

**STRUCTURAL AND BIOPHYSICAL CHARACTERIZATION OF CASSAVA
LINAMARASE AND THE ROLE OF SOLVENTS ON LINAMARIN'S PROPERTIES: A
COMPUTATIONAL STUDY**

Lucas Paul

**A Thesis Submitted in Fulfillment of the Requirements for the Degree of Doctor of
Philosophy in Materials Science and Engineering of the Nelson Mandela African
Institution of Science and Technology**

Arusha, Tanzania

June, 2022

ABSTRACT

Linamarase and linamarin mainly from cassava have many applications ranging from food, environmental to the medical industry. To better explore the potential of this enzyme and its substrate, one needs to understand its interaction mechanism at the molecular and atomistic level. In this thesis, the three-dimensional (3D) structure of linamarase was built via homology modeling. The developed model was used to determine the binding orientation and mechanism of linamarin to the enzyme using molecular docking. Molecular dynamics simulation was used to determine the stability of the built model and when complexed with the ligand. It was interesting to note that complex 1 with the low binding-free energy of -6.9 kcal/mol showed a larger Root Mean Square Deviation (RMSD) value with two maxima at 0.255 and 0.310 nm compared to complex 2 with the best binding-free energy of -7.2 kcal/mol, whose RMSD value shows the maxima at 0.19 nm. The end-point free energy method based on Molecular Mechanics Poisson Boltzmann Surface Area (MM/PBSA) was used to rescored binding free energy obtained from docking calculations. The ensemble structure was observed to be relatively stable compared to the modelled structure. Furthermore, the stability and conformational orientation preferences of linamarin in different solvents was established using classical molecular dynamics, and found to be solvent dependent. The effects of solvents on the stability and conformational preference is pronounced by different probability density maxima of the measured reaction coordinate/properties. Linamarin is observed to be stable in methanol followed by dimethyl sulfoxide (DMSO) and least stable in water. Solvent polarity was observed to influence the stability and conformation preference of the title compound. Linamarin exists in trans and gauche conformations, the former was observed to be more stable in water than other solvents and the latter in DMSO. The measured reaction coordinates, distance and dihedral angles ascertained that the conformational preference is due to rotation at $\phi = \pm 180^\circ$ and $\phi = \pm 50^\circ$. Finally, the stability of linamarin was also attributed to different numbers of inter and intra hydrogen bonds formed in different solvents. Results presented in this thesis provides atomistic insights on the role of solvents polarity on linamarase–linamarin complex interaction and stability. The findings provide important information on the application of linamarase and linamarin in different fields including food processing and in drugs.

DECLARATION

I, Lucas Paul, do hereby declare to the Senate of the Nelson Mandela African Institution of Science and Technology that this dissertation is my own original work and that it has neither been submitted nor concurrently submitted for a degree or similar award in any other institution.

Lucas Paul

Name and Signature of the Candidate

Date

The above declaration is confirmed

Prof. Revocatus Lazaro Machunda

Name and Signature of Supervisor (1)

Date

Prof. Kelvin Mark Mtei

Name and Signature of Supervisor (2)

Date

Prof. Fidele Ntie-Kang



Name and Signature of Supervisor (3)

Date

Prof. Celestin Nzanu Mudogo

CmudogoNz.

Name and Signature of Supervisor (4)

Date

COPYRIGHT

This thesis is copyright material protected under the Berne Convention, the Copyright Act of 1999 and other international and national enactments, in that behalf, on intellectual property. It must not be reproduced by any means, in full or in part, except for short extracts in fair dealing; for researcher private study, critical scholarly review or discourse with an acknowledgement, without the written permission of the office of Deputy Vice Chancellor for Academic, Research and Innovation on behalf of both the author and NM–AIST.

CERTIFICATION

The undersigned certify that they have read and hereby recommend for acceptance by the Nelson Mandela African Institution of Science and Technology a thesis titled “*Structural and Biophysical Characterization of Cassava Linamarase and the Role of Solvents on Linamarin’s properties: A Computational Study*” in fulfilment of the requirements for the degree of Doctor of Philosophy in Materials Science and Engineering of the Nelson Mandela African Institution of Science and

Prof. Revocatus Lazaro Machunda

Name and Signature of Supervisor (1)

Date

Prof. Kelvin Mark Mtei

Name and Signature of Supervisor (2)

Date

Prof. Fidele Ntie-Kang



Name and Signature of Supervisor (3)

Date

Prof. Celestin Nzanu Mudogo

CmudogoNz.

Name and Signature of Supervisor (4)

Date

ACKNOWLEDGEMENTS

First of all, I would like to thank the Almighty God for many blessings of healthy life. I would like to express my deepest gratitude to my supervisors: Prof. Revocatus Lazaro Machunda, Prof. Kelvin Mark Mtei, Prof. Fidele Ntie-Kang and Prof. Celestin Nzanzu Mudogo. I thank each of them for giving me a great space, trusting and accepting me as their student. They have been an inspiration to me by setting an example of sincerity, activeness and commitment. I extend my great appreciation to Dr. Daniel Madulu Shadrack who encouraged me as a young PhD student and quite literally introduced me to the world of molecular modelling; he has shown me what it means to be an experienced, flexible, energetic, diligent and self-motivated scientist. Dr. Daniel Madulu Shadrack helped access the CHPC lengau, South Africa, for the computational facility. A special thank goes to Prof. Andrew Paluch from Miami University for his assistance of learning a lot of computational skills and how to use Ohio Supercomputer (OSC) his experience was so valuable. I would like to express my deepest gratitude to Prof. Titus Msagati for providing Lab space at University of South Africa (UNISA). It was a good time and a new experience in learning on LC-MS/MS machines. More thanks should extend to Dr. Hlengilizwe H. Nyoni for the patient teaching me on operating the LC-MS/MS machine.

