neural networks which perform an analysis on output obtained from PSI-BLAST [96]. This server performs scalable biological analyses[97].	http://bioinf.cs.ucl.ac.uk/psipred
PEPstrMOD	Performs prediction for structure of peptides containing natural and non-natural/ modified residues. The Pepstr server predicts the tertiary structure of small peptides with sequence length varying between 7 to 25 residues[98].	http://www.imtech.res.in/raghava/pepstr/
PROTEUS2	PROTEUS2 bundles signal peptide identification, transmembrane helix prediction, transmembrane b-strand prediction, secondary structure prediction (for soluble proteins) thus highlighting its role as an avid homology modeling tool. It incorporates all this into a single prediction pipeline[99,100].	http://wks16338.biology.ualberta.ca/proteus2/
Jpred	The Jpred server can utilize either a single protein sequence or multiple sequence alignment and generates a predicted model based on the Jnet algorithm[101]	http://www.compbio.dundee.ac.uk/jpre)

3D Robot	3D Robot is a new algorithm that generates decoys of the protein structure integrating free fragment assembly accompanied by supplemented with hydrogen-bonding inclusive of compact interactions[102].	http://xhanglab.ccmb.med.umich.edu/3DRobot

Other methods include segment matching where comparative models can be constructed by utilising a subset of atomic positions such as the Cα atoms from template structures as "guiding" positions in a bid to identify and assemble short, all-atom segments that correlate to these guiding positions [103]. The search for all-atom segments that may correlate with the conformational search restrained by an energy function [18] it can also be obtained with the use of segment matching program SEGMOD[34,86] which incorporates a generalized method for modelling of 3D protein structures.

Comparative modeling by satisfaction of spatial restraints is one of the most popular methods used for model building, which is performed by computer program MODELLER[104,80,49], this method is initiated by generating a range of constraints derived from the structure of the target sequence[105].

The restraints are generated based on the corresponding distances between aligned residues in the template and the target structures assuming that their structures share structural similarity[106,44]. These derived restraints are usually determined by stereo-chemical restraints on bond lengths, bond angles, dihedral angles, and non-bonded atom contacts that are obtained from molecular mechanics force field[107]. Minimization is performed on the infractions of all the restraints in order to optimise the 3D protein model. The last method utilized is the modeling of the protein structure using artificial evolution. Once the backbone of the 3D protein model is generated it is followed by loop and side-chain modeling.

4.3.4. Modeling loops – Homologous proteins have gaps or insertions in sequences, which are often referred to as loops. However the structure of the loops are not conserved throughout evolution and are regarded as the most variable regions of a protein where insertion and deletion often transpire[108]. The loop structure predominantly determines the functional specificity of a protein structure and plays an essential role in the active and binding site of a protein[109]. The accuracy of loop modeling is a crucial factor in determining the validity and adequacy of homology models for analysing protein-ligand complex interactions[110]. The generation of loops is necessary to connect the sections within the protein and are generally much shorter in length in comparison of the whole protein chain[77]. Thus prediction of the loop structure is a complex process as the loop structure exhibits greater structural variability than strands and helices[111]. There are two widely used methods that are implemented for the construction of the loops. These methods include database-search approach or conformational search methods[112]. The data-base search approach scans all known protein structures to find segments accommodating the pivotal core regions[44]. These methods provide accuracy and efficiency, but may also be associated with limitations such as the number of possible conformations vary as the length of the loop increases[113]. As a result only 4-7 residues long have their conformations present in the protein structure databases.

The second method is the *ab initio*[114] approach also known as the conformational search approach which relies on the optimization of a scoring function [55]. The search strategies include the minimum perturbation method[115], molecular dynamics simulations[116], genetic algorithms[5], Monte Carlo [117] and simulated annealing [60].Loop prediction by optimization is applicable to both simultaneous modeling of multiple loops in combination with the generation of loops interacting with ligands. Although this is a complex process, it is much easily constructed using the ab-initio method as compared to the data-base search approach[34, 118], as fragments are extracted from unrelated structures with different environments.

4.3.5. Side-chain modeling – Side-chain modeling is an important step of homology modelling. Side-chain prediction usually involve the addition of the chains onto fixed backbone coordinates, either obtained from template structures or generated from *ab initio* modeling[119] simulations or a combination of these two methods[120]. Protein side chains tend to exist in a limited number of low energy conformations called rotamers[121,102]. In side-chain prediction methods, rotamers are selected based on the preferred protein sequence and the given backbone coordinates, by using a defined energy function and search strategy. The side-chain quality can be analysed by root mean square deviation (RMSD)[122] for all atoms or by detecting the fraction of correct rotamers found[123]. Table 4.2. Presents tools that can be used for loop and side chain modeling.

Table 4.2. Tools used for Loop and Side Chain Modeling

Loop and Side Chain Modeling Tool	**Link**
Loop Prediction	
COILS	http://www.ch.embnet.org/software/COILS_form.html
CONGEN	http://www.congenomics.com/congen/doc/index.html
Side-Chain Prediction	
RAMP	http://www.ram.org/computing /ramp/ramp.html
SWCRL	http://dunbrack.fccc.edu/SCWRL3.php

4.3.6. Model optimization – After constructing a model the next step is the optimization in order to eliminate or minimize unfavorable interactions between non-covalently bonded atoms[124]. This is usually performed by energy minimizations protocols such as molecular dynamics simulations[125,63] using force fields, implementing avid restrictions in order to avoid excessive deviations from the original templates and, therefore, loss of the experimental configuration of the model [30].

4.3.7. Model validation – the quality of the predicted model can determine the functionality of the model and the amount of data that can be extracted. Thus the accuracy of the model is of utmost importance. The model can be evaluated as a whole or validated based on various segments of the structure[117]. Determining whether the model has the correct fold is one of the key essential steps in validation. The H-factor[126] resembles the role that the R-Factor plays in X-ray crystallography. It stems from the basics of homology modeling and it incorporates all data that were included in the model building protocol to evaluate its accuracy in addition to checking for good stereochemistry. A correct fold will be established if the most optimal template is utilized for the generation of the model in accordance with precise template alignment. A high sequence similarity increases the assumption that the correct fold has been constructed[85]. This can also be attributed to a significant Z-score as well as retention of essential functional or structural residues in the target sequence. Once the correct fold has been accepted, a more detailed analysis can be conducted on the 3D protein model. Apart from sequence similarity playing a crucial role in accuracy as mentioned, the environment influences the accuracy significantly as the structure of a protein determines its functionality within a designated environment.

The stereo-chemical analysis of the protein structure is a basic requirement. Some popular programmes specialising in determining the stereochemistry of the generated models include WHATCHECK[127,55] and PROCHECK[128,129]. Stereo-chemical analysis of the model's consistency is evaluated in a similar manner in which experimental structures are performed. This is according to parameters like the distribution of the torsional angles ϕ and ψ from the main chain[130,131] and the distribution of the rotational angles of side chains[132], in order to fix eventual experimental and interpretation errors[133, 134]. Residues with stereo-chemical problems will fall in non-permitted regions of the Ramachandran plot[135, 56]. The ideal model should present more than 90% of the residues within the permitted regions of the Ramachandran plot [130].

Other programs focus on determining spatial features of the 3D protein model, based on 3D profiles and statistical potential of mean force. These programs include VERIFY3D [136,137] and PROSAII[138,139]. These programs pay substantial attention to environmental parameters in which the model was constructed in respect to the expected environment. Table 4.3 comprises of validation tools that can be utilized when evaluating the validity of the generated 3D model. Model that's have been generated utilising homology modeling can be further assessed in a biannual large-scale experiment known as the Critical Assessment of Techniques for Protein Structure Prediction, or CASP[140].

Table 4.3. Tools used for Homology Model Validation

Validation Tool	Description	Link
PROCHECK	Stereochemical property analysis which includes bond length and angle, torsional angles and chirality of residues[141]	http://www.biochem.ucl.ac.uk/~roman/procheck/procheck.html
WHATCHECK	Analysis of the models Stereochemistry, analysis of nomenclature, symmetry, identification of missing atoms or bonds[142]	http://www.swift.cmbi.ru.nl/gv/whatcheck/
VERIFY3D	Analysis of model compatibility within a designated environment prioritizing the analysis of residue interaction[143]	http://www.doe-mbi.ucla.edu/Services/Verify_3D or http://nihserver.mbi.ucla.edu/Verify_3D
PROSAII	Focuses on regions in the model that may have been incorrectly folded, as well as incorrect regions of the structural model[144, 145]	http://www.cam.sbg.ac.at
Molprobity	MolProbity is a web based 3D structure validation service that evaluates the quality of a structure at both the global and local levels. Includes steric and geometric analysis[146].	http://molprobity.biochem.duke.edu/index.php?MolProbSID=7h23k0a4ji0t4ecvmpbtbeoje5&eventID=2
VADAR	Volume, Area, Dihedral Angle and Reporter Utilises 15 different algorithms and programs for protein structure analysis including quantitative and qualitative assessment	http://vadar.wishartlab.com/VADAR
Bioinformatics Toolkit	Integrates a broad spectrum of interactive protein bioinformatics analysis[147]	http://toolkit.tuebingen.mpg.de

4.4. Applications of Homology Modeling

Homology modeling plays a focal role in the process of structure-based drug design and its importance is vastly becoming more evident as the number of available crystal structures increases.

The function of a protein is dependent upon its motion and conformational changes that may occur[148]. In order to fully understand the molecular mechanism of a protein, a concise description of versatile functional states that the protein structure can adopt dynamically must be established[8]. Some examples include allosteric conformational changes on binding events, identification of drug-binding cavities [149, 86], intermediate excited states on reaction cycles, transport and motion phenomena[150, 151]. Thus elucidating the pivotal role that homology modeling plays in exploring these alternative conformations, generating a better understanding behind protein dynamic transitional changes [152, 153, 154].

Homology modeling is a formidable application that has not only benefited many spectrums of drug design but has also enhanced our understanding on various other aspects that impact drug discovery such as analysis of mutations within the active site[155, 156] and assessing their biological role in the disease state[70]. Homology modeling has also been found to play a crucial role in compound optimization [102]

Typical applications of a homology model in drug discovery require a very high accuracy of the local side chain positions in the binding site. There are a substantial number of 3D protein structures arising from homology modeling over the years. Targets have included antibodies [157], multiple proteins involved in human biology and medicine[158]. Homology modeling also plays a crucial role in structure-based virtual screening for drug discovery[7]. Figure 4.2 displays the many applications of homology modeling in drug discovery.

Figure 4.2. Range of applications stemming from Homology Modeling accompanied by various Protein targets-orange of each application

Employing the insight extracted directly from the 3D structural details of a drug target may constitute an accurate account of the drug-protein interaction, providing an acute perspective on the mode of action (MOA) of the drug molecule which may further advance the process of drug design and discovery. Utilization of accurate 3D protein structures may ensure drug-target specificity and effectivity, which is an eminent requirement in all fields of drug design. Although homology modeling plays a pivotal role in drug discovery, there are minor hiccups associated with its application that must be overcome. These challenges along with possible solutions are summarized in Table 4.4

Table 4.4. Challenges and possible solution associated with Homology Modeling

Challenges associated with Homology Modeling	**Possible Solutions towards generating an accurate homology model**
Reduced accuracy	Improving the accuracy of the model, might encompass optimization techniques in side chain modeling as well as loop modeling.
Modeling of 3D Protein structures with low sequence similarity to other templates	Use of multiple templates, which may ultimately lead to the larger coverage of the target sequence
The use of multiple templates can led to deviations in the alignment thus leading to problems associated with convergence	Use of multiple templates that are derived from similar phylogenetic tree as that of the target sequence eliminates this challenge also the use of multiple templates alleviates the need for free modeling and thus increases the quality of the target protein
Homology models are sometimes considered incorrect	This may not be applicable if the homology model can be compared to a structure that spans the entirety of the target sequence
Alignment errors are the main cause of deviations in comparative modeling	Careful inspection and adjustment on automatic alignment may improve the quality of the modeling[51]

4.5. Conclusion

A rising number of publications have established homology modeling as a fast, trustable, and very

useful tool to access consistent 3D models for proteins in which experimental structures are not available. The main aim of homology modeling is to predict a structure from its sequence with an accuracy that is similar to the results obtained experimentally. Homology modeling provides a feasible cost-effective alternative method to generate models. Homology modeling studies are fastened through the use of visualization technique, and the differential properties of the proteins can be discovered. Advances in structural biology obtained using homology models substantiates the robustness and reliability of the software available today to build models. The recent advances in a range of software with improvements in the algorithms for alignment, modeling loops and side chains, detection of errors and validation of models, have made possible the prediction of proteins structures that, until recently, were a remote possibility. Today it is quite clear that with the appropriate software and templates, very consistent models can be built utilising this innovative tool. Models generated utilising homology modeling have effectively contributed to the field of drug design and development, and from the looks of it, homology modeling will continue to be a strong defender against the war on infectious diseases where X-ray crystal structures of drug targets are unavailable.

References

(1) Blundell, T. ., Sibanda, B. ., Sternberg, M. J. ., and Thornton, J. . (1987) Knowledge-based prediction of protein structures and the design of novel molecules. *Nat. Vol. 326*, 347–352.

(2) De Azevedo, W. F., Caceres, R. A., Pauli, I., M, L. F. S., Barcellos, G. B., Rocha, K. B., Botelho, M., and Soares, P. (2009) Protein-Drug Interaction Studies for Development of Drugs Against Plasmodium falciparum. *Curr. Drug Targets* 271–278.

(3) König, J., Müller, F., and Fromm, M. F. (2013) Transporters and Drug-Drug Interactions : Important Determinants of Drug Disposition and Effects. *Pharmacol. Rev. 65*, 944–966.

(4) Lionta, E., Spyrou, G., Vassilatis, D. K., and Cournia, Z. (2014) Structure-Based Virtual Screening for Drug Discovery : Principles , Applications and Recent Advances. *Curr. Top. Med. Chem. 14*, 1923–1938.

(5) Sliwoski, G., Kothiwale, S., Meiler, J., and Lowe, E. W. (2014) Computational Methods in Drug Discovery. *Pharmacol. Rev.* 334–395.

(6) Cavasotto, C. N., and Phatak, S. S. (2009) Homology modeling in drug discovery : current trends and applications. *Drug Discov. Today 14*.

(7) Cheng, T., Li, Q., Zhou, Z., Wang, Y., and Bryant, S. H. (2012) Structure-Based Virtual Screening for Drug Discovery : a Problem-Centric Review. *Am. Assoc. Pharm. Sci. 14*.

(8) Schmidt, T., Bergner, A., and Schwede, T. (2015) Modelling three-dimensional protein structures for applications in drug design. *Drug Discov. Today 19*, 890–897.

(9) Lambert, C., Nadia, L., and Depiereux, E. (2002) ESyPred3D : Prediction of proteins 3D structures. *Bioinformatics 18*, 1250–1256.

(10) Celmar, T., and França, C. (2015) Homology modeling : an important tool for the drug discovery. *J. Biomol. Struct. Dyn. 8*, 1780–1793.

(11) Kreiger, E., Nabuurs, S. B., and Vriend, G. (1963) Homology modeling. *Struct. Bioinforma.* (Bourne, P. E., and Weissig, H., Eds.). Wiley-Liss, Inc, New York.

(12) Centeno, N. B., Planas-iglesias, J., and Oliva, B. (2005) Comparative modelling of protein structure and its impact on microbial cell factories. *BioMed Cent. 4*, 1–11.

(13) Schwede, T., and Kopp, J. (2006) The SWISS-MODEL Repository : new features and

functionalities. *Nucleic Acids Res. 34*, 315–318.

(14) Terwilliger, T. C. (2011) The success of structural genomics. *J. Struct. Funct. Genomics 12*, 43–44.

(15) Söding, J. (2005) Sequence analysis Protein homology detection by HMM – HMM comparison. *Bioinformatics 21*, 951–960.

(16) Remmert, M., Biegert, A., and Hauser, A. (2012) HHblits : lightning-fast iterative protein sequence searching by HMM-HMM alignment. *Nat. Methods 9*.

(17) Zhou, Y., and Johnson, M. E. (1999) Comparative molecular modeling analysis of 5-amidinoindole and benzamidine binding to thrombin and trypsin : specific H-bond formation contributes to high 5-amidinoindole potency and selectivity for thrombin and factor Xa. *J. Mol. Recognit. 12*, 235–241.

(18) Vyas, V. K., Ukawala, R. D., and Chintha, C. (2016) Homology Modeling a Fast Tool for Drug Discovery : Current Perspectives. *PubMed Cent.* 1–20.

(19) Ferreira, L. G., Santos, R. N., Oliva, G., and Andricopulo, A. D. (2015) Molecular Docking and Structure-Based Drug Design Strategies. *Molecules 20*, 13384–13421.