I am genuinely grateful to Prof. Revocatus Lazaro Machunda (Dean, School of Materials, Energy, Water and Environmental Sciences [MEWES]) and Dr. Thomas Kivevele for providing a supportive environment to carry out my studies. I extend my heartfelt thanks to the Development Bank (AfDB) Project Management at NM-AIST, Mr. Julius Lenguyana and Ms. Victoria Ndossi for providing an excellent atmosphere to carry out my research activities. Furthermore, I wish to thank Ms. Analyce Ichwekeleza (Dean of Students), Ms. Victoria Ndossi, and Ms. Silvia Ndakidemi (Students Welfare Office), Mr. Adam Mawenya, Ms. Leah Gonda and Ms. Joyce Martin for their dedicated assistance from IT department. I gain my record and my sincere gratitude to Mr. Haji Chomba, Ms. Christina Dominick, Ms. Florence George and Ms. Eunice Mwakang'ata for their exceptional work in MEWES Dean's Office; I also owe the entire NM-AIST community.

Without neglecting the importance of personal support, I also would like to thank all friends at MEWES and COCSE Lab for their friendliness and enthusiasm, in particular Mr. Denis Mteremko and Mr. Naseriani Kambaine (University of Dodoma [UDOM]) for their time we have been discussing and sharing new concepts from molecular dynamics and computational chemistry in general. I extend my sincere appreciation to colleagues Dr. Geradius Deogratius, Dr. Eric Mutegoa, Mr. Rene Costa, Mr. Michael Samweli Mollel, Mr. Macton Wasiwasi Mgonzo, Mr. Aristide G. Lambura, Dr. Catherine Paschal, Mr. Clemence Ansbert, Dr. Edwin R. Ndibalema,

Mr. Lucas Theodori, Ms. Faith Mpondo, Ms. Fina Lesafi, Ms. Geni Juma, Ms. Godiana H. Philipo, Mr. Hegespo H. Mwanyika, Mr. Plassidius J. Chengula, Ms. Siri Abihudi, Dr. Stephano Hanolo, Ms. Tusekile Alfred, Mr. Wilson Mahene and many others. I wish to thank all the people whose assistance was such a long journey of three years with a lot of sacrifice.

I wish to acknowledge the support and great love of my family, especially my beloved wife, Ms. Agnes Sadock Bitungwa. Thank you for taking care of our beautiful family throughout this long commitment of your time, you are such a special person and I am so fortunate to call you my wife. My sincere appreciation goes to the family of my father: Mr. Mzee Paul Maziku Luchemba and my mother Ms. Gaudensia Stanslaus Mlindi; my sisters Ms. Thersia Paul maziku, Ms. Rachel Paul maziku, Ms. Govina Paul Maziku and my brothers: Mr. Stanslaus Paul Maziku, Mr. Godfrey Paul Maziku, Mr. David Paul Maziku, Mr. Nestory Paul Maziku and Mr. Fredirick Paul Maziku and the entire of Luchemba family.

I would like to appreciate colleagues and my previous co-workers at Monduli Teachers College for their lovely support. We spent nine years working together, it will remain a memorable experience throughout my life.

Lastly, I would like to express my sincere appreciation to my employer, Dar es Salaam University College of Education, for providing me a study leave. I gratefully acknowledge the AfDB's financial support, the United Republic of Tanzania through project number P-Z1-IA0-016 and grant number 2100155032816.

DEDICATION

To my wonderful daughter Uwera Lucas Luchemba, my loving grandparents, the late Mr. Babu Stanslaus Mlindi and Late Mr. Masanja Luchemba, also Ms. Bibi Mary Sozi and Ms. Theresia Kazoya, my mentors: Dr. Samwel Mchael Irira and Dr. Jesca Moiro.

TABLE OF CONTENTS

ABSTRACT.....	i
DECLARATION	ii
COPYRIGHT.....	iii
CERTIFICATION	iv
ACKNOWLEDGEMENTS	iv
DEDICATION	vii
TABLE OF CONTENTS.....	viii
LIST OF TABLES	xii
LIST OF FIGURES	xiii
LIST OF ABBREVIATIONS AND SYMBOLS	xvi
CHAPTER ONE	1
INTRODUCTION	1
1.1 Background of the Problem	1
1.1.1 Brief Introduction of Cassava as a Leading Source of Linamarase a Hydrolyzing Agent of Cyanogenesis.....	1
1.1.2 Cyanogenic Glycosides in Cassava.....	1
1.1.3 Nutrient Content of Cassava	2
1.1.4 Cyanide Toxification and Health Effects	3
1.1.5 Cassava Processing	4
1.1.6 Catalytic Properties, Functional Attributes and Industrial Applications of β – glucosidases.....	6
1.1.7 Linamarin as a Glycoside and its Pharmacological Potential	8
1.2 Statement of the Problem	10
1.3 Rationale of the Study	11
1.4 Research Objectives	11
1.4.1 General Objective.....	11

1.4.2	Specific Objectives.....	11
1.5	Research Questions	11
1.6	Significance of the Study	12
1.7	Delineation of the Study.....	12
CHAPTER TWO		13
LITERATURE REVIEW		13
2.1	Experimental Determination of Protein	13
2.2	Structure Determination by Homology Modeling	13
2.2.1	Identification of Templates	14
2.2.2	Sequence Alignment	14
2.2.3	Model Building	15
2.2.4	Loop Modeling.....	15
2.2.5	Sidechain Modelling	16
2.2.6	Model Optimization	16
2.2.7	Model Validation.....	16
2.2.8	Application of Homology Modelling in Different Aspects	17
2.2.9	Challenges of Homology Modeling	17
2.3	Molecular Docking.....	18
2.3.1	Docking Algorithm	19
2.3.2	Scoring Function	20
2.3.3	Empirical Scoring Function	21
2.3.4	Knowledge-based Scoring Function.....	21
2.4	Molecular Dynamics	22
2.4.1	Energy Minimization.....	23
2.4.2	Force Fields	23
2.4.3	Integrating the Equation of Motion	24

2.4.4	Molecular Mechanics Poisson–Boltzmann (Generalized Born) Surface Area (MM–PB(GB)SA)	25
2.4.5	Limitation of Molecular Mechanics Poisson –Boltzmann Surface Area	27
2.5	Some Successful Examples and Applications of Computational Approaches in Structural Determination as well as Protein–ligand Interaction	27
2.5.1	Application of Homology Modelling and Molecular Docking	27
2.5.2	Applications of Molecular Dynamics	29
2.5.3	Relaxed Complex Scheme	31
2.5.4	Application of Free Energies Methods.....	32
CHAPTER THREE		34
MATERIALS AND METHODS.....		34
3.1	Homology Modeling of Linamarase	34
3.2	Molecular Docking.....	35
3.2.1	Protein (Receptor) Preparation.....	35
3.2.2	Ligand Preparation (Linamarin).....	35
3.2.3	Docking Bench Making and Method Validation	35
3.3	Molecular Dynamic (MD) Simulation	36
3.4	Determination of the Binding Free Energy by Molecular Mechanics Poisson –Boltzmann Surface Area Analysis.....	36
3.5	Linamarin Conformation in a Vacuum and Different Solvents	37
CHAPTER FOUR.....		38
RESULTS AND DISCUSSION		38
4.1	Homology Modeling of Linamarase	38
4.2	Molecular Docking Benchmark and Validation	46
4.3	Molecular Dynamics Simulation	53
4.4	Molecular Mechanics Poisson–Boltzmann Surface Area Binding–free Energy Analysis	55
4.5	Stability and Conformational Preference of Linamarin in Different Solvents	57
4.5.1	Pair Distribution Function of Polar Groups Around Linamarin and Different Solvents ..	57

4.5.2	Distributions of Dihedral Angles	61
4.5.3	Distance and Angles Distribution	63
4.5.4	Hydrogen Bonds Analysis	65
CHAPTER FIVE		69
CONCLUSION AND RECOMMENDATIONS		69
5.1	Conclusion	69
5.2	Recommendations	70
REFERENCES:		72
RESEARCH OUTPUTS.....		99

LIST OF TABLES

Table 1:	Summarizing cassava processing techniques and their efficiency.....	5
Table 2:	Ramachandran plot statistics analysis of modelled linamarase from SWISS-MODEL, MODELLER and I-TASSER.....	43
Table 3:	Summary of main chain parameters for the structure generated by SWISS-MODEL	43
Table 4:	Summary of main chain parameters for the structure generated by MODELLER-MODEL	43
Table 5:	Summary of main chain parameters for the structure generated by I-TASSER-MODEL	44
Table 6:	Summary of side-chain parameters for the structure generated by SWISS-MODEL	44
Table 7:	Summary of side chain parameters for the structure generated by MODELLER-MODEL	44
Table 8:	Summary of side chain parameters for the structure generated by I-TASSER-MODEL	45
Table 9:	Summary of Experimental and calculated binding energies (kcal mol ⁻¹)	46
Table 10:	The amino acid residues at the active site of modelled, holo and ensemble structure	48
Table 11:	The binding-free energy (kJ/mol) for the two complexes calculated by the MM/PBSA method.....	56
Table 12:	Summarizes the arrangement of the pair distribution functions and the first coordination numbers.....	61
Table 13:	Naming of the dihedral angles of linamarin molecules	61
Table 14:	The average number of HBs contributed by each group of linamarin in different solvents.....	67

LIST OF FIGURES

Figure 1:	Worldwide production of Cassava by 2019 ranging from blue to red representing highest to lowest producers, respectively	1
Figure 2:	A complete reaction mechanism of linamarin and linamarase during cyanogenesis ..	2
Figure 3:	Schematic representation of the hydrolysis of cellulose using a series of reaction, cellulosic fibres broken down by Exo and endoglucanase to oligosaccharides; these small molecule oligosaccharides hydrolyzed to a final product that is glucose (Dan Li <i>et al.</i> , 2013)	6
Figure 4:	A reaction mechanism involving glycosylation and deglycosylation, hydrolysis and cleavage of the glycosidic bond (Singh <i>et al.</i> , 2016)	8
Figure 5:	The flow chart of molecular docking	18
Figure 6:	Relationship between energy function and position of atoms in a system	22
Figure 7:	Schematic representation of molecular dynamics process.....	23
Figure 8:	The structure of beta-glucosidase B with protein data base code 2O9T	30
Figure 9:	Protein structure generated by SWISS-MODEL.....	38
Figure 10:	Structure validation of modelled Linamarase: (A) Local quality estimate of the residues of the predicted Linamarase model; (B) comparison of the predicted Linamarase structure with a nonredundant set of protein data base structures.....	39
Figure 11:	Multiple sequence alignment for modelled linamarase template from rice, sorghum and maize. The red highlighted represents identical; the secondary structures are presented on the top while the residues' surface accessibility to solvent is displayed at the bottom (the blue for surface accessible, cyan for partially accessible and white for buried residues)	40
Figure 12:	Structure validation using Ramachandran plot of modelled protein from SWISS-MODEL	41
Figure 13:	Structure validation using Ramachandran plot of modelled protein from MODELLER (A) and I-TASSER (B).....	42

Figure 14: Correlation between theoretical and experimental binding–free energy (kJ/mol) ($r^2 = 0.69$)	46
Figure 15: Binding free energy of different protein conformations	47
Figure 16: General interaction between linamarin and modelled structure using proteins plus (A) and Discovery Studio Visualizer (B)	49
Figure 17: Linamarin interaction by a hydrogen bond with modelled structure (A) for 3D and (B) for 2D	50
Figure 18: Hydrophobic interaction between linamarin and modelled structure 3D for (A) and 2D for (B).....	50
Figure 19: General interaction between linamarin and ensemble structure using proteins plus (A) and Discovery Studio Visualizer (B)	51
Figure 20: Linamarin interaction by a hydrogen bond with ensemble structure (A) for 3D and (B) for 2D	51
Figure 21: Hydrophobic interaction between linamarin and ensemble structure 3D for (A) and 2D for (B).....	52
Figure 22: The probability values for RMSD, R_g and SASA of the apo protein and the complexes	54
Figure 23: Root Mean Square Fluctuation difference between apo and complex 1	55
Figure 24: The energy contribution on protein–ligand binding (a) complex 1 and (b) complex 2.	57
Figure 25: The chemical structure of linamarin used in this study, atoms are colour by colour: Oxygen in red, nitrogen in blue, carbon in light blue and hydrogen in white	58
Figure 26: Pair distribution function of oxygen groups around linamarin molecule in water, MeOH and DMSO	59
Figure 27: Distribution of dihedral angles of linamarin molecule in a vacuum and different solvents (a) 1–Dihedral, (b) 2–Dihedral, (c) 3–Dihedral, (d) 4–Dihedral and (e) 5–Dihedral.....	62