(20) Nguyen, T. L., Gussio, R., Smith, J. A., Lannigan, D. A., Hecht, S. M., Scudiero, D. A., Shoemaker, H., and Zaharevitz, D. W. (2006) Homology model of RSK2 N-terminal kinase domain , structure-based identification of novel RSK2 inhibitors , and preliminary common pharmacophore. *Bioorg. Med. Chem. 14*, 6097–6105.

(21) Mundaganur, D. S., Mundaganur, D. Y., and Kannarath, A. (2014) Bioinformatics Study on Zaire Ebolavirus (EBOV) Protein for Better Understanding the Vaccine Development. *Int. J. Appl. Sci. Biotechnol. 2*, 544–552.

(22) Balmith, M., Faya, M., and Soliman, M. E. S. (2016) Ebola virus: A gap in drug design and discovery - experimental and computational perspective. *Chem. Biol. Drug Des.* 1–12.

(23) Ramharack, P., and Soliman, M. E. S. (2016) Zika virus drug targets: a missing link in drug design and discovery – a route map to fill the gap. *RSC Adv. 6*, 68719–68731.

(24) Hasegawa, K. (2012) Introduction to single crystal X-ray analysis. *Rigaku J. 28*, 14–18.

(25) Tjioe, E., Schneidman-, D., Peterson, B., and Sali, A. (2012) Putting the Pieces Together : Integrative Modeling Platform Software for Structure Determination of Macromolecular Assemblies. *PLOS Biol. 10*, 1–5.

(26) Zheng, H., Handing, K. B., Zimmerman, M. D., Almo, S. C., Minor, W., Physics, B., and Avenue, J. P. (2016) X-ray crystallography over the past decade for novel drug discovery – where are we heading next? *Expert Opin. Drug Discov. 10*, 975–989.

(27) Schneider, M., Fu, X., and Keating, A. E. (2010) X-ray vs. NMR structures as templates for computational protein design. *Journa of Proteins 77*, 97–110.

(28) Honarparvar, B., Govender, T., Maguire, G. E. M., Soliman, M. E. S., and Kruger, H. G. (2014) Integrated Approach to Structure-Based Enzymatic Drug Design : Molecular Modeling , Spectroscopy , and Experimental Bioactivity. *Chem. Rev. 114*, 493–537.

(29) Khor, B. Y., Tye, G. J., Lim, T. S., and Choong, Y. S. (2015) General overview on structure prediction of twilight-zone proteins. *Theor. Biol. Med. Model. 12*, 1–11.

(30) Andrade, O., Filho, S., and Alencastro, B. De. (2003) Homology modeling of proteins. *Quim.Nova 26*, 253–259.

(31) Fernandez-fuentes, N., Rai, B. K., Madrid-aliste, C. J., and Fajardo, J. E. (2007) Structural bioinformatics Comparative protein structure modeling by combining multiple templates and optimizing sequence-to-structure alignments. *Bioinformatics 23*, 2558–2565.

(32) Tsujikawa, H., Sato, K., Wei, C., Saad, G., and Sumikoshi, K. (2016) Development of a protein – ligand-binding site prediction method based on interaction energy and sequence conservation. *J. Struct. Funct. Genomics 17*, 39–49.

(33) Pieper, U., Webb, B. M., Dong, G. Q., Schneidman-duhovny, D., Fan, H., Kim, S. J., Khuri, N., Spill, Y. G., Weinkam, P., Hammel, M., Tainer, J. A., Nilges, M., Sali, A., and Roux, D. (2013) ModBase , a database of annotated comparative protein structure models and associated resources. *Nucleic Acids Res.* 1–11.

(34) Fiser, A. (2010) Template-Based Protein Structure Modeling. *Methods Mol. Biol. 673*, 1–20.

(35) Claessens, M., Cutsem, E. Van, Lasters, I., and Wodak, S. (1989) Modelling the polypeptide backbone with ’ spare parts ’ from known protein structures. *protein Eng. 2*, 335–345.

(36) Greer, J. (1990) Comparative Modeling Methods : Application to the Family of the Mammalian Serine Proteases. *PROTEINS Struct. Funct. Genet. 334*, 317–334.

(37) Pearson, W. R. (2014) An Introduction to Sequence Similarity (“ Homology ”) Searching. *Curr. Protoc. Bioinforma.* 1–9.

(38) Rao, V. S., Srinivas, K., Sujini, G. N., and Kumar, G. N. S. (2014) Protein-Protein Interaction Detection : Methods and Analysis. *Int. J. Proteomics 2014*, 1–12.

(39) Berman, H., Henrick, K., Nakamura, H., and Markley, J. L. (2007) The worldwide Protein Data Bank (wwPDB): ensuring a single , uniform archive of PDB data. *Nucleic Acids Res. 35*, 2006–2008.

(40) Altschup, S. F., Gish, W., Pennsylvania, T., and Park, U. (1990) Basic Local Alignment Search Tool. *J. Mol. Biol. 215*, 403–410.

(41) Benson, D. A., Boguski, M., Lipman, D. J., and Ostell, J. (1994) GenBank. *Oxford Univ. Press 22*, 3441–3444.

(42) Pearson, W. R., and Lipmant, D. J. (1988) Improved tools for biological sequence comparison. *Natl. Biomed. Res. Found. 85*, 2444–2448.

(43) Bastian, F. B., Chibucos, M. C., Gaudet, P., Giglio, M., Holliday, G. L., Huang, H., and Lewis, S. E. (2015) Original article The Confidence Information Ontology : a step towards a standard for asserting confidence in annotations. *Database* 1–11.

(44) Eswar, N., Webb, B., Marti-renom, M. A., Madhusudhan, M. S., Eramian, D., Shen, M., Pieper, U., and Sali, A. (2014) Comparative Protein Structure Modeling Using Modeller. *J. Protoc. Bioinforma. 9*, 1–47.

(45) Madhusudhan, M. S., and Marti-Renom, M. A. (2011) Comparative Protein Structure Modeling, in *The Proteomics Protocol Handbook*, pp 831–837. Human Press Inc, Tokotowa, NJ.

(46) Rost, B., and Sander, C. (1996) Bridging the protein sequence-structure gap by structure predictions. *Annu. Rev. Biophys. Biomol. Struct. 25*, 113–136.

(47) Webb, B., Sali, A., Sciences, T., Chemistry, P., Biosciences, Q., Francisco, C. S., Francisco, S., and Head, R. (2013) Protein structure modeling with.

(48) Forrest, L. R., Tang, C. L., and Honig, B. (2006) On the Accuracy of Homology Modeling and Sequence Alignment Methods Applied to Membrane Proteins. *Biophys. J. 91*, 508–517.

(49) Yang, Z., and Sitao, W. (2009) A comprehensive assessment of sequence-based and template-based methods for protein contact prediction. *J. Bioinforma. 24*, 924–931.

(50) Dunbrack, L. R. J. (2006) Sequence comparison and protein structure prediction. *Curr. Opin. Struct. Biol. 16*, 374–384.

(51) Ashkenazy, H., Unger, R., and Kliger, Y. (2011) Hidden conformations in protein structures. *J. Bioinforma. 27*, 1941–1947.

(52) Felsenstein, J. (1981) Evolutionary Trees from DNA Sequences : A Maximum Likelihood Approach. *J. Mol. Evol. 17*, 368–376.

(53) Peng, J. (2013) Statistical Inference for template-based protein structure prediction Statistical Inference for template-based protein structure prediction. Toyota Technological Institue at Chicago.

(54) Marti-renom, M. A., Yerkovich, B., and Sali, A. (2002) Modeling Protein Structure From Its Sequence. *Lab. Mol. Biophys.* 1–44.

(55) Saxena, A., Sangwan, R. S., and Mishra, S. (2013) Fundamentals of Homology Modeling Steps and Comparison among Important Bioinformatics Tools : An Overview. *Sci. Int. 7*, 237–252.

(56) Ramachandran, S., Dokholyan, N. V, and Need, H. M. (2012) Homology Modeling : Generating Structural Models to Understand Protein Function and Mechanism, in *Computational Modeling of Biological Sciences: From Molecules to Pathways, Biological and Medical Physics, Biomedical Engineering.*, pp 97–116.

(57) Brunger, A. T., and Alto, P. (2012) 1.7 Refinement of X-ray Crystal Structures, in *Comprehensive Biophysics*, pp 105–115.

(58) Buchner, L., and Gu, P. (2015) Increased Reliability of Nuclear Magnetic Resonance Protein Structures by Consensus Structure Bundles Article Increased Reliability of Nuclear Magnetic Resonance Protein Structures by Consensus Structure Bundles. *Elsevier 23*, 1–10.

(59) Tainer, J. A. (2014) Why Align Strings ? John Wiley and Sons, New Jersey.

(60) Joo, K., Lee, J., and Lee, J. (2012) Chapter 7 Methods for Accurate Homology Modeling by Global Optimization, in *Homology Modeling: Methods and Protocols, Methods in Molecular Biology*, pp 175–188.

(61) Hui, S. L., and Wonpil, I. (2013) Identification of Ligand Templates using Local Structure Alignement for Structure-based Drug Design. *J. Chem. Inf. Model. 52*, 2784–2795.

(62) Neuwald, A. F. (2009) Rapid detection , classification and accurate alignment of up to a million or more related protein sequences. *J. Bioinforma. 25*, 1869–1875.

(63) Bryant, S. H., and Lawrence, C. E. (1993) An Empirical Energy Function for Threading Protein Sequence Through the Folding Motif. *J. Proteins Struct. Funct. Genet. 112*, 92–112.

(64) Fineklstein, A. V, and Reva, B. A. (1991) A search fro the most stable folds of protein chains. *Lett. to Nat. 351*, 497–499.

(65) Jones, D. ., Taylot, R. ., and Thornton, J. . (1992) A new approach to protein fold recognition. *Lett. to Nat. 358*, 86–89.

(66) Sippl, M. J., and Weitckus, S. (1992) Detection of Native-Like Models for Amino Acid Sequences of Unknown Three-Dimensional Structure in a Data Base of Known Protein Conformations. *J. Proteins Struct. Funct. Genet. 271*, 258–271.

(67) Chakravarty, S., Wang, L., and Sanchez, R. (2005) Accuracy of structure-derived properties in simple comparative models of protein structures. *Nucleic Acids Res. 33*, 244–259.

(68) Thampi, S. M. Bioinformatics, in *Bioinformatics*, pp 1–155. Wiley-Liss, Inc, Houston.

(69) Fernandez-fuentes, N., Madrid-aliste, C. J., Rai, B. K., and Eduardo, J. (2007) M4T : a comparative protein structure modeling server. *Nucleic Acids Res. 35*, 363–368.

(70) Sun, W., Gerth, C., Maeda, A., Lodowski, D. T., Palczewski, K., Kraak, L. Van Der, Saperstein, D. A., and He, E. (2007) Novel RDH12 mutations associated with Leber congenital amaurosis and cone-rod dystrophy : Biochemical and clinical evaluations. *Elsevier 47*, 2055–2066.

(71) Xu, D., Jaroszewski, L., Li, Z., and Godzik, A. (2013) Structural bioinformatics FFAS-3D : improving fold recognition by including optimized structural features and template re-ranking. *Bioinforma. Adv. Access* 1–8.

(72) Jaroszewski, L., Li, Z., Cai, X., Weber, C., and Godzik, A. (2011) FFAS server : novel features and applications. *Nucleic Acids Res. 39*, 38–44.

(73) Braberg, H., Webb, B. M., Tjioe, E., Pieper, U., Sali, A., and Madhusudhan, M. S. (2012) SALIGN : a web server for alignment of multiple protein sequences and structures. *Struct. Bioinforma. 28*, 2072–2073.

(74) Thompson, J. D., Plewniak, F., and Poch, O. (1999) A comprehensive comparison of multiple sequence alignment programs. *Nucleic Acids Res. 27*, 2682–2690.

(75) Katebi, A. R., Kloczkowski, A., and Jernigan, R. L. (2010) Structural interpretation of protein-protein interaction network. *BMC Struct. Biol. 10*, 1–11.

(76) Browne, J. W., North, A. C. ., and Phillips, D. . (1969) A Possible Three-dimensional Structure of Bovine a-Lactalbumin based on that of Hen's Egg-White Lysozyme. *J. Mol. Biol. 42*, 65–86.

(77) Collura, V., Garnier, J., and Higo, J. (1993) Modeling of protein loops by simulated annealing. *J. Protein Sci. 2*, 1502–1510.

(78) Nielsen, M., Lundegaard, C., Lund, O., and Petersen, T. N. (2010) structure-guided sequence profiles. *Nucleic Acids Res. 38*, 576–581.

(79) Gomes, H., Romeiro, N. C., Braz, G. R. C., Alves, E., Oliveira, G. De, Rodrigues, C., Nunes, R., Githaka, N., Isezaki, M., Konnai, S., Ohashi, K., Vaz, S., Logullo, C., and Moraes, J. (2013) Identification and Structural-Functional Analysis of Cyclin-Dependent Kinases of the Cattle Tick Rhipicephalus (Boophilus) microplus. *PLoS One 8*, 1–12.

(80) Webb, B., and Sali, A. (2014) Comparative Protein Structure Modeling Using MODELLER, in *Modeling Structure from Sequence*, pp 1–32. John Wiley and Sons, New York.

(81) Arnold, K., Bordoli, L., and Schwede, T. (2006) Structural bioinformatics The SWISS-MODEL workspace : a web-based environment for protein structure homology modelling. *Bioinformatics 22*, 195–201.

(82) Kopp, J., Battey, J. N. D., Podvinec, M., Westbrook, J. D., Berman, H. M., Bordoli, L., and Schwede, T. (2009) The Protein Model Portal. *J. Struct. Funct. Genomics 10*, 1–8.

(83) Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., Kiefer, F., Cassarino, T. G., Bertoni, M., Bordoli, L., and Schwede, T. (2014) SWISS-MODEL : modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res. 42*, 252–258.

(84) Bordoli, L., Kiefer, F., Arnold, K., Benkert, P., Battey, J., and Schwede, T. (2009) Protein structure homology modeling using SWISS-MODEL workspace. *Nat. Protoc. 4*.

(85) Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., and Sternberg, M. J. E. (2015) The Phyre2 web portal for protein modeling , prediction and analysis. *Nat. Protoc. 10*, 845–858.

(86) Larsson, P., Skwark, M. J., Wallner, B., and Elofsson, A. (2011) Improved predictions by Pcons . net using multiple templates. *Bioinformatics 27*, 426–427.

(87) Hildebrand, A., Remmert, M., Biegert, A., and So, J. (2009) PREDICTION REPORT Fast and accurate automatic structure prediction with HHpred ¨. *J. Proteins Struct. Funct. Bioinforma. 9*, 128–132.

(88) Shi, J., Blundell, T. L., and Mizuguchi, K. (2001) FUGUE : Sequence-structure Homology Recognition Using Environment-specific Substitution Tables and Structure-dependent Gap Penalties. *J. Mol. Biol. 310*, 243–257.

(89) Das, R., and Baker, D. (2008) Macromolecular Modeling with Rosetta. *J. Annu. Rev. 77*, 363–382.

(90) Kim, D. E., Chivian, D., and Baker, D. (2004) Protein structure prediction and analysis using the Robetta server. *Nucleic Acids Res. 32*, 526–531.

(91) Yang, J., Wang, Y., and Zhang, Y. (2015) ResQ : An Approach to Unified Estimation of B -Factor and Residue-Specific Error in Protein Structure Prediction. *J. Mol. Biol.* 1–9.

(92) Zhou, H., and Skolnick, J. (2007) Ab Initio Protein Structure Prediction Using Chunk-TASSER. *Biophys. J. 93*, 1510–1518.

(93) Haas, J., Roth, S., Arnold, K., Kiefer, F., Schmidt, T., Bordoli, L., and Schwede, T. (2013) Original article The Protein Model Portal — a comprehensive resource for protein structure and model information. *Oxford Univ. Press 2013*, 1–8.

(94) Krivov, G. G., Shapovalov, M. ., and Dunbrack, R. L. jr. (2010) Improved prediction of protein side-chain conformations with SCWRL4. *Journa of Proteins 77*, 778–795.

(95) Roche, D. B., Buenavista, M. T., Tetchner, S. J., Mcguffin, L. J., Beamline, B., and Ode, D. O. X. (2011) The IntFOLD server : an integrated web resource for protein fold recognition , 3D model quality assessment , intrinsic disorder prediction , domain prediction and ligand binding site prediction. *Nucleic Acids Res. 39*, 171–176.

(96) Gibbons, T. R., Mount, S. M., Cooper, E. D., and Delwiche, C. F. (2015) Evaluation of BLAST-based edge-weighting metrics used for homology inference with the Markov Clustering algorithm. *BMC Bioinformatics 16*, 1–14.

(97) Buchan, D. W. A., Minneci, F., Nugent, T. C. O., Bryson, K., and Jones, D. T. (2013) Scalable web services for the PSIPRED Protein Analysis Workbench. *Nucleic Acids Res. 41*, 349–357.