Figure 28: The probability distribution of the distance of the selected atoms and end to end for linamarin in a vacuum and different solvents	64
Figure 29: The probability distributions of the number of HBs between linamarin and different solvents.....	65
Figure 30: Contribution of hydrogen bonds between linamarin and different solvents	66
Figure 31: Contribution of different groups of linamarin during hydrogen bonding and different solvents.....	67

LIST OF ABBREVIATIONS AND SYMBOLS

3D	Three-Dimensional Structure
BglG	β -Glucosidase
BLAST	Basic Local Alignment Search Tool
CADDD	Computer-Aided Drug Design and Discovery
CASTp	Computer Atlas of Surface Topography of Proteins
DMSO	Dimethylsulphoxide
DNA	Deoxyribonucleic Acid
FAO	Food and Agriculture Organization
FEP	Free Energy Perturbation
fs	Femtosecond
GA	Genetic Algorithm
GDH	Glutamate Dehydrogenase
Glc	Glucose
GMQE	Global Model Quality Estimation
HCN	Hydrogen Cyanide
HNL	Hydroxynitrile
HPLC	High Performance Liquid Chromatography
IC	Incremental Construction
K _i	Binding Constant
LIE	Linear Interaction Energies
MC	Monte Carlo
MD	Molecular dynamics
MeOH	Methanol
mg	Milligram
MM-GBSA	Molecular Mechanics Generalized Born Surface area
MM-PBSA	Molecular Mechanics Poisson-Boltzmann Surface Area
NCBI	National Center for Biotechnology Information
NEST	An integrated model- building application that takes as input a sequence alignment of a target to one or multiple template PDB files and produces a mode
NMR	Nuclear Magnetic Resonance
NPT	Constant Number of Particles, Pressure and Temperature
ns	Nanosecond

NVT	Constant Number of Particles, Volume and Temperature
PB	Poisson–Boltzmann
PBS	Phosphate–Buffered Solution
PDB	Protein Data Bank
pNPG	4–Nitrophenyl β –D–glucopyranoside
ppm	Parts per Million
ps	Picosecond
QMEAN	Quality Model Energy Analysis
RCS	Relaxed Complex Scheme
R_g	Radius of Gyration
RMSD	Root Mean Square Deviation
RMSF	Root Mean Square Fluctuation
RNA	Ribonucleic Acid
S	Entropy
SASA	Solvent Accessible Surface Area
TI	Thermodynamic Integration
WHO	World Health Organization
β	Beta
ΔG	Binding free energy
ΔG_0	Independent Binding Energy
ΔG_{hb}	Binding Energy from Hydrogen Bonds
ΔG_{lip}	Lipophilic Interaction
ΔH	Enthalpy
μM	Micro Molar

CHAPTER ONE

INTRODUCTION

1.1 Background of the Problem

1.1.1 Brief Introduction of Cassava as a Leading Source of Linamarase a Hydrolyzing Agent of Cyanogenesis

Cassava (*Manihot esculenta* Crantz) is a perennial, dicotyledonous plant belonging to the Euphorbiaceae (Zou & Yang, 2019). It has underground starchy tubers (roots) grown in subtropical, hotter and lowland tropics (Falade & Akingbala, 2010). It is an essential food source for more than 500 million people, specifically in Africa, Asia, and Latin America (Fig. 1). The map shows the worldwide cassava production by 2019; the blue to red colors represent the highest to lowest producers, respectively. Cassava is ranked as a significant food crop after rice, wheat and maize (Nambisan, 2011). It is a valuable crop due to its high yield in poor soil, drought tolerance, disease resistance and the starch-rich tuber can stay in the soil for a longer time and can be harvested only on demand. Therefore, these attributes make cassava a valuable food security crop and insurance against famine (Jørgensen *et al.*, 2011).

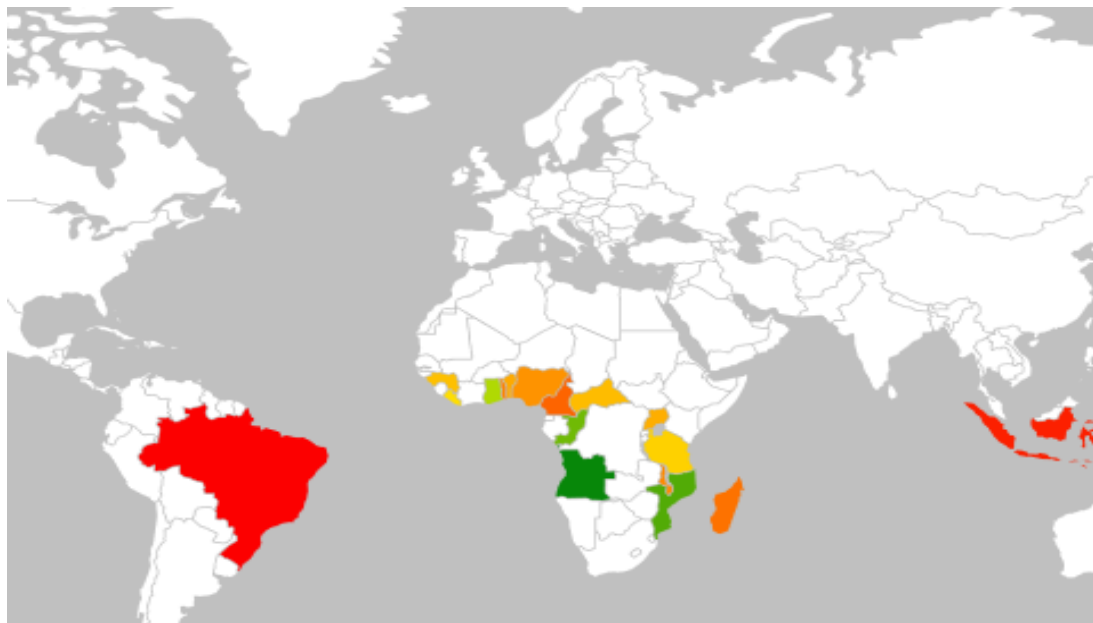


Figure 1: Worldwide production of Cassava by 2019 ranging from blue to red representing highest to lowest producers, respectively

1.1.2 Cyanogenic Glycosides in Cassava

Cassava has a high level of cyanogenic glucosides, which act as antinutrients, limiting its consumption as a staple food. Almost all cassava parts (leaves, stems, roots) contain cyanogenic

glucosides, mainly: linamarin, lotaustralin and amygdalin, linamarin being the dominant among the cyanogenic (Lykkesfeldt & Møller, 1994). Cassava leaves have been reported to contain higher cyanide contents of about 53 to 1300 cyanide equivalents/kg of dry matter (Wobeto *et al.*, 2007) than the roots, approximately 10 to 500 mg/kg of the dry matter (Siritunga & Sayre, 2003).

The process of cassava releasing toxic cyanide is a defensive mechanism against predators (called cyanogenesis). This process is initiated when predators consume cassava tissue. During the chewing process, cassava tissues rupture, the vacuole releasing cyanogenic (mainly linamarin), which is hydrolysed by cell wall enzyme linamarase (β -glucosidase) (McMahon *et al.*, 1995). Hydrolysis of linamarin releases glucose, unstable intermediate acetone cyanohydrin and hydroxy nitrile, as shown in Fig. 2.

The intermediate compounds can spontaneously decompose to the final products (acetone and Hydrogen Cyanide [HCN]). These products are either by the enzyme Hydroxynitrile (HNL) or under favourable conditions ($\text{pH} > 4$) or temperature (>30) (Cutler & Conn, 1981; Wajant *et al.*, 1994; Yemm & Poulton, 1986).

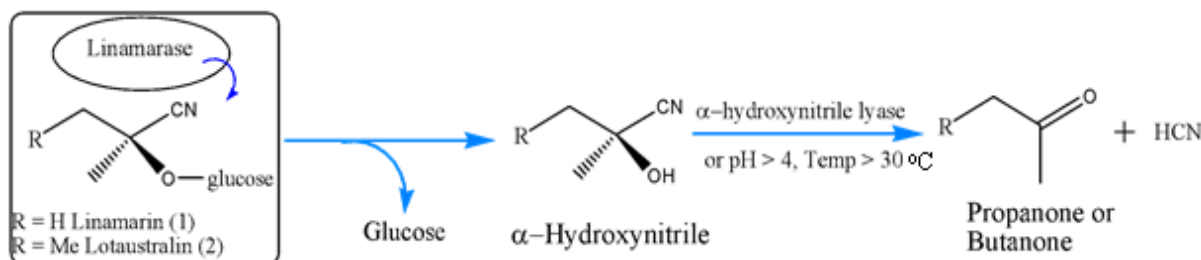


Figure 2: A complete reaction mechanism of linamarin and linamarase during cyanogenesis

1.1.3 Nutrient Content of Cassava

Cassava has two mainly edible parts; roots and leaves. The roots contain carbohydrates ranging from 80% to 90% on dry weight, and starch is about 80% (Montagnac *et al.*, 2009b). There is a small amount of maltose, sucrose, glucose and fructose (Montagnac *et al.*, 2009a; Sun *et al.*, 2014). Cassava leaves have a higher vitamin A, vitamin C, and protein content and are also a good source of dietary fibre. Most protein from cassava comprises linamarase enzymes (Bokanga *et al.*, 1994; Commission, 2000). It is reported that the protein and essential amino acid available in leaves is at a higher level even than that recommended by Food and Agriculture Organization's (FAO) (Ngudi *et al.*, 2003). Cassava leaves have a good amount of magnesium, manganese, zinc, iron, calcium and vitamin B1, B2, C, and carotenoids (Ngudi *et al.*, 2003). Combination of roots and

leaves provides a balanced meal with almost all requirements in terms of calories, vitamins, minerals and protein (Montagnac *et al.*, 2009a).

1.1.4 Cyanide Toxification and Health Effects

Cyanide in cassava is not available in a free state, so it is found in a bound state (nitrile group) to a complex compound called linamarin. Hydrolysis and decomposition of linamarin release cyanide as one of its products. The compounds, which release cyanide upon decomposition, are known as cyanogenic. Some cyanogenic decomposes only by boiling; however, those from cassava are less affected by boiling. Naturally, linamarin itself is not toxic, and if it happens, it is absorbed into the bloodstream and can probably be excreted unchanged in the urine. Either of these two ways can release cyanide from linamarin; first, cassava product with residues of linamarin is ingested, it may interact with an enzyme produced by microbes in the gut, thus liberating free cyanide (Domínguez *et al.*, 2013). Secondly, interaction by endogenous linamarase that previously compartmentalised in the cell wall while linamarin in vacuole so tissue disruption brings the interaction active (Nambisan, 1999). The World Health Organisation (WHO) set the safe level of the cyanogenic for cassava flour consumption to be 10 ppm (Djazuli & Bradbury, 1999). When cyanide gets into the body, if it is not more than 10 mg, it binds to Sulphur in blood and is carried to the liver to be detoxified and then excreted in the urine. A protein diet always provides Sulphur which is essential for cyanogenic detoxification. In places when the diet lacks protein, cyanide detoxification becomes very minimal. So, in places where cassava has been established as the dominant staple food, a considerable amount of fish should be consumed as part of the diet to supplement the amount of protein.

The following are possible reasons for cassava–cyanide toxification that needs to be considered for healthcare: Firstly, sweet varieties may produce a high amount of cyanide. Secondly, unknown varieties may be introduced to the market, which may confuse people on determining the sweet cassava variety and thirdly, due to hunger, innocent children are likely to eat toxic cassava regardless of their bitterness. Furthermore, cyanide toxification can lead to the following long–time health effects: Dizziness, weakness, headache, and diarrhoea, goitre and cretinism in iodine–deficient, drop in blood pressure, mental confusion, stupor, blue discolouration of the skin because of deficiency of oxygen, convulsions and twitching in most cases leads to konzo and seldom death (Delange *et al.*, 1994; Mlingi *et al.*, 1992).