(98) Singh, S., Singh, H., Tuknait, A., Chaudhary, K., Singh, B., Kumaran, S., and Raghava, G. P. S. (2015) PEPstrMOD : structure prediction of peptides containing natural , non-natural and modified residues. *Biol. Direct* 1–19.

(99) Montgomerie, S., Cruz, J. A., Shrivastava, S., Arndt, D., Berjanskii, M., and Wishart, D. S. (2008) PROTEUS2 : a web server for comprehensive protein structure prediction and structure-based annotation. *Nucleic Acids Res. 36*, 202–209.

(100) Montgomerie, S., Sundararaj, S., Gallin, W. J., and Wishart, D. S. (2006) using structural alignment. *J. Biomed. Cent. 13*, 1–13.

(101) Cole, C., Barber, J. D., and Barton, G. J. (2008) The Jpred 3 secondary structure prediction server. *Nucleic Acids Res. 36*, 197–201.

(102) Deng, H., Jia, Y., and Zhang, Y. (2015) Structural bioinformatics 3DRobot : automated generation of diverse and well-packed protein structure decoys. *Bioinformatics* 1–10.

(103) Chen, T. S., and Keating, A. (2012) Design of Protein-Protein Interaction Specificity Using by. *Des. Protein-Protein Interact. Specif. Using Comput. Methods Exp. Libr. Screen.*

(104) Fiser, A., and Sali, A. (2003) MODELLER: Generation and Refinement of Homology-Based Protein Structure Models. *Comp. protein Struct. Model. 374*, 461–491.

(105) Meier, A., and Söding, J. (2015) Automatic Prediction of Protein 3D Structures by Probabilistic Multi-template Homology Modeling. *PLOS Comput. Biol. 11*, 1–20.

(106) Gront, D., Grabowski, M., Raynor, J., Tkaczuk, K. L., and Minor, W. (2014) NIH Public Access. *J. Funct. Genomics 13*, 213–225.

(107) Feyfant, E., and Sali, A. (2007) Modeling mutations in protein structures. *J. Protein Sci. 16*, 2030–2041.

(108) Blouin, C., Butt, D., and Roger, A. J. (2004) Rapid evolution in conformational space : A study of loop regions in a ubiquitous GTP binding domain. *J. Protein Sci. 13*, 608–616.

(109) Steinberg, S. F. (2010) Structural Basis of Protein Kinase C Isoform Function. *J. Physiol. Rev. 88*, 1341–1378.

(110) Lee, J., Lee, D., Park, H., Coutsias, E. A., and Seok, C. (2011) Protein loop modeling by using fragment assembly and analytical loop closure. *Journa of Proteins 78*, 3428–3436.

(111) Regad, L., Martin, J., Nuel, G., and Camproux, A. (2010) Mining protein loops using a structural alphabet and statistical exceptionality. *BMC Bioinformatics 11*, 1–18.

(112) Xuan-Yu, M., Hong-Xing, Z., Mihaly, M., and Meng, C. (2012) Molecular Docking: A powerful approach for structure-based drug discovery. *J. Curr. Comput. Aided Drug Des. 7*, 146–157.

(113) Kmiecik, S., Gront, D., Kolinski, M., Wieteska, L., Dawid, A. E., Kolinski, A., Models, S., and Field, F. (2016) Coarse-Grained Protein Models and Their Applications. *Chem. Rev. 106*.

(114) Wu, S., Skolnick, J., and Zhang, Y. (2007) Ab initio modeling of small proteins by iterative TASSER simulations. *BMC Biol. 5*, 17.

(115) Taylor, P., Ohsaki, M., and Nakamura, T. (2007) MIinimum constraint pertubation method for system topology optimization, in *Engineering Optimization*, pp 37–41. Gordon and Breach Science Publishers SA, Amsterdam.

(116) Shen, J., Zhang, W., Fang, H., Perkins, R., Tong, W., and Hong, H. (2013) Homology modeling , molecular docking , and molecular dynamics simulations elucidated a -fetoprotein binding modes. *BM 14*.

(117) Kauffman, C., Rangwala, H., and Karypis, G. (2007) Improving Homology Models for Protein-Ligand Binding Sites, pp 1–12.

(118) Trotov, M. (2012) Loop Modeling, in *Homology Modeling: Methods and Protocols, Methods in Molecular Biology*.

(119) Wang, Q., Canutescu, A. A., and Jr, R. L. D. (2008) PROTOCOL SCWRL and MolIDE : computer programs for side-chain conformation prediction and homology modeling. *Nat. Protoc. 3*, 1832–1847.

(120) Levitt, M. (1992) Accurate Modeling of Protein Conformation Segment Matching Michael Levitt by Automatic. *J. Mol. Biol. 226*, 507–533.

(121) Holm, L., and Sander, C. (1991) Database Algorithm for Generating Protein Backbone and Side-chain Co-ordinates from a C-alpha trace: Application to model building and detection of co-ordinate errors. *J. Mol. Biol. 218*, 183–194.

(122) Bordoli, L., Schwede, T., Orry, A. J. W., and Abagyan, R. (2012) Homology Modeling. *Methods Mol. Biol. 857*, 107.

(123) Daley, M. E., and Sykes, B. D. (2003) The role of side chain conformational flexibility in surface recognition by Tenebrio molitor antifreeze protein. *J. Protein Sci. 12*, 1323–1331.

(124) Gregory, S. (2010) Monte Carlo Optimization, in *Introducing Monte Carlo Methods with R*.

(125) Bruccoleri, R. E. Ab Initio Loop Modeling and Its Application to Homology Modeling, in *Protein structure prediciton: methods and protocols*, pp 247–264.

(126) Luccio, E., and Koehl, P. (2011) A quality metric for homology modeling : the H-factor. *BMC Bioinformatics 12*.

(127) Vriend, G. (1990) WHAT IF : A molecular modeling and drug design program. *J. Mol. Graph. 8*, 52–56.

(128) Lemieux, M. J. (2007) Eukaryotic major facilitator superfamily transporter modeling based on the prokaryotic GlpT crystal structure. *J. Mol. Membr. Biol. 24*.

(129) Laskowski, R. A., Rullman, A. ., MacArthur, M. W., Kaptein, R., and Thornton, J. M. (1996) AQUA and PROCHECK-NMR: Programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR 8*, 477–486.

(130) Kuster, D. J., Liu, C., Fang, Z., Ponder, J. W., and Marshall, G. R. (2015) High-Resolution Crystal Structures of Protein Helices Reconciled with Three-Centered Hydrogen Bonds and Multipole Electrostatics. *PLoS One 10*, 1–37.

(131) Depristo, M. A., Bakker, P. I. W. D. E., Shetty, R. P., and Blundell, T. O. M. L. (2003) Discrete restraint-based protein modeling and the C- alpha trace problem. *J. Protein Sci. 12*, 2032–2046.

(132) Arnautova, Y. A., Abagyan, R., and Totrov, M. (2015) All-Atom Internal Coordinate Mechanics (ICM) Force Field for Hexopyranoses and Glycoproteins. *J. Chem. Theory Comput. 11*, 2167–2186.

(133) Chu, Y. M., and Feng, K. K. (1991) Concluding remarks. *Cryst. Chem. 47*, 392–400.

(134) Cowan, S. W. (1991) Improved Methods for Building Protein Models in Electron Density Maps and the Location of Errors in these Models. *Cryztal Chem. 47*, 110–119.

(135) Chen, Y. (2014) Beware of docking ! *Trends Pharmacol. Sci.* 1–18.

(136) Profiles, T. (1997) Verify3D: Assessment of protein models with three-dimensional profiles, in *Methods in enzymology*, pp 396–404.

(137) Elofsson, A. (2006) Identification of correct regions in protein models using structural , alignment , and consensus information. *J. Protein Sci. 15*, 900–913.

(138) Wiederstein, M., and Sippl, M. J. (2007) ProSA-web : interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res. 35*, 407–410.

(139) Sippl, M. J. (1993) Recognition of Errors in Three-Dimensional Structures of Proteins. *PROTEINS Struct. Funct. Genet. 362*, 355–362.

(140) Kryshtafovych, A., Monastyrskyy, B., and Fidelis, K. (2015) CASP predition center infrastructure and evalution measures in CASP10 and CASP roll *82*, 7–13.

(141) Rasouli, H., and Fazeli-nasab, B. (2014) Structural Validation and Homology Modeling of Lea 2 Protein in Bread Wheat Agricultural Biotechnology , Graduated from Kermanshah University , Iran.

Amerixan-Eurasian J. Agric. Evironmental Sci. 14, 1044–1048.

(142) Richardson, J. S., Prisant, M. G., and Richardson, D. C. (2013) Crystallographic model validation : from diagnosis to healing. *Curr. Opin. Struct. Biol. 23*, 707–714.

(143) Sasin, J. M., and Bujnicki, J. M. (2004) COLORADO3D , a web server for the visual analysis of protein structures. *Nucleic Acids Res. 32*, 586–589.

(144) Bagaria, A., Jaravine, V., Huang, Y. J., Montelione, G. T., and Gu, P. (2012) Protein structure validation by generalized linear model root-mean-square deviation prediction. *J. Protein Sci. 21*, 229–238.

(145) Fasnacht, M., Zhu, J., and Honig, B. (2007) Local quality assessment in homology models using statistical potentials and support vector machines. *J. Protein Sci. 16*, 1557–1568.

(146) Chen, V. B., Bryan, W., Iii, A., Headd, J. J., Keedy, D. A., Robert, M., Kapral, G. J., Murray, L. W., Jane, S., and David, C. (2010) research papers MolProbity : all-atom structure validation for macromolecular crystallography research papers. *Acta Crystalloraphica 66*, 12–21.

(147) Alva, V., Nam, S., Johannes, S., Lupas, A. N., Mpi, T., and Toolkit, B. (2016) The MPI bioinformatics Toolkit as an integrative platform for advanced protein sequence and structure. *Nucleic Acids Res. 44*, 410–415.

(148) Rai, B. K., Tawa, G. J., Katz, A. H., and Humblet, C. (2010) Brajesh K. Rai,* Gregory J. Tawa, Alan H. Katz, and Christine Humblet. *PROTEINS Struct. Funct. Genet. 78*, 457–473.

(149) Fritz, H., Rojas, A. M., Schwarzenbacher, R., Proell, M., and Riedl, S. J. (2008) The Nod-Like Receptor (NLR) Family : A Tale of Similarities and Differences. *PLoS One 3*, 1–11.

(150) Ochi, T., Wu, Q., Chirgadze, D. Y., Gu, J., Bolanos-garcia, V. M., and Blundell, T. L. (2012) Article Structural Insights into the Role of Domain Flexibility in Human DNA Ligase IV. *Cell Press Struct. 20*, 1212–1222.

(151) Yoshikawa, Y., Oishi, S., Noriko, T., Fujii, N., and Furuya, T. (2013) An optimized method of G-protein coupled receptor homology modeling : its application to the discovery of novel CXCR7 ligands.

(152) Nussinov, R. (2011) Review Dynamic Allostery : Linkers Are Not Merely Flexible. *Cell Press Struct. 13*, 907–917.

(153) Nygaard, R., Zou, Y., Dror, R. O., Mildorf, T. J., Arlow, D. H., Manglik, A., Pan, A. C., Bokoch,

M. P., Thian, F. S., Kobilka, T. S., Shaw, D. E., Liu, C. W., Jose, J., Mueller, L., Prosser, R. S., and Kobilka, B. K. (2013) The Dynamic Process of b 2 -Adrenergic Receptor Activation. *Cell 152*, 532–542.

(154) Weinkam, P., Pons, J., and Sali, A. (2012) Structure-based model of allostery predicts coupling between distant sites. *PNAS 109*, 4875–4880.

(155) Mueckler, M., and Makepeace, C. (2008) Transmembrane Segment 6 of the Glut1 Glucose Transporter Is an Outer Helix and Contains Amino Acid Side Chains Essential for Transport Activity. *J. Biol. Chem. 283*, 11550–11555.

(156) Autin, L., and Steen, M. (2006) Proposed Structural Models of the Prothrombinase (FXa – FVa) Complex. *J. Proteins Struct. Funct. Bioinforma. 450*, 440–450.

(157) Kurella, V. B., and Gali, R. (2014) Structure guided homology model based design and engineering of mouse antibodies for humanization. *Biomed. Informatics 10*, 180–186.

(158) Schwede, T. (2014) Protein Modelling: What Happened to the “Protein Structure Gap”? *Structure 21*.

CHAPTER 5

Size does not matter: A molecular insight into the biological activity of chemical fragments using thermodynamic calculations

Geraldene Munsamy[a], Ndumiso N. Mhlongo[a], Abraham J. Niehaus[a*] and Mahmoud E. S. Soliman[a*]

[a] Molecular Modelling and Drug Design Research Group, School of Health Sciences, University of KwaZulu-Natal, Westville, Durban 4000, South Africa

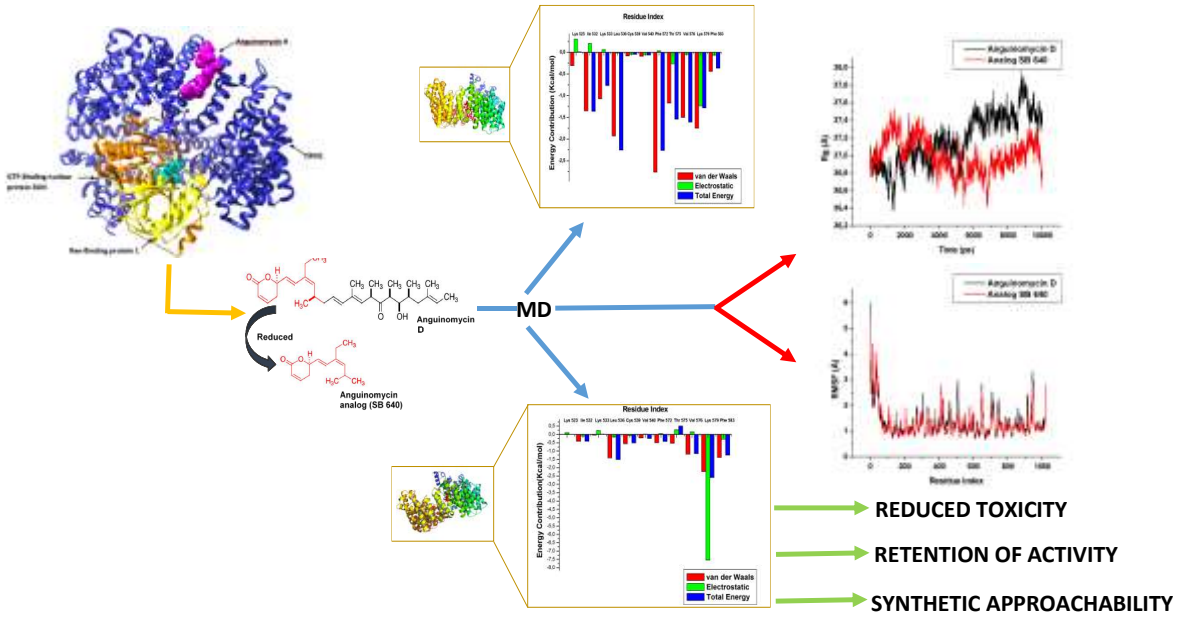
*Corresponding author: Mahmoud E. Soliman

Email: soliman@ukzn.ac.za

Telephone: +27 0312607413, Fax: +27 031260 779

Webpage: http://soliman.ukzn.ac.za/

5.1. Graphical abstract



5.2. Abstract

This study aims to give molecular insight into the activity of chemical fragments obtained from natural sources, elucidating how the reduction in their structure, may be a contributing factor to retention or enhanced biological activity, reduced structural complexity as well as an overall exhibition of desired pharmacophore activity displayed. All of which are key aspects of drug design and development. Natural compounds have timelessly been a choice in drug design and development however due to factors such as limited bioavailability and strenuous synthetic protocols, manipulation of the chemical structures obtained from these natural compounds are vastly becoming the new approach in drug synthesis. In a bid to understand and validate the newly established concept of *"reduce to maximum*" highlighting the truncation of a chemical compounds structure derived from natural sources, molecular dynamic ensembles were generated to obtain per-residue energy decomposition footprints as well as thermodynamic integrations as previously introduced in our publications. From the simulations it can be seen that the activity of the analog SB 640 derived from its parent compound Anguinomycin D displayed a slight decrease in binding affinity in comparison to the parent compound although being reduced by more than half of the C skeleton the analog still displays retention of activity due to the presence of the α, β-unsaturated δ-lactone moiety, in accordance with experimental data. The analog SB 640 also displayed reduced toxicity in comparison to the parent compound with improved chemical tractability and a significant reduction in the number of synthetic steps.

Keywords: Chemical fragment; Natural products; Anguinomycin D; Analog SB 640; Pharmacokinetic properties; MD simulation,

5.3. Introduction

Despite prominent advances in drug design and development, there has been a decline in the advancement of new drugs, threatening new therapeutic advances as well as commerciality of drug companies. With factors such as limited resources and adamant time constraints, the discovery of viable lead drug candidates is still an ongoing challenge in the field of drug development[1]. The transition of screening hits to the actual selection of potential drug targets is a task that requires both discernment as well as expertise[2,3,4,5]. Other underlying factors that influence the observable rise in the incidence of prevalent diseases can be attributed to the lack of potential drugs, limited bioavailability or accessibility of potential drug sources and lack of expedient synthetic approachability [6,7,8].

Drug resistance also plays a pivotal role in the observed increase in prevalence. It has been reported that between 44000-98000 annual deaths , an estimated 7000 deaths were due to drug resistance and adverse drug response caused by harmful side effects of drugs, lack of effectiveness, as well as the administration of interacting drugs often leading to toxicity [9].

Natural products encapsulate the wisdom of evolution as their derivatives have timelessly been sought after on a molecular level due to their pivotal role in biological processes, elucidating their contributing efforts in the advancement of human quality of life and healthcare[10]. Natural products have also been highlighted as the ideal starting point in drug discovery due to their ability to selectively and optimally bind to the target proteins[11,12]. Natural products tend to have a high binding affinity to protein targets whilst still maintaining constant entropy with limited loss and are still being sought after due to their flexibility of adopting different conformations in aqueous and lipophilic environments[13].

The structural diversity and beneficial therapeutic properties possessed by compounds obtained from plant derivatives as well as other various natural sources including marine organisms and microorganisms, inaugurate their inherent use in the development of pharmacological drugs thus unravelling a new era of drug design[14,15,16]. Chemical fragments isolated specifically from natural products have made a resurgence over the past 10 decades in the field of drug development due to the exiguous amount of innovative and effective drugs emerging from the field of pharmacological drug research and design[17,18]. Currently, over 50% of new drugs being synthesized by pharmaceutical companies are derived directly from secondary metabolites obtained from natural products. [19]

Compounds isolated from natural sources were not the first option in drug design due to the inability to determine their chemical tractability, as well as the inability to adhere to the stringent criteria following Lipinski's "rule of five"[20] and related ADME (absorption, distribution, metabolism and excretion)/pharmacokinetic aspects, which all led to the phasing out of natural products in the production of next generation therapeutics[21,22]. Although there are ample amounts of chemical fragments derived from natural compounds that are currently being used in drug development, there are still evident pitfalls as they lack adequate effectivity, reduced solubility and display observed toxicity[23,24,25].

However recent advances in technology in both organic and computational chemistry have initiated a vital transition with the use of synthetic organic chemistry, allowing the traceability of complex chemical structures of natural products which has been used to design simplified structural analogs with retained biological activity, enhanced biological properties, reduced toxicity, as well as an overall reduction in the number synthetic steps [26,27]. Such advances have allowed the modification and synthesis of compounds derived from natural resources to develop more potent drugs. The biological evaluation of the structural intricacy of natural products exemplifies the importance of reducing the chemical structure of a complex natural product in order to generate an analog structure for drug discovery [28,29].This process involves the reduction of unwanted fragments of natural products that are contributing factors to the lack of effectivity and toxicity. This may result in retention of essential fragments required to induce biological activity. Such strategic process may also aid in providing favourable structurally simplified compounds that can be easily manipulated.[30] Analysis of drugs and recognition of target

proteins through computational modeling and drug design tools may provide vital insight and guidance into the design of pharmaceutical drugs[31]. This could aid substantially in understanding the effect of structure reduction of compounds derived from natural resources and how it may further enhance optimal pharmacodynamics and desired pharmacokinetic properties essential in drug synthesis.

5.3.1. *"Reduce to maximum Concept"*

A study performed by Crane *et al*.[32] encompassing the *"reduce to maximum concept"* reported the use of fragments derived from natural compounds denoting a significant decline in molecular weight, structural intricacy as well as reduced number of synthetic steps all of which contributed to the overall increase in biological activity thus indicating potency and effectivity[14]. Derived compounds from natural products typically possess a large molecular weight, encompassing a large structure that may not be directly involved in the biological activity exhibited by the compound[33]. This large structure of natural compounds may also be a contributing factor to the observed toxicity quintessentially displayed by natural compounds[34,35]. Such structure may also induce inhibition or suppression of the biologically active fragment of the compound from exhibiting its full potent effect. A contributing factor could be the limitation of the flexibility of the fragment altering optimal binding at the active site of the receptor[36,37] Simplified analogs derived from natural products may however bind to the target protein at more optimal orientations and conformations due to the reduced molecular size, enabling the ligand to direct itself eloquently into the receptor's active site where it undergoes various internal conformational alterations, such as torsion angle rotations and translations[38]. The reduced size of the ligand may allow an adequate fit into the binding site illustrating improvement or retention of binding affinity and chemical specificity[39].

A key example would be the use of Eribulin mesylate, a synthetic analogue of Halichondrin B isolated from the marine sponge *Halichondria okadai* Kadota[40]. The synthetic analog displayed synergistic activity as the parent compound, however still superseding the desired pharmacokinetic properties of the parent compound comprising of a 35% reduction in molecular weight, requiring only half the number of synthetic steps to retain the potency elucidating its key role as an avid cancer treatment [41]. Another study encompassing the use of anti-malarial drug artemisinin and its structural analogues displayed comparable *in vitro* anti-malarial activities of the structural analogs as compared to the parent compound and were furthermore effective in animal models [42, 14].

Analog (SB 640) derived from parent compound Anguinomycin D displayed a 60% reduction in molecular weight in comparison to the parent compound presented in Figure 5.2 [14] Biological studies conducted on Anguinomycin D isolated from *Streptomyces cerevisiae* revealed that this natural compound exhibited inhibition of the chromosome maintenance region 1 (CRM1), a 120 KDa protein

that plays a crucial role in nucleocytoplasmic transport, in addition to signal transduction, immune response as well as cell differentiation[43,44]. Studies conducted on CRM1 isolated from several organisms which include human, mouse, fungi *C. thermophilum* and *S. cerevisiae*, the architecture of CRM1 appears conserved across these homologs[19,45,46]. The nuclear export signal (NES)-binding groove of CRM1 is responsible for binding to NES peptides that direct proteins out of the nucleus mediated by the CRM1 export pathway[43,47]. The NES-binding groove consists of five hydrophobic PKIαNES residues (ϕ0, ϕ1, ϕ2, ϕ3, and ϕ4) to which the NES peptide fragments in protein cargoes attach to the CRM1 for transport out of the nucleus[48]. Excessive levels of CRM1 in the cytosol have been linked to cancer, either inactivation of its tumour suppressive function or an excess of anti-apoptotic activity (onco-protein)[49,44]. In addition, aberrant CRM1 mediated transport results in the further prognosis of disease state which include cancer, viral and anti-inflammatory diseases[47,50]. Inhibition of CRM1 is an activity shared by close structural relative of the Anguinomycin family, Leptomycin B, however studies on this compound were abandoned due to its elevated levels of toxicity[19]. Natural compound Anguinomycin D exhibits remarkable biological activity which includes induction of inactivated retinoblastoma protein (pRB) glia cell death and in the process, pRB is inactivated all of which are key points in cancer progression and development.[32]

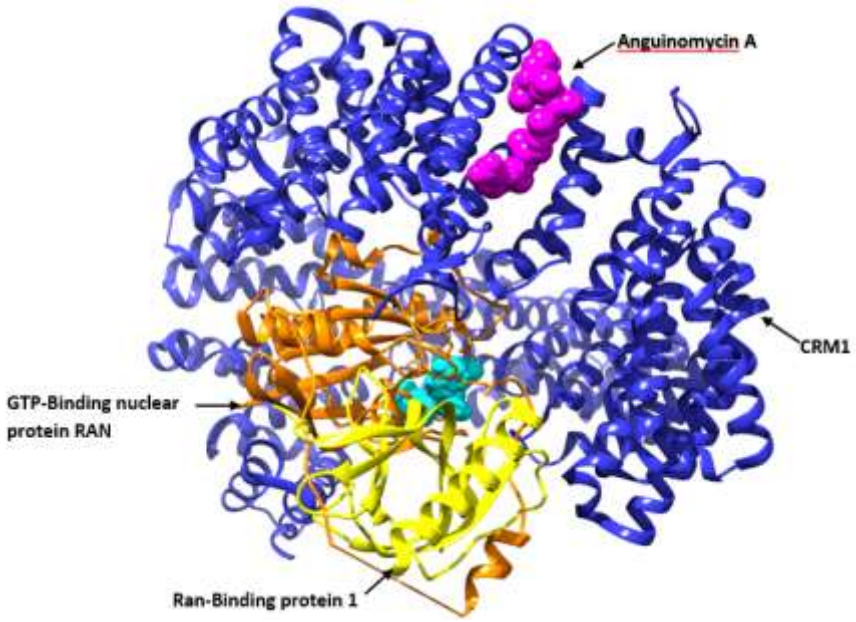


Figure 5.1. The crystal structure of CRM1 inhibitor Anguinomycin A (magenta) in complex with CRM1-RAN-RANBP1; CRM1-meduim blue, RAN-orange, RANBP1-yellow

Biological assays conducted by Gademann and colleagues illustrated biological activity retention by the synthesized simplified analog SB 640 of Anguinomycin D, displaying a substantial decrease in molecular weight, presence of one stereo-center and reduced unsaturation[10], thus displaying a substantial significance of reducing the structure of a chemical compound to generate smaller and more optimally functioning compounds that surpass or retain the biological activity of the parent compound[10].

Experimental studies conducted by Bonazzi and colleagues illustrated retention in the biological activity of the simplified structural analog (SB 640) as compared to the parent compound[19]. Due to lack of resources and in a bid to reduce time constraints, the SB 640 poses as an optimal choice in synthesis and effectivity in comparison to the naturally derived parent compound.

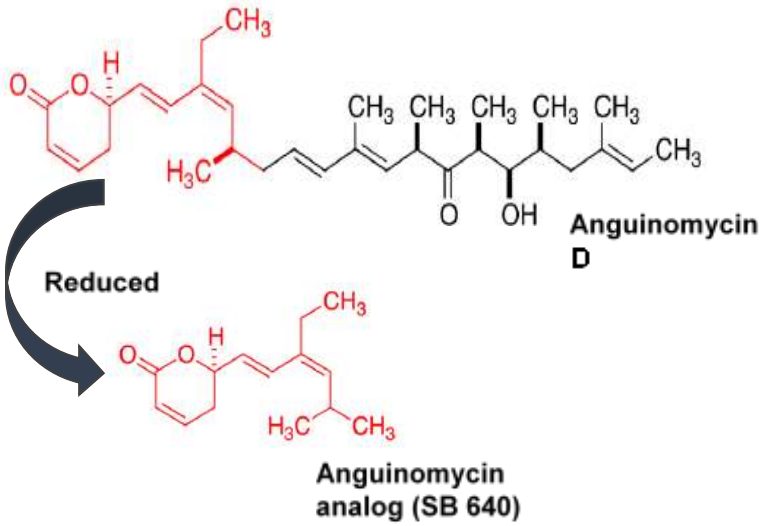


Figure 5. 2. Structural comparison of Anguinomycin D isolated from Streptomyces sp. and its simplified structural analog Anguinomycin D analogue (were constructed using Chemdraw Ultra[51])

In this study, we apply computational approaches to validate the *"reduce to maximum concept"* and provide further insight in determining which fragments of the parent compound can be truncated, as well as to aid in the synthesis of smaller and more concise compounds which exhibit a higher degree of potency and effectivity. Understanding the mechanism behind the latter concept may also aid in further development of next generation therapeutics.

We also aim to determine if the reduced compound will generate a lower binding affinity as opposed to the parent compound, as an evaluation of the total energy of interaction between the therapeutic drug and target may be minimal, in order to ensure full and adequate efficacy of the therapeutic drug.

Computational tools such as molecular docking and thermodynamic calculations render this an easy task as it advances the analysis of the level of interaction between the drug-protein complexes, ensuring adequate determination of the binding affinity. A computational approach may also aid in determining which fragment of the chemical structure is essential for biological functionality and which fragment can be truncated in order to ensure optimal activity and retention or enhanced potency.

There are various computational tools that cogently estimate ligand binding affinities with an extent of precision and efficacy[52]. Molecular dynamics simulation (MD) is identified as a key computational tool utilized in the theoretical inspection of biological molecules[53,54]. MD simulations provide intricate information on the fluctuations and conformational changes that may occur during drug-protein interactions.[55] Application of molecular dynamics enables the distinctive identification of highly contributing amino acid residues based on thermodynamic calculations incorporating per-residue energy decomposition technique. This approach allows the identification of highly contributing amino acid residues towards the total binding energy in this study MD simulations were performed for Anguinomycin D and analog SB 640 each in complex with CRM1. Post MD analyses were further performed on the respective systems including thermodynamic calculations to determine binding free energy contributions of the parent compound Anguinomycin D and the analog SB 640-CRM1 systems. Substantiating the role of MD and thermodynamic calculations in order to understand the structural composition, dynamics and thermodynamics of biological molecules of each system. It is also utilized to further identify commonly shared amino acid residues that contribute substantially towards the total binding energy of each system The RMSD, RMSF and Rg of each system were also determined as well as the predicted toxicity profile of each compound.

5.4. Computational Methods

5.4.1. Systems Preparation

The crystal structure of CRM1 in complex with Anguinomycin A was obtained from the Protein Data Bank, PDB code: 4HAV[48]. UCSF Chimera[56] was used to observe and manipulate the crystal structure, where chains A and B, H_2O molecules, Mg atoms, Cl atoms and a phosphoaminophosphonic acid-Guanylate ester (GNP) as well as Anguinomycin A were removed.

The ligands Anguinomycin D and derived analog SB 640 were drawn using ChemDraw Ultra 9.0 [51]. Each ligand was individually analysed in MMV[57], Chimera [56] as well as Gaussview 5.0 [58]software to ensure the correct angle bonds and hybridization were visible. The ligands were individually minimized incorporating the steepest descent method and MMFF94S force field in Avogadro[59]. Receptor modification visualization was accomplished using Chimera[56] and MMV [57,60,61].

5.4.2. Molecular Docking

Molecular docking aims to provide the most energetically favourable binding pose as the ligand positions itself within the binding cavity of the target protein. Molecular docking was performed using AutoDock [62]a well-established docking program. Autodock Vina[62] was used to generate the calculations obtained from the docking scores. During the process of docking, Geister partial charges were added whilst Autodock atom types were defined using the Autodock Graphical user interface supplied by MGL tools.[63] The Lamarckian Genetic Algorithm was applied to determine the docked conformations,

a docking technique that is considered rather reliable and adequate[64,65]. Raccoon [62]software was used to convert the files into pdbqt format in assembly for docking[66]. The gridbox was defined using Autodock Vina with the grid parameters being X = -32.777, Y = 72.797 and Z = 32.985 for the dimensions and X = 36, Y = 30 and Z = 30 for the center grid box[67]. Molecular visualization of the docked complexes was conducted using Chimera software and the LigPlot[68] program.

5.4.3. Molecular Dynamic Simulations

MD simulations were performed on four systems: a parent compound Anguinomycin D-CRM1 complex; derived structural analog SB 640-CRM1 complex, docked Anguinomycin A-CRM1 complex as well as the original complex of Anguinomycin A bound to CRM1 retrieved from the RSCB Protein Data Bank. Using the GPU version of PMEMD incorporated in the Amber 14 package, MD simulations were conducted for these systems [69, 70]. Protein systems were modelled using the standard AMBER (FF99SB) force field[71] present in the Amber 14 package. The LEaP module of Amber 14 was employed to add hydrogen atoms and counter ions to aid in the stabilisation of the system[70]. Optimisation of ligands was performed by addition of partial atomic charges, which were calculated encompassing the restrained electrostatic potential (RESP)[59,60]. Neutralisation was achieved by the addition of Na^+ ions which were performed on all systems. Atomic partial charges comprising of General Amber Force Field (GAFF) were prompted by the ANTECHAMBER module[72]. Complete solvation of the systems was attained within a TIP3P waterbox consisting of a buffering distance of 8 Å in the midst of water and the protein surface at box extremity.[73] Long-range electrostatics interactions were performed adhering to the particle mesh Ewald (PME) method administered in the Amber 14 package with a van der Waals limitation of a distance of 10 Å.

Systems were subjected to consecutive partial partial minimization and full minimization steps. Initial energy minimization with a 500 kcal/mol $Å^2$ restraint potential related to the solute, were performed incorporating the steepest descent method for 1000 iterations which were finally followed by 1000 iterations of conjugate gradient minimization. MD simulations were carried out for 10 ns during which the system was heated moderately between a range of 0-300K regulated with the aid of the Langevin thermostat[74]. Systems were equilibrated at 300K under 1 atm pressure whilst retaining force constraints on the restrained solute for 500ps prior to production runs, followed by removal of restraints and maintenance of a constant pressure (1 bar) using a Berendson barostat.[75] All atoms covalently bound to a hydrogen atom were subjected to the SHAKE algorithm throughout the MD simulation.[76] From experimental studies, the preferred pH of the system was at 6.6 which was kept constant and validated in accordance with data projected from the H++[77] tool which computes the pKa value of ionisable groups.[78] All simulations were run at a 2 fs time step and the SPFP precision module. Trajectories were saved and analysed every 1 ps. The root mean square deviation (RMSD)[79], root mean square fluctuation

(RMSF)[80] and radius of gyration (Rg)[81] were calculated using the PTRAJ and CPPTRAJ modules found in the Amber 14 package. Results were analyzed and plotted using Origin software[82].