Severe toxicity is gaining popularity and is more documented in cassava–consuming countries (Teles, 2002). Most African countries use cassava as a staple food that contributes to severe

toxicity; for example, in Nigeria, it is estimated that the prevalence of tropical ataxic neuropathy is 6% overall, while a higher proportion is the women (7.7%) than men with only (3.9%). This massive occurrence is linked to cassava consumption (Oluwole *et al.*, 2000).

1.1.5 Cassava Processing

To effectively utilize cassava and related products, it needs to process the products before consumption to avoid cyanide intoxication. This processing helps obtain cassava products that can be easily preserved and improved qualities to be incorporated in diets. Cyanogenic in cassava is mainly found in three forms; bound cyanogenic, intermediate cyanohydrins and free cyanides (Cressey & Reeve, 2019). These three different forms behave differently during cassava processing. Different processing techniques have been employed to reduce cyanogenic content in cassava; these techniques include drying, fermentation and cooking. Fermentation is the most common cassava processing technique, in which cassava can either be submerged or undergo solid–state fermentation. Solid–state fermentation uses grated or sliced pieces of cassava, and these are typified for gari production. These techniques are performed by exposing to the natural atmosphere or by pressing in a bag.

Furthermore, fermentation may involve three ways; soaking the peeled whole root, soaking peeled and cut into pieces or unpeeled starchy roots, then can be left in water for an average of 4 to 6 days; these allow effective cyanogenic removal. These fermentation techniques are commonly used in Nigeria to produce *lafun* and *fufu* (local names in Nigeria). In some places where there is a high food shortage, famine and war fermentation can be less than two days, but these attributes to most reported cases of food poisoning.

Other traditional processing methods contribute to the reduction of cyanogenic in cassava and related products. Crushing and sun–drying remove about 96% to 99%. Pounding and boiling cassava leaves in water removes 99% of cyanogen; sun–drying of cassava leaves with or without steaming removes about 60% (Cooke & Maduagwu, 1978; Nambisan, 1994). The proposed traditional processing techniques are observed to be effective in cyanogenic removal. However, they are known to temper and lead to nutrient loss (protein, carbohydrates, vitamins, and minerals); especially, soaking and fermentation dissolves most of the water–soluble nutrients (Hotz & Gibson, 2007). According to Silano *et al.* (1981) around half of the nitrogen in cassava is lost in free amino acids during processing. Linamarin is an abundant cyanogenic compound whose glycosidic bond can be broken under complicated conditions like high pressure and high temperature and may require mineral acids. However, enzymatic hydrolysis breaks the glycosidic

bond easily to glucose and cyanohydrin (Fig. 2). Several researchers have investigated the use of linamarase to mechanically disrupted cassava tissue to facilitate hydrolysis of linamarin to release free cyanide (Ikediobi & Onyike, 1982; Petruccioli *et al.*, 1999; Yeoh & Sun, 2001). Therefore, enzymatic hydrolysis like additional linamarase to cassava products during processing is enough to make cassava products free from cyanide, retaining the nutritinients contents while simultaneously improving the protein content linamarase (Murugan *et al.*, 2012).

The use of catalytic enzymes in cyanogenic removal is associated with biotechnology, which either describes the microorganisms or provides a way to increase the amount of hydrolyzing enzymes. The use of microorganisms like *bacillus subtilis*, *klebsiella* sp, specifically lactic acid bacteria, impacts the fermentation process and leads to the release of bound minerals like magnesium, calcium, and notably the endogenous linamarase enzyme. Biotechnology aims to develop a starter culture for cassava processing that produces the enzyme linamarase, which intern hydrolyse linamarin. A broad range of microorganism from those that produce amylase enzymes for breaking down starch of lactobacillus Plantarum can produce linamarase purified and observed to have optimal properties at pH 5 to 8 temperature at 30° C to 40° C. So, the genetic manipulation is done, the plasmid profiles of lactobacilli have to be investigated to see the potential of establishing linamarase production. Sanni *et al.* (2013) reported that genetic manipulation and biotechnology are the best alternative and hope for advanced and improved cassava processing. This technique will ensure the products with high protein content, short processing time, improved stable qualities, high economic returns, products with stable desired qualities.

Table 1: Summarizing cassava processing techniques and their efficiency

S/ N	Technique/ method	Working principle	Efficiency range (%)	References
1	Boiling	By cooking is applicable for low cyanide cultivar.	50–80	Cooke and Maduagwu (1978), Nambisan (1994), Oke (1994), Nambisan (2011)
2	Drying(sun or oven)	Dehydration allows enzymatic reaction <50°C	82–99	Padmaja and Steinkraus (1995), Hotz and Gibson (2007)
3	Fermentation	By microbial process by socking, retting allows anaerobic process	70–98	Gidamis <i>et al.</i> (1993), Dufour (1994), Westby and Choo (1994)
4	Steam distillation	Apply distillate to fresh or fermented pulm	90–99	Meuser and Smolnik (1980), Satora and Tuszyński (2010)

1.1.6 Catalytic Properties, Functional Attributes and Industrial Applications of β – glucosidases

Linamarase is among the β -glucosidase enzymes (E.C.3.2.1.21), which according to glycosyl hydrolases, belong to GH 1, GH 3, GH 5, GH 9, GH 30 GH 116 families sharing amino acid sequence. The β –glucosidase is essential for hydrolysis of carbohydrate moiety (complex compound) to simple compounds like oligosaccharides by breaking the glycosidic bond (Bhatia *et al.*, 2002; Cairns & Esen, 2010; Morant *et al.*, 2008). These enzymes are available to almost all organisms like bacteria, eukaryotes, and archaea. They have several roles like lignification, biomass conversion, breaking down glycolipids and degradation of the cell wall of plants and microbes. The β -glucosidase is essential for metabolising cellulose hydrolysis's final stages (Dan Li *et al.*, 2013). The β -glucosidase enzymes receive products from cellulose enzymes, which break the complex cellulose to cellobiose and oligosaccharides, then β -glucosidase hydrolyse to final simple sugar (glucose). These enzymes that participate in the hydrolysis of cellulose are cellulose; please consider Fig. 3.