5.4.4. Thermodynamic calculations

Thermodynamic calculations assist in determining the binding free energy as an endpoint calculation which provides indispensable information about the ligand-protein interaction. There are various computational tools used in the coherent determination of binding affinities namely thermodynamic integration (TI)[83], free energy perturbation (FEP)[84], Molecular Mechanics Generalized Born Surface Area (MM/GB-SA)[85], Molecular Mechanics Poisson-Boltzmann Surface Area (MM/PB-SA)[86], linear response (LR)[87], and fluctuation-dissipation theorem (FDT)[78,79]. All these methods provide meaningful insight, however there is a generated focus on the MM/PBSA and MM/GBSA methods as they are more computationally efficient and encompass a versatile range of parameters for each energy term. The MM/PBSA and MM/GBSA methods are employed to calculate binding free energies for macromolecules integrating continuum solvation models merging with molecular mechanics calculations[86]. The MM/GB-SA and MM/PB-SA methods are applied to determine the binding free energy estimation whilst the MM/GB-SA method presented in Figure 5.3 is employed to establish the per-residue energy decomposition of the highest contributing amino acid residues to the total binding free energy[88,89]

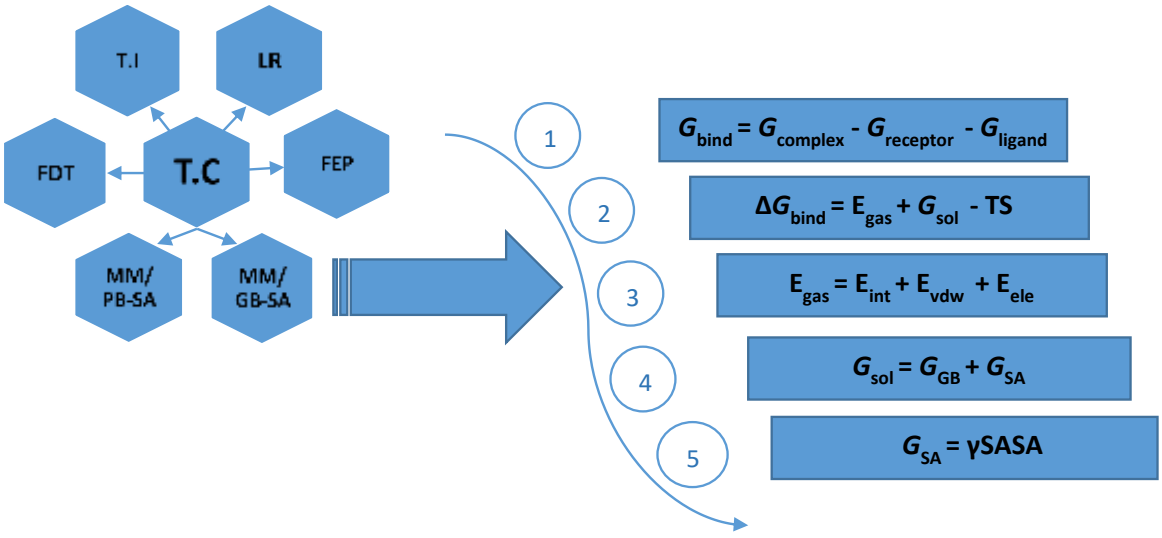


Figure 5.3. The different computational approaches used to determine binding affinity, highlighting the MM-GBSA approach

E_{gas} denotes the gas-phase energy; E_{int} indicates the internal energy; and E_{ele} and E_{vdW} are the Coulomb and van der Waals energies, respectively. Egas is determined incorporating the FF99SB force field terms. The solvation free energy, denoted by G_{sol}, is a direct contribution of the polar and non-polar states where the polar solvation contribution G_{GB} can be further calculated by resolving the GB

equation. The contributing value of the non-polar solvation can be approximated directly from the Solvent-accessible surface area (SASA) estimated using a water probe radius of 1.4 Å. TS subsequently correspond to temperature and total solute entropy. In order to determine the individual amino acid contribution towards total binding free energy between CRM1 and parent compound Anguinomycin D and derived analogue SB 640 a decomposition analysis of the interaction energy for each residue was computed by using the MM/GBSA binding free energy decomposition protocol in Amber 14 package.

5.4.5. PASS Toxicity Prediction

Compound toxicity may result from a series of interactions that may occur in relation to a range of biological targets within an organism. Toxicity can be characterized by an interaction with a single target or from a range of interactions with a versatile group of targets within an organism. Therefore determination of the toxicity of a particular chemical compound remains an intricate process.[90] The online tool PASS (Prediction of Activity Spectra for Substances)[91] incorporates an *in-silico* approach that generates the predicted biological activities of selected chemical compounds, their mechanisms of action and related levels of toxicity interacting with a single or a range of biological targets within an organism. PASS prediction is based on the analysis of the chemical compounds structure relative to the biological activity exhibited for about 60 000 biologically active compounds[92,91]. The Pa and Pi values vary from 0.000 to 1.000 and the probabilities Pa +Pi≠1 are calculated independently. The selected structures of Anguinomycin D and SB 640 were generated using ChemDraw Ultra 9.0.[51] The sdf files were then submitted for toxicity prediction in the PASS online programme.

5.5. Results and Discussion

5.5.0. Post MD analysis

Post MD analysis is utilized to analyse the results obtained from Molecular dynamic simulations. It can be used to understand the relation between protein secondary structure and internal motions[59].

5.5.1. Validation of docking

Molecular docking validation is pivotal as it ensures the ligand is in an accurate conformational pose as it binds to the active site of the target protein. This can be established by validating the grid box size and coordinates center along the binding pocket. During the process of molecular docking, the size and center of the coordinates of the grid box where the ligands bind must be validated to ensure certainty that the correct conformation pose is established[93]. In a bid to validate our docking approach, we removed Anguinomycin A from the crystal structure of CRM1 in complex with RAN-RANBP1 (PDB: 4HAV) and re-docked it into a low energy structure of the same active site. The docked Anguinomycin

A conformational pose was similar to that of the crystallographic pose with an RMSD less than 0.05 Å, implying that the applied AutoDock docking parameters were conducive for this system. A superimposed image of the crystallized complex along with an energy-minimized structure are presented in the supplementary material (S4) implying that the protocol implemented during simulations are of adequate reliability and genuine authenticity.

5.5.2. Root mean square fluctuation

The biological properties of a protein may be reliant on its physical interaction with other molecules, as exhibited upon ligand binding and dissociation that may induce conformational changes essential to a CRM1 function. Root mean square fluctuation (RMSF) is calculated to determine the mobility of individual residues within a protein[59]. Figure 5.4 displays the RMSF of Anguinomycin D and analog SB 640-CRM1 complexes over the 10 ns simulation.

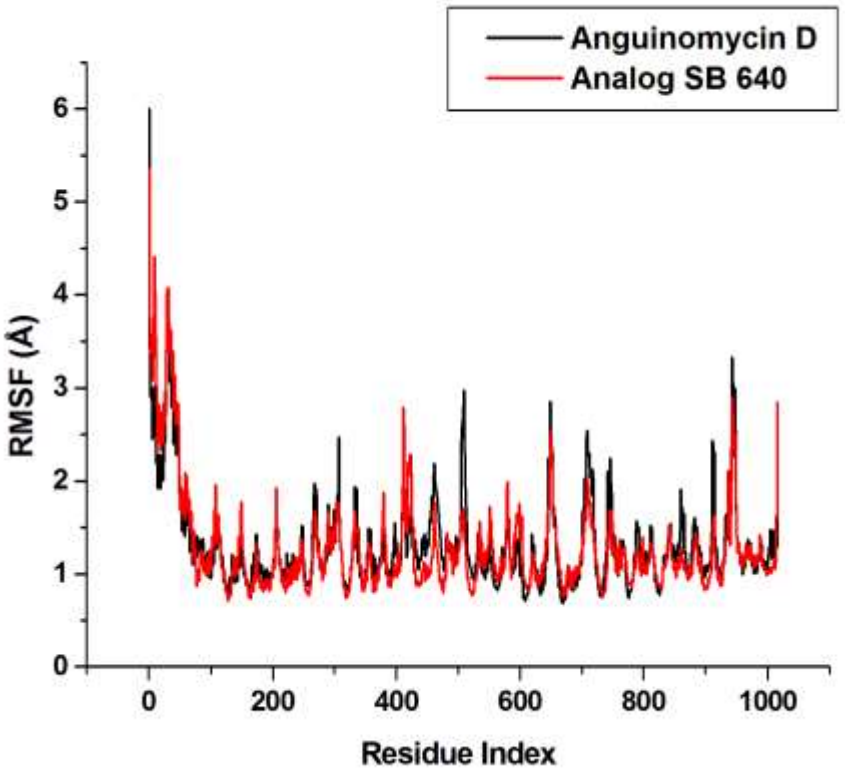


Figure 5.4. RMSF of Anguinomycin D-CRM1 and derived analogue SB 640-CRM1 complexes.

The average RMSF values of Anguinomycin D and SB 640 are 1.34 Å and 1.29 Å respectively, exhibiting a slight difference of 0.057Å. Both systems displayed rigid stability throughout the simulations. From visual inspections, there is evident residue fluctuation predominantly in non-binding

residues. The structural and active site residues display minimal fluctuation throughout the simulation of the two-systems, substantiating that the binding activity of analog SB640 may be analogous to that observed from the parent compound Anguinomycin D whilst displaying marginal perturbation in the stability of the protein backbone.

5.5.3 Radius of gyration

The radius of gyration (Rg) measures the thermodynamics and kinetics of protein folding and may aid in determining the compactness and stability of protein complexes[94]. Figure 5.5 displays the radius of gyration of Anguinomycin D and analog SB 640 both in complex with CRM1.

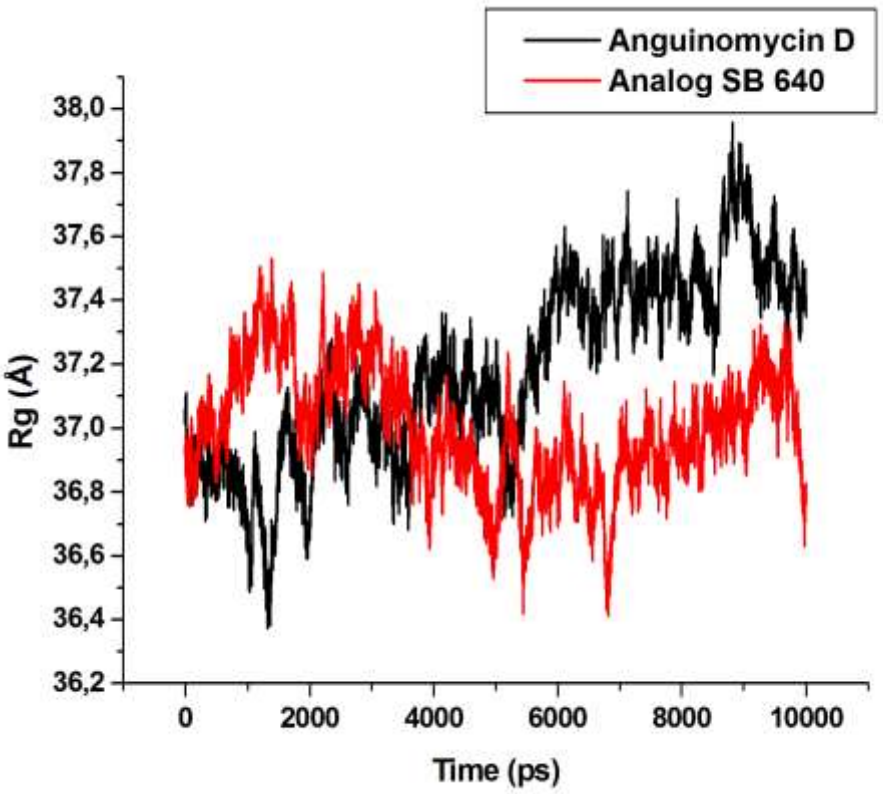


Figure 5.5. Radius of gyration of C-alpha atoms of Anguinomycin D-CRM1 and analog SB 640-CRM1 complexes

The Rg of Anguinomycin D and analog SB 640 complexed with CRM1 was determined at 300 K. From the conformational analysis, the Rg of both systems displayed significant difference with an average Rg value of 37.18 Å for the parent compound Anguinomycin-CRM1 complex as compared to the analog-CRM1 complex which was 37.00 Å. SB 640 displays reduced fluctuation of 0.18 Å in comparison to Anguinomycin D in complex with CRM1, depicting an overall stable SB640-CRM1 system as opposed to the parent Anguinomycin D-CRM1 system. The analog SB 640 displayed reduced stability during the time interval of 0-2000 ps as opposed to the parent compound, showing that

Anguinomycin D-CRM1 complex reached stability much earlier in the simulation as opposed to the analog. However between 3500-8000 ps the Anguinomycin D-CRM1 system displayed minimal stability in comparison to the analog SB 640-CRM1 system. The observed decline in the Rg may be attributed to the hydrophobic interactions presented in Figure 5.7 as the analog SB 640 positions itself within the active site of CRM1 protein resulting in a system displaying greater stability.

5.5.4. MM/GBSA Binding Free Energy Calculations

Binding free energy calculations were computed using the MM/GBSA protocol. MM/GBSA calculations are based on molecular dynamics simulations of the receptor-ligand complex and hence are intermediate in both accuracy and computational effort between empirical scoring and precise alchemical perturbation methods[95,96]. This method determines the magnitude of the binding affinity of the contributing amino acid residues.[97,39] . The calculated binding free energy (ΔG_{bind}) of the parent compound Anguinomycin D and analog SB 640 in complex with CRM1 are -35.10 kcal/mol and -29.13 kcal/mol respectfully.The analog SB 640 displayed a minimal reduction of ~6 Kcal/mol , this is in great correspondence with experimental studies conducted whereby the analog displayed a marginal decrease, however managed to retain biological activity comparable to that displayed by the parent compound[19]. Despite drastic simplification and total reduction in the polyketide chain of Anguinomycin D and SB 640, the structurally simplified analog (SB 640) still displayed high activity which can be attributed to the essential presence of the α, β-unsaturated δ-lactone moiety. From a synthetic perspective, the synthesis of the analog SB 640 is reduced by 60% thus resulting in a gain of time as well as resources. From previous studies conducted on natural products, a reduction of more than one-half of the Carbon skeleton often leads to complete loss of bioactivity[19]. However, from the binding energy exhibited by SB 640 and experimental studies conducted, this has proven to challenge this concept, introducing the possibility of a drastic reduction in the total number of synthetic steps as well as retention of biological activity. The calculated van der Waals contributions (ΔE_{vdW}) to the total binding free energy in the Anguinomycin D-CRM1 complex (-40.83 kcal/mol) are higher than that of the SB 640-CRM1 complex (-29.35kcal/mol). The observed van der Waals interactions between the parent compound and CRM1, as presented in Figure 5.6, can be attributed to the presence of a polyketide chain consisting of a carbonyl group at C_{17} and hydroxyl group at C_{18}, which is totally omitted in SB 640. Although there was a drastic reduction in size of the analog SB 640, the electrostatic interactions (ΔE_{elec}) of SB 640 was -11.95kcal/mol superseding that of the parent compound Anguinomycin D which was -4.11kcal/mol. From Table 1, it can observed that the electrostatic force between the analog SB 640-CRM1 complexes is greater than that of the Anguinomycin D-CRM1 system with electrostatic forces of -11.95±7.03 kcal/mol and -4.11±5.08 kcal/mol respectively. The difference in the electrostatic forces between these systems can be attributed to the ability of SB 640 to optimally position itself deeper within the hydrophobic binding groove of the CRM1 protein as opposed

to the conformational pose taken by the parent compound Anguinomycin D presented as depicted in Figure 5.6.

Table 5.1. MM/GBSA binding free energies

Hit	**Energy Components (kcal/mol)**				
Rank	**ΔE_{vdW}**	**ΔE_{elec}**	**ΔG_{gas}**	**ΔG_{solv}**	**ΔG_{bind}**
Anguinomycin D	-40.83±3.48	-4.11±5.08	-44.94±6.38	9.84±4.71	35.10±3.28
Analog SB 640	-29.35±3.18	-11.95±7.03	-41.94±7.80	12.16±5.77	29.13±3.49

Figure 5.6 presents Anguinomycin D bound to the NES binding groove of the CRM1 protein.

Anguinomycin D displays higher van der Waals contributions to the binding of the NES binding groove of CRM1. This is due to the presence of the polyketide chain which consists of a carbonyl group at C_{17} and hydroxyl group at C_{18}, absent in the analog SB 640.

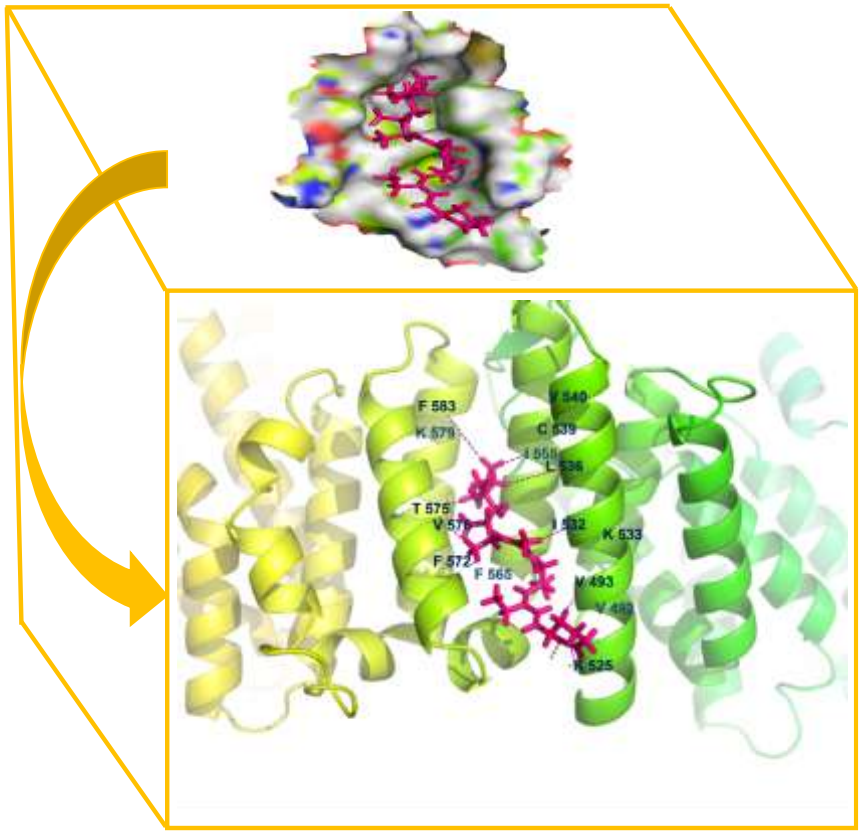


Figure 5.6. 3D structure of Anguinomycin D in complex with CRM1 showing interactions with binding site residues

Figure 5.7 presents the analog SB 640 bound to the NES binding groove of the CRM1 protein. The minimized structure of the analog SB 640, in which the polyketide chain is omitted allows the analog to eloquently penetrate much deeper into the groove within the hydrophobic binding groove of the CRM1 as presented in Figure 5.7.

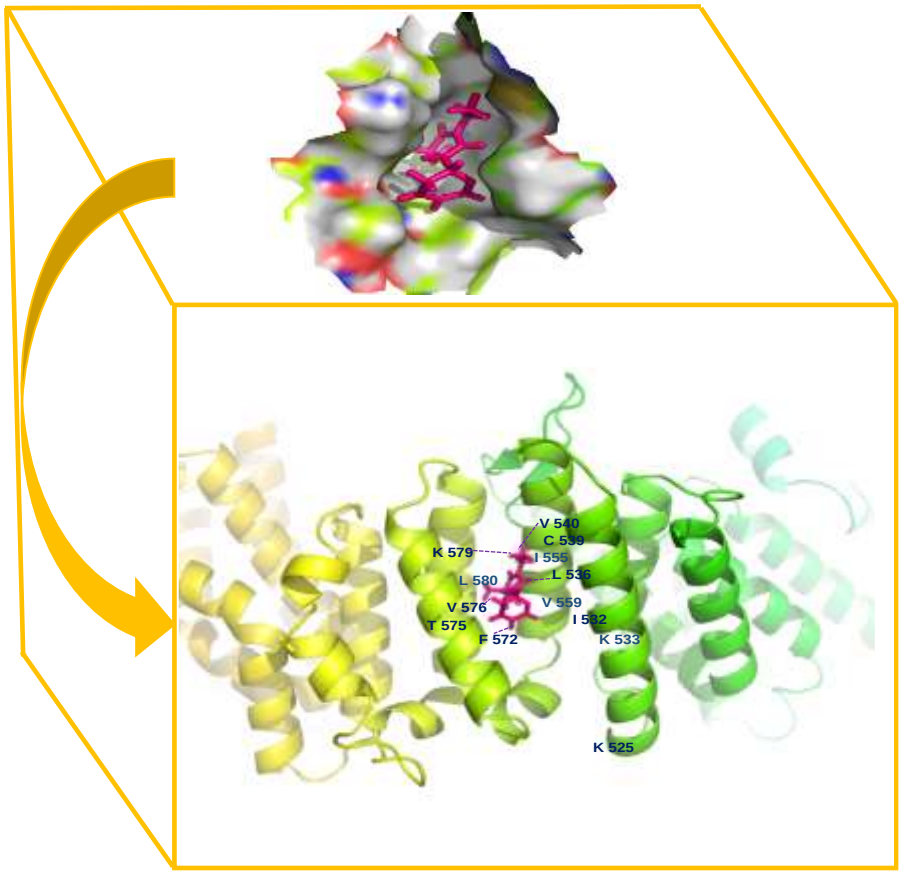


Figure 5.7 . Representative structure for analog SB 640 in complex with CRM1 Anguinomycin D with 2D and 3D graphical representation of the different binding forces.

5.5.5. Per–residue Energy Decomposition Analysis

Per-residue energy decomposition enables the analysis of the protein-drug interaction establishing the highest contributing amino acid residues towards the total binding energy. Analysis of the most energetically favourable residues highlighted major contributions from amino acid residues Leu 536, Thr 575, Val 576 and Lys 579 favoured by van der Waals forces in combination with electrostatic interactions commonly shared amongst both ligands. Electrostatic forces aid in inhibitor molecules gaining binding energy, contributing substantially to the overall total binding energy. Amino acid Lys 579 provided favourable electrostatic energy contributions of -1.282 and -2.583 Kcal/mol to the total binding energy, which can be observed between Anguinomycin and analog SB 640 CRM1 bound complexes respectively. From experimental studies Cys 539 is highlighted as an essential hot spot residue, which plays a pivotal role in the binding of potential inhibitors to the NES binding region of CRM1. However from MM/GBSA calculations, Cys 539 displayed relatively less favourable energy

contributions of 0.05 and 0.50 kcal/mol towards the binding of Anguinomycin D and analog SB 640 to CRM1 respectively. Underlying factors such as conformational plasticity and steric repulsion explain why molecular modelling of Anguinomycin D and related analogs into a rigid NES-binding groove may be different from that depicted in experimental studies. As steric repulsion between certain atoms pairs, may be intimately related to the chosen functional form for the non-bonded energy. Another contributing factor may be the balance between the functional form and the angle bend/torsional terms ultimately affecting the shape, conformation and reactivity of the conformational pose of the ligand as it fits itself within the active site of the target protein.

The NES peptide fragments in protein cargoes are essential as they attach to the CRM1 for transport out of the nucleus and restriction of the NES peptide fragments would result in limited transportation of onco-proteins out of the nucleus[49,50]. Natural compound Anguinomycin D occupies four of five hydrophobic PKIαNES residues (ϕ0, ϕ1, ϕ2, ϕ3, and ϕ4) at the NES binding site, whereas analog SB 640 being reduced 60%, occupies three of the five PKIαNES residues despite its massive reduction size. Extensive inhibition of the NES groove of CRM1 is attributed to the ability of inhibitors to overlap and occupy the majority of the groove displacing the residing NES peptides, thus explaining the extensive spectrum of nuclear export blocking exhibited by Anguinomycins and its derivatives.

Table 5.2. MM/GBSA per-residue energy decomposition

Residue	**ΔE_{vdw}**	**ΔE_{ele}**	**$\Delta G_{polar\ solv}$**	**$\Delta G_{non\text{-}polar\ solv}$**	**ΔG_{bind}**
Lys 525	-0.304 ±0.264	0.306±0.868	0.060±0.900	-0.051±0.057	0.011±0.219
	-0.002± 0.001	0.111±0.112	-0.100±0.112	0.000±0.000	0.009±0.002[a]
Ile 532	-1.350±0.319	0.206±0.137	-0.065±0.169	-0.152±0.040	-1.362±0.303
	-0.402±0.190	-0.083±0.081	0.101±525	-0.025±0.020	-0.410±0.194[a]
Lys 533	-1.071±0.475	0.063±3.657	0.477±3.439	-0.233±0.114	-0.764±0.452
	-0.040±0.010	0.224±0.383	-0.157±0.380	0.000±0.000	0.026±0.015[a]
Leu 536	-1.925±0.43	-0.013±0.122	0.152±0.112	-0.467±0.071	. -2.253±0.469
	-1.401±0.352	-0.150±0.268	0.356±0.315	-0.311±0.057	-1.506±0.334[a]
Cys 539	-0.002±0.062	-0.046±0.053	0.089±0.056	-0.005±0.008	-0.045±0.063
	-0.544±0.325	-0.069±0.518	0.239±0.378	-0.115±0.030	-0.499±0.298[a]
Val 540	-0.089±0.07	-0.053±0.055	0.094±0.062	-0.009±0.017	-0.058±0.084

	-0.205±0.225	0.020±0.142	-0.009±0.138	-0.048±0.061	-0.241±0.281[a]
Phe 572	-2.579±0.531	0.038±0.274	0.831±0.222	-0.368±0.058	-2.259±0.454
	-0.489±0.163	0.051±0.138	0.117±0.162	-0,093±0.034	-0.415±0.135[a]
Thr 575	-1.172±0.476	-0.261±0.921	0.119±0.619	-0.232±0.066	-1.543±0.646
	-0.520±0.237	0.281±0.350	-0.178±0.310	-0.085±0.047	-0.503±0.241[a]
Val 576	-1.496±0.405	-0.050±0.270	0.080±0.213	-0.141±0.029	-1.606±0.439
	-1.173±0.365	0.158±0.123	-0.013±0.144	-0.118±0.028	-1.146±0.351[a]
Lys 579	-1.745±0.303	-1.231±1.071	2.010±1.081	-0.316±0.041	-1.282±0.453
	-2.227±0.464	-7.538±3.989	7.572±3.432	-0.392±0.053	-2.583±0.771[a]
Phe 583	-0.440±0.208	-0.067±0.076	0.240±0.090	-0.095±0.053	-0.362±0.195
	-1.379±0.367	-0.278±0.202	0.616±0.233	-0.194±0.043	-1.234±0.343[a]

[a] SB 640

The highest contributing amino acid residues towards the total binding energy of the Anguinomycin D-CRM1 complex. Substantial energy contributions were evident from residues Phe 572, Leu 536, Val 576, Thr 575 and Ile 532, contributing -2.26, -2.25, -1.61, -1.54 and -1.36 kcal/mol to the total binding energy respectively (Figure 5.8). Phe 572 is identified as the energetically favourable residue with a substantial energy contribution of -2.259 Kcal/mol. Its functional aromatic ring forms a hydrophobic interaction with the carbonyl group present on the aromatic ring present in Anguinomycin D. Lys 579 is another energetically favourable residue as the charged NH group of Lys 579 interacts with C_1 and C_{13} atoms present in the Anguinomycin D. Contributions from residues Thr 575 and Ile 532 were generated through hydrophobic interactions with Anguinomycin D presented in Figure 5.8.

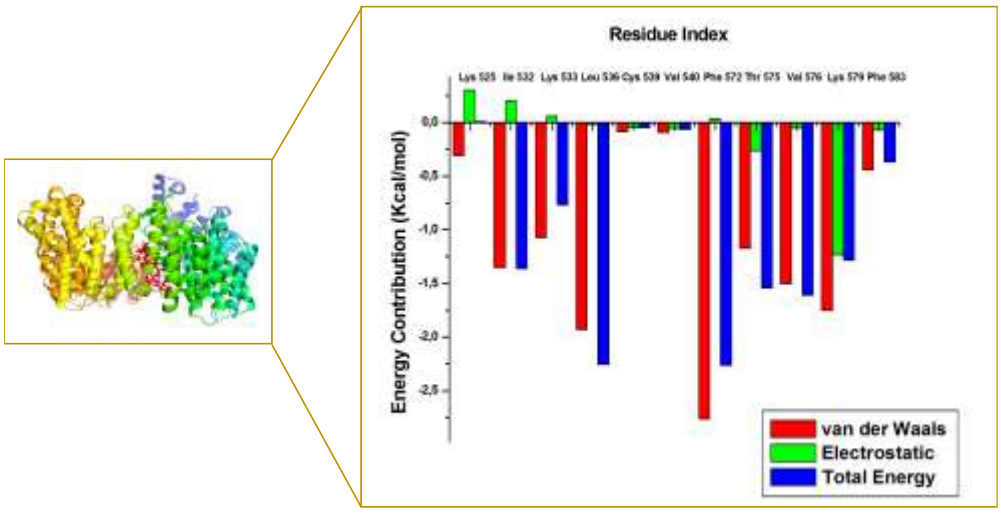


Figure 5.8. Per-residue energy decomposition contributions of Anguinomycin D.

Its functional aromatic ring forms a hydrophobic interaction with the carbonyl group present on the aromatic ring present in Anguinomycin D. Lys 579 is another energetically favourable residue as the charged NH group of Lys 579 interacts with C_1 and C_{13} atoms present in the Anguinomycin D. Contributions from residues Thr 575 and Ile 532 were Amino acids Leu 536, Thr 575, Val 576, and Phe 583 all contributed substantially with electrostatic and van der Waal forces to the total binding energy observed between analog SB 640 and CRM1 complex presented in Figure 5.9. Lys 579 is identified as an energetically favourable amino acid residue, with a significant electrostatic energy contribution of -2.583 Kcal/mol to the total binding energy. The charged NH group of Lys 579 interacts with the terminal carbon atoms present in the analog SB 640 as present in Figure 5.7.

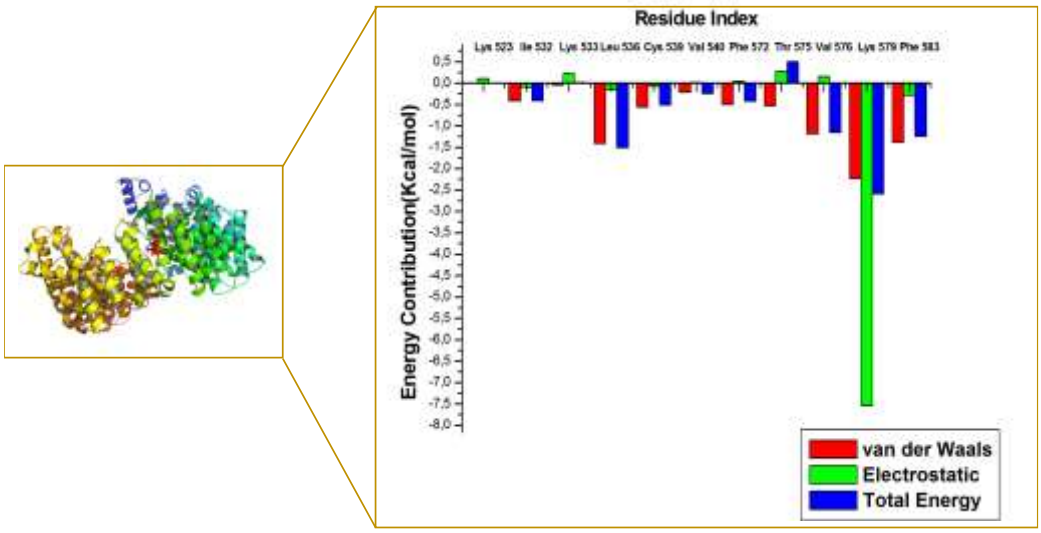


Figure 5.9. Per residue energy decomposition contributions of Analog SB 640.

5.5.6. PASS Toxicity Prediction

Prediction of activity spectra for substances (PASS)[98] is an online software tool that can be utilised to generate the predicted spectrum of a compound's potential biological activity in addition to predicted toxicity and computes it as probable activity (Pa) and probable inactivity (Pi). The PASS predicted toxicity profile for Anguinomycin D and analog SB 640 are presented in Table 5.3. It can be observed from Table 5.3 that the toxicity profile whereby Pa>Pi for Anguinomycin D displays a higher probability of predicted toxicity in comparison to its counterpart SB 640. The range of toxicity varies, using an index whereby Pa>0.7 indicates that significant activity that can be observed experimentally; 0.5<Pa<0.7 indicates that activity can be observed however minimal activity can be observed experimentally. Although this may be the case, the compound holds great biological significance as it may not be similar to current pharmaceutical agents. From the results projected in Table 5.3, it can be observed that the analog SB 640 displayed an overall reduction in the toxicity levels in comparison to the parent compound Anguinomycin D. Anguinomycin D displays increased predicted toxicity as an immunosuppressant, DNA and RNA synthesis inhibitor, which supersedes that predicted by the analog SB 640. The observed increase in toxicity levels displayed by Anguinomycin D may be attributed to the polyketide chain present in its structure which is omitted in the analog SB 640. Although the polyketide chain is predominantly associated with a vast range of biological activity, recent studies have unveiled that its synthesis is affiliated with toxin production[99]. The presence of the polyketide chain is also found in Leptomycin B which is a close structural relative of the Anguinomycin family. Leptomycin B was found to be associated with dose-limiting toxicity and thus prompted its discontinuation as an inhibitor of CRM1 exportin protein[14].