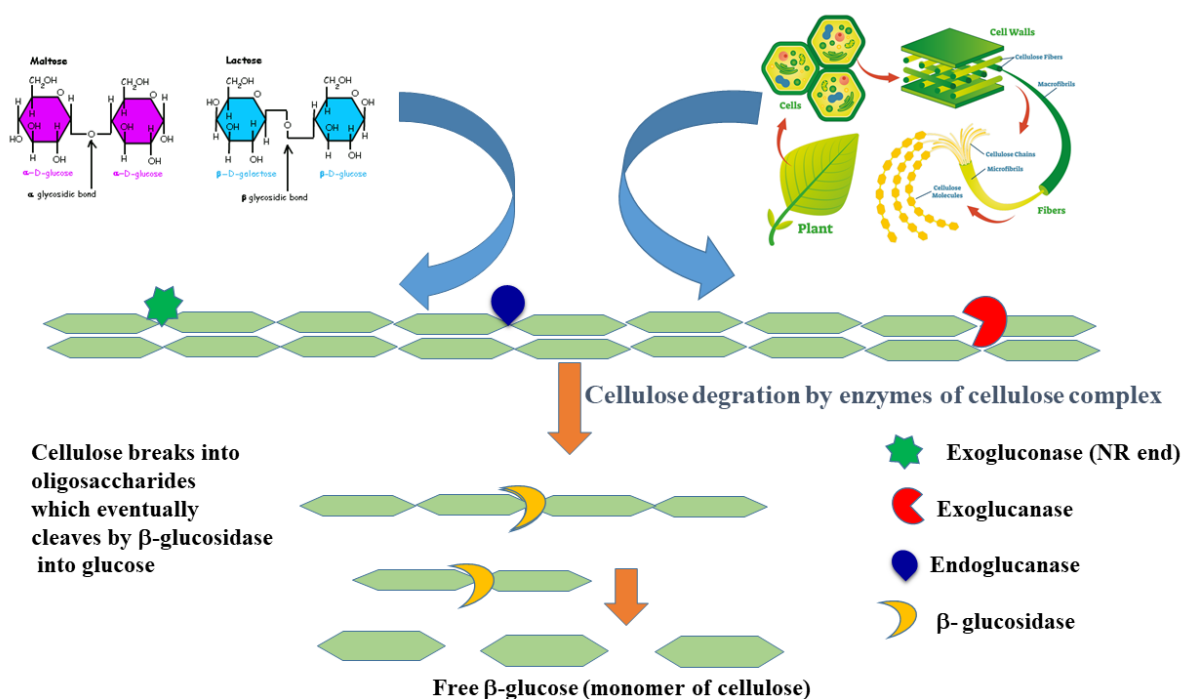


Figure 3: Schematic representation of the hydrolysis of cellulose using a series of reaction, cellulosic fibres broken down by Exo and endoglucanase to oligosaccharides; these small molecule oligosaccharides hydrolyzed to a final product that is glucose (Dan Li *et al.*, 2013)

The enzymes work in a dependent manner from endoglucanase (end-1, 4- β -glucanase [E.C.3.2.1.4], exoglucanase (cellobiohydrolase (exo-1, 4- β -glucanase [E.C.3.2.1.91]) and β -glucosidase (β -D-glucoside glycohydrolase [E.C.3.2.1.21]) (Teeri, 1997). The hydrolysis of cellulose starts with endoglucanase, which randomly hydrolyses from the middle portion of

cellulose, breaking the β -1-4 bonds. The exoglucanase acts at the ends of the cellulose to release oligosaccharides which finally are converted to glucose by β -glucosidase (Bhat & Bhat, 1997). The enzyme β -glucosidase is currently the research subject due to its potential application in many aspects. The following are some of the critical applications of β -glucosidase in biotechnology:

- (i) They are used in ethanol and biofuel production, specifically using cellulosic agricultural waste materials (Li *et al.*, 2013).
- (ii) Used in the extraction of juice and releasing of aroma from wine grapes, this is done by hydrolysis of bitter compounds (Gueguen *et al.*, 1998).
- (iii) It is helpful in the flavour industry; the aromatic compounds, when hydrolyzed, release flavours. It also uses any compound with glycosidic bonds like fruits and fermenting products (Krisch *et al.*, 2010). So, these are important in the food industry since they fasten the flavouring in tea, fruits juice and wine.
- (iv) Food, feed, textiles, detergents, and pharmaceuticals all benefit from the production of bioactive aglycones by β -glucosidases (Yin *et al.*, 2019).

Cassava β -glucosidase works using the same principle of cleaving the β -glycosidic bonds of the substrate linamarin to release glucose and unstable intermediate acetone cyanohydrin, which decomposes under favourable conditions to the carbonyl group and toxic cyanide, consider Fig. 2. The release of cyanide is the defensive mechanism of the cassava against predators. (<http://www.cazy.org/Glycoside-Hydrolases.html>) (Singh *et al.*, 2016).

The catalytic mechanism of β -glucosidase is described by the presence of Glutamate, which is the active site residue and conserved (Davies & Henrissat, 1995). The catalytic reaction of β -glucosidase is performed by two Glutamate; one works as a nucleophile (a conserved as 'I/VTENG' motif), and another residue act as acid/base catalyst (conserved as a 'TFNEP' motif) (Davies & Henrissat, 1995). The catalytic mechanism starts with glycosylation, where Glutamate acts as a nucleophile attacking anomeric carbon of glucose, resulting in intermediate glucose-enzyme. The second stage is deglycosylation, which involves water molecules that act as nucleophiles attacking the glycosidic bond and glutamate residue activates acid/base catalysis, resulting in cleavage of glycosidic bond and release glucose, consider Fig. 4.