Table 5.3. Toxicity Prediction of Anguinomycin D and SB 640.

Activity	**Pa**	**Pi**
Immunosuppressant	0.801	0.005
	0.745	0.012[a]
DNA synthesis inhibitor	0.672	0.007
	0.402	0.028[a]
RNA synthesis inhibitor	0.642	0.002
	0.488	0.008[a]
HMG CoA synthase inhibitor	0.574	0.001

	0.506	0.002[a]
Protein synthesis inhibitor	0.509	0.006
	0.500	0.006[a]
ATPase inhibitor	0.496	0.002
	0.350	0.006[a]
Acylaminoacyl-peptidase inhibitor	0.443	0.023
	0.282	0.042[a]
Lactase inhibitor	0.413	0.055
	0.294	0.146[a]
Electron Transport Complex I inhibitor	0.047	0.022
	0.045	0.026[a]
HIF1A expression inhibitor	0.444	0.079
	0.348	0.135[a]

Pa: Coefficient of Activation **Pi**: Coefficient of Inhibition

[a] SB 640

5.6. Conclusion

Natural product analogs have timelessly displayed improved characteristics of solubility as well as pharmacokinetics and are thus constantly being targeted in drug design and synthesis. From the thermodynamics calculations projected in this study, it is observed that the derived analog SB 640 exhibited slight reduction in binding affinity whilst still maintaining an overall retention of biological activity. Although being reduced by 60% in molecular weight, the analog SB 64 illustrated retention of biological activity in comparison with the parent compound Anguinomycin D, which is in correspondence with experimental studies conducted [14]. The toxicity profile of SB 640 in comparison to the parent compound was substantially reduced, further highlighting that a reduction in the size of a compound may result in reduced toxicity. These fragments provide an ideal initial platform to assist medicinal chemists in drug design as they portray vital interactions thus reducing time constraints whilst and production costs. The use of structural analogs would aid in synthesis being much more time

efficient and thus bridging the gap of a lack of bioavailability of natural products. Further investigations into the “reduce to maximum” concept may be pivotal in expanding the horizon of the drug discovery.

Supplementary Information

RMSD of the C-alpha backbone of CRM1 in complex with Anguinomycin D and analog SB 640 provided in the supplementary material.

Acknowledgments

The authors like to acknowledge the National Research Foundation of South Africa and the School of Health Sciences, University of KwaZulu-Natal, Westville, for financial support and the Centre for High Performance Computing (CHPC) for computational resource.

Conflicts of Interest

Autors declare no potential conflicts of interest.

References

(1) Garg, G., Gibbs, J., Belt, B., Powell, M. A., Mutch, D. G., Goedegebuure, P., Collins, L., Piwnica-worms, D., Hawkins, W. G., and Spitzer, D. (2014) Novel treatment option for MUC16-positive malignancies with the targeted TRAIL-based fusion protein Meso-TR3 Novel treatment option for MUC16-positive malignancies with the targeted TRAIL-based fusion protein Meso-TR3. *BMC Cancer 14*, 1471–2407.

(2) Paul, S. M., Mytelka, D. S., Dunwiddie, C. T., Persinger, C. C., Munos, B. H., Lindborg, S. R., and Schacht, A. L. (2010) How to improve R&D productivity: the pharmaceutical industry's grand challenge. *Nat. Rev. 9*, 203–214.

(3) Sams-dodd, F., and Sams-dodd, F. (2005) Target-based drug discovery : is something wrong ? define rational drug discovery programs . REVIEWS. *Drug Discov. Today 10*, 139–147.

(4) Bax, R., Mullan, N., and Verhoef, J. (2000) The millennium bugs — the need for and development of new antibacterials. *Int. J. Antimicrob. Agents 16*, 51–59.

(5) Projan, S. J. (2003) Why is big Pharma getting out of antibacterial drug discovery ? *Curr. Opin. Microbiol. 6*, 427–430.

(6) Murray, C. J. L., and Lopez, A. D. (1997) Global mortality , disability , and the contribution of risk factors : Global Burden of Disease Study. *Lancet 349*, 1436–1442.

(7) Morse, S. S. (1995) Factors in the Emergence of Infectious Diseases. *Emerg. Infect. Dis. 1*, 7–15.

(8) Cuatrecasas P. (2006) Drug Discovery in Jeopardy. *Sci. Soc. 116*, 2837.

(9) Alomar, M. J. (2014) Factors affecting the development of adverse drug reactions (Review article). *Saudi Pharm. J. 22*, 83–94.

(10) Gademann, K. (2015) Copy, Edit, and Paste: Natural Product Approaches to Biomaterials and Neuroengineering. *Acc. Chem. Res. 48*, 731–739.

(11) Raskin, I., Ribnicky, D. M., Komarnytsky, S., Ilic, N., Poulev, A., Borisjuk, N., Brinker, A., Moreno, D. A., Ripoll, C., Yakoby, N., Neal, J. M. O., Cornwell, T., Pastor, I., Fridlender, B., Ribnicky, D. M., Moreno, D. A., and Raskin, I. (2002) Plants and human health in the twenty-first century. *Trends Biotechnol. 20*, 522–531.

(12) Bathula, S. R., Akondi, S. M., Mainkar, P. S., and Chandrasekhar, S. (2015) "Pruning of biomolecules and natural products (PBNP)": an innovative paradigm in drug discovery. *Org. Biomol. Chem. 13*, 6432–6448.

(13) p92813-natural-product-pharmaceuticals-the-third-generation;winter-08 @ www.ddw-

online.com.

(14) Crane, E. A., and Gademann, K. (2016) Capturing Biological Activity in Natural Product Fragments by Chemical Synthesis. *Angew. Chemie Int. Ed. 55*, 2–23.

(15) Lahlou, M. (2013) The Success of Natural Products in Drug Discovery. *Sci. Res. 2013*, 17–31.

(16) Sams-dodd, F., and Sams-dodd, F. (2005) Optimizing the discovery organization for innovation innovation in the early drug discovery process . REVIEWS. *Drug Discov. Today 10*, 1049–1056.

(17) Newman, D. J., and Cragg, G. M. (2007) Natural Products as Sources of New Drugs over the Last 25 Years ⊥. *J. Nat. Prod. 70*, 461–477.

(18) Atanasov, A. G., Waltenberger, B., Pferschy-wenzig, E., Linder, T., Wawrosch, C., Uhrin, P., Temml, V., Wang, L., Schwaiger, S., Heiss, E. H., Rollinger, J. M., Schuster, D., Breuss, J. M., Bochkov, V., Mihovilovic, M. D., Kopp, B., Bauer, R., Dirsch, V. M., and Stuppner, H. (2015) Discovery and resupply of pharmacologically active plant-derived natural products : A review. *Biotechnol. Adv. 33*, 1582–1614.

(19) Bonazzi, S., Eidam, O., Gu, S., Wach, J., Zemp, I., Kutay, U., and Gademann, K. (2010) Anguinomycins and Derivatives : Total Syntheses , Modeling , and Biological Evaluation of the Inhibition of Nucleocytoplasmic Transport. *J. Antimicrob. Chemother. 132*, 1432–1442.

(20) Lipinski, C. A., Lombardo, F., Dominy, B. W., and Feeney, P. J. (2001) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. drug Deliv. 46*, 3–26.

(21) Gu, J., Gui, Y., Chen, L., Yuan, G., Lu, H., and Xu, X. (2013) Use of Natural Products as Chemical Library for Drug Discovery and Network Pharmacology. *Peer-reviewed open accesses J. 8*, 1–10.

(22) Keller, T. H., Pichota, A., and Yin, Z. (2006) A practical view of " druggability ." *Curr. Opin. Chem. Biol. 10*, 10;357-361.

(23) Basnet-Skalko, N. (2014) Biologics : the role of delivery systems in improved therapy. *Dovepress 8*, 107–114.

(24) Brown, D. (2007) Unfinished business : target-based drug discovery. *Drug Discov. Today 12*, 1007–1012.

(25) Spellberg, B. (2008) Author ' s reply. *Lancet 8*, 45–47.

(26) Cragg, G. M., and Newman, D. J. (2014) Natural Products: A Continuing Source of Novel Drug Leads. *Biochim. biophys acta J. 1830*, 3670–3695.

(27) Harvey, A. L. (2008) Natural products in drug discovery. *Drug Discov. Today 13*, 894–901.

(28) Hong, J. (2014) Natural Product Synthesis at the Interface of Chemistry and Biology. *Chem. A Eur. J. Concept* 10204–10212.

(29) Kaburgi, and Kishi, Y. (2008) An Operationally Simple and Efficient Work-up Procedure for TBAF-mediated Desilylation; Application to Halichondrin Synthesis. *Org. Lett. 9*, 723–726.

(30) Yao, Xiao Z, Lui R, Li T, Jiao W, and Y. C. (2013) Aldehyde or a , b -Dibromoaldehyde with Isatins : An Efficient Synthesis of. *Chem. A Eur. J. 19*, 456–459.

(31) Durrant, J. D., and Mccammon, J. A. (2011) Molecular dynamics simulations and drug discovery. *BMC Biol. 9*, 71.

(32) Crane, E. A., and Gademann, K. (2016) Capturing Biological Activity in Natural Product Fragments by Chemical Synthesis. *Angew. Chemie Int. Ed. 55*, 2–23.

(33) Pathania, S., Ramakrishnan, S. M., and Bagler, G. (2015) Original article Phytochemica : a platform to explore phytochemicals of medicinal plants. *Database Sci. Journals 2015*, 1–8.

(34) Stonik, V. A., and Fedorov, S. N. (2014) Marine Low Molecular Weight Natural Products as Potential Cancer Preventive Compounds. *J. Mar. Drugs 12*, 636–671.

(35) Bernhoft, A. (2010) Bioactive compounds in plants – benefits and risks for man and animals, in *The Norwegian Academy of Science and Letters*, pp 13–14.

(36) Nagle, G. D., and Zhou, Y.-D. (2010) Natural Product-Derived Small Molecule Activators of Hypoxia-inducible Factor (HIF-1). *Curr. Pharm. Des. 12*, 2673–2688.

(37) Harvey, L. A., Edrada-Ebel, R., and Quinn, J. R. (2011) The re-emergence of natural products for drug discovery in the genomics era. *Biotechnol. Adv.* 1–52.

(38) Prachayasittikul, V., Worachartcheewan, A., and Shoombuatong, W. (2015) Computer-Aided Drug Design of Bioactive Natural Products Computer-Aided Drug Design of Bioactive Natural Products. *Curr. Top. Med. Chem. 15*, 1780–1800.

(39) Chang, C. A., Chen, W., and Gilson, M. K. (2007) Ligand configurational entropy and protein binding. *J. Proceeding Natl. Acad. Scienes 104*, 1534–1539.

(40) Dabydeen, D. A., Burnett, J. C., Bai, R., Verdier-pinard, P., Hickford, S. J. H., Pettit, G. R., Blunt, J. W., Munro, M. H. G., Gussio, R., and Hamel, E. (2006) Comparison of the Activities of the Truncated Halichondrin B Analog NSC 707389 (E7389) with Those of the Parent Compound and a Proposed Binding Site on Tubulin. *Mol. Pharmacol. 70*, 1866–1875.

(41) Smith, J. A., Wilson, L., Azarenko, O., Zhu, X., Lewis, B. M., Littlefield, A., and Jordan, M. A. (2011) Eribulin Binds at Microtubule Ends to a Single Site on Tubulin to Suppress Dynamic Instability.

Biochemistry 49, 1331–1337.

(42) Srivastava, V., Negi, A. S., Kumar, J. K., Gupta, M. M., and Khanuja, S. P. S. (2005) Plant-based anticancer molecules : A chemical and biological profile of some important leads. *Biorgan. Med. Chem. 13*, 5892–5908.

(43) Xu, D., Marquis, K., Pei, J., Fu, S., Cag, T., Chook, Y. M., and Grishin, N. V. (2015) Sequence analysis LocNES : a computational tool for locating classical NESs in CRM1 cargo proteins. *Oxford Univ. Press 31*, 1357–1365.

(44) Nguyen, K. T., Holloway, M. P., and Altura, R. A. (2012) The CRM1 nuclear export protein in normal development and disease. *Int. J. Biochem. 3*, 137–151.

(45) Zhao, C. L., Mahboobi, S. H., Moussavi-baygi, R., and Mofrad, M. R. K. (2014) The Interaction of CRM1 and the Nuclear Pore Protein Tpr. *J. Peer-reviewed open access 9*, 1–13.

(46) Mahboobi, S. H., Javanpour, A. A., and Mofrad, M. R. K. (2015) The Interaction of RNA Helicase DDX3 with HIV-1 Rev-CRM1-RanGTP Complex during the HIV Replication Cycle. *J. Peer-reviewed open access 10*, 1–27.

(47) Turner, G. J., Dawson, J., Cubitt, C., Baz, R., and Sullivan, M. D. (2015) Inhibition of CRM1-dependent nuclear export sensitizes malignant cells to cytotoxic and targeted agents. *Semin. Cancer Biol. 0*, 62–73.

(48) Sun, Q., Carrasco, Y. P., Hu, Y., Guo, X., Mirzaei, H., and Macmillan, J. (2013) Nuclear export inhibition through covalent conjugation and hydrolysis of Leptomycin B by CRM1. *Proceeding Natl. Acad. Sci. 110*, 1303–1308.

(49) Nobuaki, K., Nobuaki, M., Hiroshi, T., Daisuke, F., Erwin, P. S., Wolff, B., Yoshida, M., and Horinouchi, S. (1999) Leptomycin B inactivates CRM1 ¯exportin 1 by covalent modification at a cysteine residue in the. *Proceeding Natl. Acad. Sci. 96*, 9112–9117.

(50) Guttler, T., and Gorlich, D. (2011) Ran-dependent nuclear export mediators : a structural perspective. *Eur. Mol. Biol. Orgnanization. 30*, 3457–3474.

(51) Cousins, Kimberley R. (California State University, S. B. (2005) ChemDraw Ultra 9.0. CambridgeSoft, 100 CambridgePark Drive, Cambridge, MA 02140. *J. Am. Chem. Soc. 127*, 4115–4116.

(52) Lee, E. H., Hsin, J., Sotomayor, M., Comellas, G., and Schulten, K. (2010) Discovery Through the Computational Microscope. *NIH Public Access 17*, 1295–1306.

(53) Moonsamy, S., Dash, R. C., and Soliman, M. E. S. (2014) Integrated Computational Tools for Identification of CCR5 Antagonists as Potential HIV-1 Entry Inhibitors: Homology Modeling, Virtual Screening, Molecular Dynamics Simulations and 3D QSAR Analysis. *J. Mol. Sci. 19*, 5243–5265.

(54) Pearlman, D. A., Case, D. A., Caldwell, J. W., Ross, W. S., Cheatham, T. E., Debolt, S., Ferguson, D., Seibel, G., and Kollman, P. (1995) AMBER , a package of computer programs for applying molecular mechanics , normal mode analysis , molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. *Comput. Phys. Commun. 91*, 1–41.

(55) Freddie, S. J. R. (2011) Molecular Dynamics Simulations of Protein Dynamics and their relevance to drug discovery. *NIH Public Access 10*, 738–744.

(56) 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. *J. Comput. Chem. 25*, 1605–1612.

(57) c4d1df0ec401dea9aa42e3a511a1716cfa52fd20 @ www.clcbio.com.

(58) M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Had. (2009) gv5citation @ www.gaussian.com.

(59) Hanwell, M. D., Curtis, D. E., Lonie, D. C., Vandermeersch, T., Zurek, E., and Hutchison, G. R. (2012) Avogadro : an advanced semantic chemical editor , visualization , and analysis platform 1–17.

(60) Chetty, S., and Soliman, M. E. S. (2015) CHEMISTRY Possible allosteric binding site on Gyrase B , a key target for novel anti-TB drugs : homology modelling and binding site identification using molecular dynamics simulation and binding free energy calculations. *Med. Chem. Res. 24*, 2055–2074.

(61) Arodola, O. A., and Soliman, M. E. S. (2016) Molecular Dynamics Simulations of Ligand-Induced Flap Conformational Changes in Cathepsin-D — A Comparative Study. *J. Cel. Biochem. 15*, 1–15.

(62) Trott Oleg, O. J. A. (2011) AutoDock Vina: improving the speed and accuracy of docking with a new scoring function,efficint optimization and multithreading. *NIH Public Access 31*, 455–461.

(63) Sanner, M. F., and Jolla, L. (2000) Python: a programming language for software integration and development. *Scripps Res. Inst.*

(64) Huey, R., Morris, G. M., Olson, A. J., and Goodsell, D. S. (2007) Software News and Update A Semiempirical Free Energy Force Field with Charge-Based Desolvation. *J. Comput. Chem. 28*, 1145–1152.

(65) Morris, M. G., Goodsell, S. D., Halliday, S. R., and Huey, R. (2009) Automated Docking Using a Lamarckian Genetic Algorithm and an Empirical Binding Free Energy Function. *J. Comput. Chem. 14*, 1639–1662.

(66) Morris, G. M., Huey, R., and Lindstrom, W. (2009) Autodock :Raccoon. California.

(67) Berinyuy, E., and Soliman, M. E. S. (2016) Identification of Novel Potential gp120 of HIV-1 Antagonist Using Per-Residue Energy Contribution-Based Pharmacophore modelling. *Interdiscip. Sci. Comput. Life Sci. 10*, 174–179.

(68) Wallace A C, L. R. A. & T. J. M. (1995) LIGPLOT: A program to generate schematic diagrams of protein-ligand interactions. *protein Eng. Des. Sel. 8*, 127–134.

(69) Go, A. W., Williamson, M. J., Xu, D., Poole, D., Grand, S. Le, and Walker, R. C. (2012) Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs . 1 . Generalized Born. *J. Chem. Theory Comput. 8*, 1542–1555.

(70) Case, D. A., Iii, T. E. C., Darden, T. O. M., Gohlke, H., Jr, K. M. M., Onufriev, A., Simmerling, C., and Woods, R. J. (2007) NIH Public Access *26*, 1668–1688.

(71) Lindorff-larsen, K., Piana, S., Palmo, K., Maragakis, P., Klepeis, J. L., Dror, R. O., and Shaw, D. E. (2010) Improved side-chain torsion potentials for the. *J. Int. Sci.* 1950–1958.

(72) Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A., and Case, D. A. (2004) Development and Testing of a General Amber Force Field. *J. Comput. Chem. 25*, 1157–1174.

(73) Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., Klein, M. L., Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (2001) Comparison of simple potential functions for simulating liquid water Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys. 79*, 926–935.

(74) Moonsamy, S., Bhakat, S., Walker, R. C., and Soliman, M. E. S. (2016) Single Active Site Mutation Causes Serious Resistance of HIV Reverse Transcriptase to Lamivudine : Insight from Multiple Molecular Dynamics Simulations. *Cell Biochem. Biophys. 74*, 35–48.

(75) Berendsen, H. J. C., Postma, J. P. M., Gunsteren, W. F. Van, Dinola, A., Haak, J. R., Berendsen, H. J. C., Postma, J. P. M., Gunsteren, W. F. Van, Dinola, A., and Haak, J. R. (2012) Molecular dynamics with coupling to an external bath Molecular dynamics with coupling to an external bath. *J. Chem. Phys. 3684*, 926–935.

(76) Ryckaert, J.-P., Cioccti, G., and Berendensen, J. H. (1977) Numerical integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkanes. *J. Comput. Phys. 23*, 327–431.

(77) Ramu Anandakrishnan, B. A. and A. V. O. (2012) automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulation", Nucleic Acids Res., http://biophysics.cs.vt.edu/H++.

(78) index @ biophysics.cs.vt.edu.

(79) Mao, Y. (2011) Dynamical Basis for Drug Resistance of HIV-1 Protease. *BMC Struct. Biol. 11*, 1–9.

(80) Kumar, A., and Purohit, R. (2014) Use of Long Term Molecular Dynamics Simulation in Predicting Cancer Associated SNPs. *Peer-reviewed open access J. Comput. Chem. 10*, 1–14.

(81) Burton, B., Zimmermann, M. T., Jernigan, R. L., and Wang, Y. (2012) A Computational Investigation on the Connection between Dynamics Properties of Ribosomal Proteins and Ribosome Assembly. *Peer-reviewed open access J. Comput. Chem. 8*, 1–13.

(82) Laura A. Deschenes and David A. Vanden BoutUniversity of Texas, A. (2000) Origin 6.0: Scientific Data Analysis and Graphing Software Origin Lab Corporation (formerly Microcal Software, Inc.). *J. Am. Chem. Soc. 122*, 9567–9568.

(83) Lawrenz, M., Baron, R., and Mccammon, J. A. (2009) Free-Energy Changes for Biomolecular Systems : Determinants of H5N1 Avian Influenza Virus Neuraminidase Inhibition by Peramivir. *J. Chem. Theory Comput. 5*, 1106–1116.

(84) Lawrenz, M., Wereszczynski, J., Amaro, R., Walker, R., Roitberg, A., and Mccammon, J. A. (2010) Impact of calcium on N1 influenza neuraminidase dynamics and binding free energy. *J. Proteins 5*, 2523–2532.

(85) Eldridge, M. D., Murray, C. W., Auton, T. R., Paolini, G. V, and Mee, R. P. (1997) Empirical scoring functions : I . The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *J. Comput. Aided. Mol. Des. 11*, 425–445.

(86) 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. *J. Chem. Inf. Model. 51*, 69–82.

(87) Sun, H., Hou, T., Wang, J., and Xu, L. (2013) Assessing the performance of MM/PBSA and MM/GBSA methods. 3. The impact of force fields and ligand charge models. *J. Phys. Chem.*

(88) Gohlke, H., Kiel, C., and Case, D. A. (2003) Insights into Protein – Protein Binding by Binding Free Energy Calculation and Free Energy Decomposition for the Ras – Raf and Ras – RalGDS Complexes. *J. Mol. Biol. 2836*, 891–913.

(89) Hou, Tingjun, Li Nan, W. W. (2013) Characterization of Domain–Peptide Interaction Interface: Prediction of SH3 Domain-Mediated Protein–Protein Interaction Network in Yeast by Generic Structure-Based Models. *NIH Public Access 11*, 2982–2995.

(90) Pramely, R., and Raj, T. L. S. (2012) Prediction of biological activity spectra of a few phytoconstituents of Azadirachta indicia A . Juss. *J. Biochem. Tech. 3*, 375–379.

(91) Kumar, R., and Damanpreet, G. (2011) CHEMISTRY PASS-assisted exploration of new therapeutic potential of natural products. *Acc. Chem. Res. 20*, 1509–1514.

(92) Lagunin, A., Zakharov, A., Filimonov, D., and Poroikov, V. (2011) QSAR Modelling of Rat Acute Toxicity on the Basis of PASS Prediction. *J. Mol. Informatics 30*, 241–250.

(93) Lim, S. V., Rahman, M. B. A., and Tejo, B. A. (2011) Structure-based and ligand-based virtual screening of novel methyltransferase inhibitors of the dengue virus. *BMC Bioinformatics 12*, 2–12.

(94) Filimonov, D. A., Lagunin, A. A., Gloriozova, T. A., Rudik, A. V, Druzhilovskii, D. S., Pogodin, P. V, and Poroikov, V. V. (2014) Prediction of the biological activity spectra of organic compounds uing the PASSonline website resource. *Chem. Heterocycl. Compd. 50*, 444–457.

(95) Genheden, S., and Ryde, U. (2016) Expert Opinion on Drug Discovery The MM / PBSA and MM / GBSA methods to estimate ligand-binding affinities The MM / PBSA and MM / GBSA methods to estimate ligand-binding affinities. *Expert Opin. Drug Discov. 441*, 449–461.

(96) Desheng, L., Jian, G., Yuanhua, C., Wei, C., Huai, Z., and Mingjuan, J. (2011) Bioorganic & Medicinal Chemistry Letters Molecular dynamics simulations and MM / GBSA methods to investigate binding mechanisms of aminomethylpyrimidine inhibitors with DPP-IV. *Bioorg. Med. Chem. Lett. 21*, 6630–6635.

(97) Srivastava, M., Singh, H., and Naik, K. P. (2010) Molecular modeling evaluation of the antimalarial activity of artemisinin analogues : molecular docking and rescoring using Prime / MM-GBSA Approach. *Curr. Res. J. Biol. Sci.* 83–102.

(98) Langunin, A., Stepanchikova, A., Filimonov, D., and Poroikov, V. (2000) Prediction of activity spectra for biologically active substances. *Bioinforma. Appl. NOTE 16*, 747–748.

(99) Kohli, G. S., John, U., Figueroa, R. I., Rhodes, L. L., Harwood, D. T., Groth, M., Bolch, C. J. S., and Murray, S. A. (2015) Polyketide synthesis genes associated with toxin productionin two species of Gamierdiscus (Dinophyceae). *BMC Genomics* 1–10

CHAPTER 6

6.0. General conclusion and future study recommendations

6.1 General conclusions

The major aims of this study were to validate the *"reduce to maximum concept"* utilising computational modeling approaches in a bid to contrive new alternative protocols that can be followed in next generation drug discovery. To generate a protocol that can be utilised to distinguish which fragments of a chemical compound must be retained for biological activity and which fragments can be omitted based on thermodynamic integrated calculations. In addition to providing significant molecular insight into the binding affinities and molecular mechanism interaction of the parent compound Anguinomycin D and derived analog SB 640. Estimating the toxicity profiles of parent compound Anguinomycin D in comparison to its structural analog SB 640. Another pivotal aim of this study was to highlight the use computational tools such as homology modeling in the drug design and discovery process. Results from this work have led to the following conclusions:

1. Although being reduced by more than 60 % in its structural composition, the derived analog of Anguinomycin D, SB 640 displayed only a minimal reduction in the binding affinity in comparison to the parent compound which corresponds with experimental studies conducted. The retention of biological activity exhibited by SB 640 may be attributed to the retention of the crucial α, β unsaturated δ-lactone moiety.
2. The convergence of post-dynamic simulation of Anguinomycin D-CRM1 and analog SB 640-CRM1 systems were validated by RMSF and Rg potential energy plots. The RMSF of both the parent compound and analog SB 640 in complex with the NES binding groove of the CRM1 protein displayed a relatively rigid and stable systems. With the parent compound Anguinomycin D-CRM1 complex displaying a greater amino acid fluctuation in comparison to the analog SB 640.
3. From the Rg potential energy plot, the analog SB-640-CRM1 complex displayed an overall stable system as compared to Anguinomycin D-CRM1 system which exhibited prominent flexibility. The high level of stability exhibited by the analog SB 640 can be attributed to its compact size and ability to position itself deeper, as displayed by Figure 5.7 within the active site.
4. From the thermodynamics calculations performed it can observed, that the electrostatic interaction exhibited by the analog SB 640 supersedes that of the parent compound, the observed increased in the electrostatic interactions can be correlated to the conformational pose

undertaken by the analog SB 640 within the hydrophobic NES binding groove of the CRM1 protein. However the parent compound Anguinomycin D displayed a higher van der Waal contribution to the binding of the NES binding groove of the CRM1, which can be attributed to the presence of the polyketide chain which is omitted in the structure of the analog SB 640.

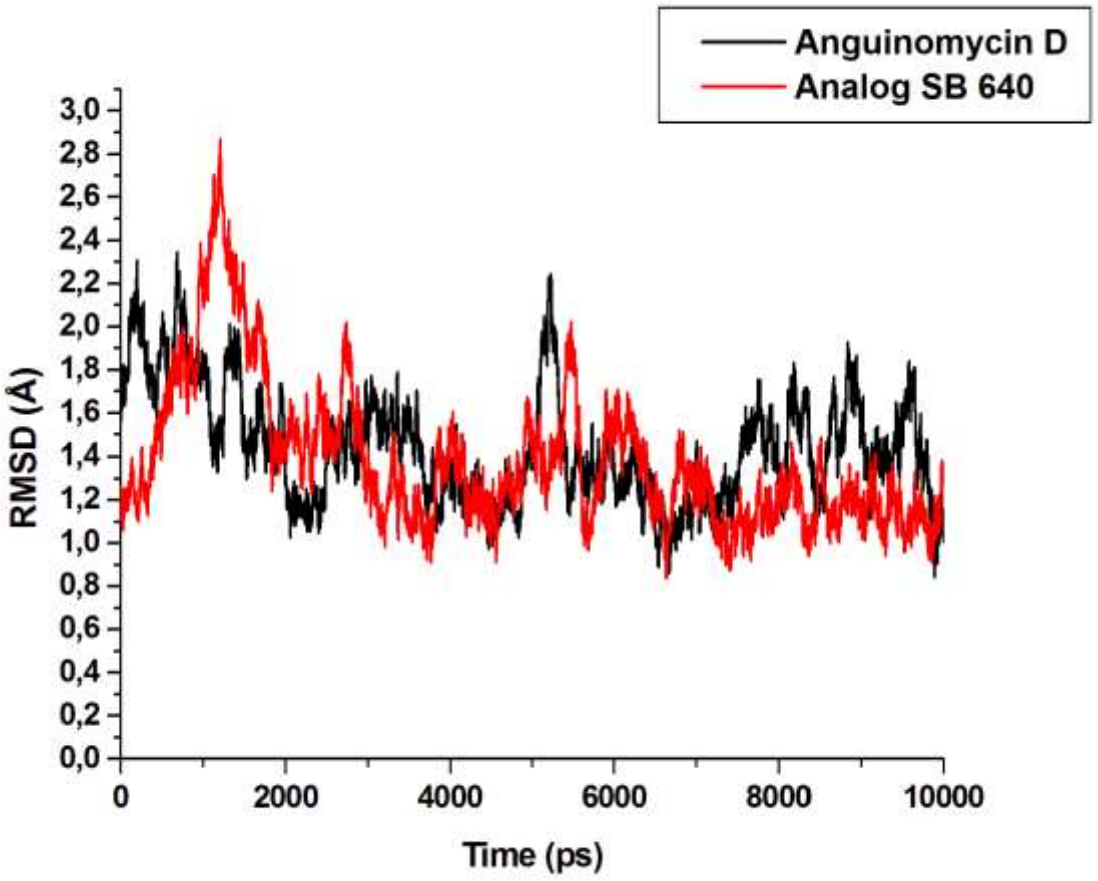
5. Per-residue energy decomposition analysis displayed energetically favourable contributions from amino acids Leu 536, Thr 575, Val 576 and Lys 579, these prominent amino acid residues were conserved in the analog SB 640-CRM1 and Anguinomycin D-CRM1 complexes. Elucidating which fragments of a chemical compounds is essential for retention or enhanced biological activity and which fragments can be omitted.
6. Toxicity profiling of the parent compound Anguinomycin D displayed a greater level of toxicity of the parent compound in comparison to the analog SB 640, which can be allocated to the presence of the polyketide chain present in the parent compound. Highlighting that the reduction in the size of a compound may result in reduced toxicity exhibited by the analog SB 640.
7. The *reduce to maximum concept* demonstrates the potent activity of compact structural analogues, which was further validated in this study, the execution of this concept may enable better synthetic approachability of chemical fragments derived from natural products.
8. Homology modeling was identified as a prominent tools in drug discovery, identifying its various applications as well as providing a conclusive protocol that can be followed to generate the most accurate yet optimal 3D protein structure.

6.2 Future study recommendations

Strategic computational techniques presented in this work will serve as beneficial tools to enhance novel drug discovery and development process. This study displays the potential use for the "*reduce to maximum concept* "to be implemented in all spheres of drug design and discovery. As the data extracted from this study displays the retention of biological activity, reduced toxicity and enhanced synthetic approachability of the derived analog although being reduced by more than 60% in its structural composition form that of the parent compound Anguinomycin D. This study displays the use of computational approaches such as thermodynamic calculations that can be utilised to isolate and reproduce only the most essential chemical fragments derived from natural products that can be used in drug discovery. As well as highlights which fragments can omitted as their presence may result in the toxic effect displayed by many pharmacological drugs and so the implementation of this study may prevent the synthesis of unwanted and toxic component of chemical fragments, this study in combination with chemical synthesis and prospective biological testing of lead compounds identified may alleviate the pending crisis of lack of new prospective drug candidates required for rational drug. The implementation of homology modeling in drug design and discovery protocols to generate the most accurate model, increasing the specificity of potential drug candidates ultimately leading to the effective

treatment of disease conditions. Further validation of the data extracted from this study may provide insight into the direct molecular interactions with regard to the distribution of motion, providing further into the exact binding mode adopted by the analog in comparison to the parent compound. This can be accomplished by performing Principal Component Analysis (PCA), Residue Interaction Network (RIN), and Substrate Envelope Analysis (SEA). Application of these methods may provide vital insight, elaborating the enzyme dynamics, drug-enzyme interactions and conformational structural changes.

APPENDIX



A1. RMSD of the C-alpha backbone of CRM1 in complex with Anguinomycin D (black) and analog SB 640 (red).