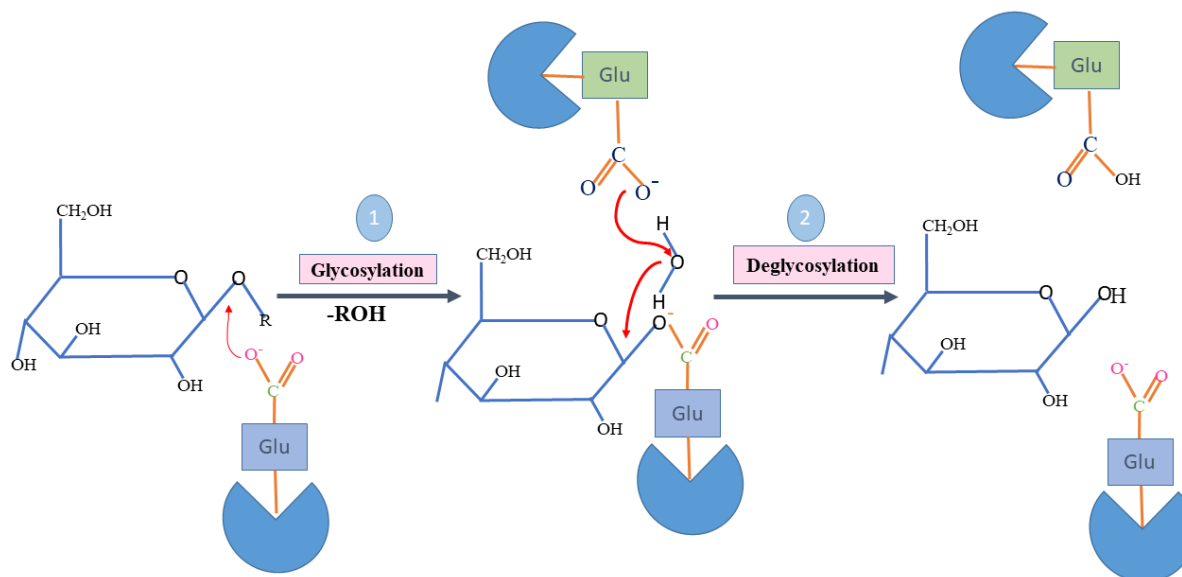


Figure 4: A reaction mechanism involving glycosylation and deglycosylation, hydrolysis and cleavage of the glycosidic bond (Singh *et al.*, 2016)

The unique property of linamarase differentiating it from the rest GH1 family is the ability to effectively catalyze the transglycosylation using primary, secondary and tertiary alcohol as acceptors (Svasti *et al.*, 2003) but cannot synthesize oligosaccharides and glycosides by reverse hydrolysis (Srisomsap *et al.*, 1999). To better understand the reaction mechanism and exhaust the potential of β -glucosidase, there is a need to determine the 3D structure. It will enable the determination of potential interaction modes and binding affinities of natural linamarin substrate (Kongsaeree *et al.*, 2010).

1.1.7 Linamarin as a Glycoside and its Pharmacological Potential

The natural products considered secondary metabolites are composed of a sugar moiety and non-sugar consistent called aglycone, containing a range of functional groups composed of glycosylated compounds; there are three main groups: cyanogenic glycosides, flavonoid glycosides and cardiac glycosides. The herbal glycosides have the pharmaceutical effect applicable in the medical and treatment of different diseases (Vetter, 2000). They are water-soluble metabolites responsible for regulating cell signalling and plant defence, and plant growth (Vetter, 2000). The glycosylated compounds have been observed to be potential for anti-tumour research (Kepp *et al.*, 2012).

The glycosylation process contributes significantly to the bioavailability, bioactivity, solubility and stability of the glycosides compounds. These properties are attributed to the blocking and hindrance of electron donation from aglycone to other reactive species/molecules (Poppenberger *et al.*, 2003). So, glycosylation hinders and makes the nucleophile aglycone to be inactive in the

chemical reaction. So eventually, the compound becomes stable (Osbourn, 1996). Because the sugar moiety can be easily polarized, it influences the solubility of hydrophobic compounds (Osbourn, 1996).

Cyanogenic glycoside is among the most crucial group of herbal glycosides, and it can be found in various plants, including cassava, which is edible to humans. Amygdalin and linamarin, the best known cyanogenic glycosides, are primarily extracted from sorghum and cassava plants (Kato & Terada, 2014; White *et al.*, 1998). These plants use cyanogenic glycosides for self-defence upon insect, animal or parasite consumption by releasing hydrogen cyanide, a toxic compound (Kato & Terada, 2014). When hydrolysed by β -glucosidase from the human intestine releases hydrogen cyanide, the cyanogenic glycosides cause toxicity (Shim *et al.*, 2016). Many findings have reported that cyanogenic glycosides have a potential therapeutic effect on different diseases such as hypertension, cancer, and Asthma (Lee & Moon, 2016).

Linamarin is among the cyanogenic glycosides, is mainly found in cassava and is highly protected from enzymatic degradation by compartmentalization in plant tissue. However, upon plant tissue rupture, cyanide is released as the results of linamarase hydrolysis. Linamarin has proved to have anti-cancer properties due to its ability to release cyanide. For linamarin to express the therapeutic properties, it is co-applied with linamarase; the possible mechanism here is the hydrolysis of linamarin and finally releasing cyanide, which blocks the mitochondrial respiratory chain which causes the death of the cancer cell (García-Escudero & Gargini, 2008). Linamarin and linamarase were observed to have cytotoxicity impact on MCF-7 and HT-29 the cancer cell lines, but also to Caov-3 (ovarian) and Hela (cervical adenocarcinoma) cancer cells, the effect was higher as compared to solitary linamarin (Idibie *et al.*, 2007; Yusuf *et al.*, 2006). When linamarase degrade linamarin, the toxic cyanide can be used for suicide gene therapy, commonly in cancer treatment. Suicide gene therapy incorporates the particular gene within a cancer cell to convert non-toxic to toxic, which eventually kills the tumour site (Zarogoulidis *et al.*, 2013).

A study by Girald *et al.* (2011) used a co-treatment linamarin/linamarase, which demonstrated the ability to cause oxidative stress, specifically glioblastoma cell line, which ends up cell death. In the same study, U87MG cells were genetically modified to induce expression of linamarase, which were grafted in Rats brain and the effect of linamarin was examined after local and systemic administration (Girald *et al.*, 2011).

Extraction and purification of linamarin from natural sources uses several methods and involve different solvents. Linamarin needs to be available for different applications like drug activities

and food processing (analytical analysis). A study by Idibie *et al.* (2007) extracted and purified linamarin for cytotoxicity activities. In this study, activated carbon batch adsorption was used, followed by ultrafiltration to obtain pure linamarin. The solvents used were a mixture of phosphate-buffered solution (PBS), methanol and ammonium sulphate. Other components were mixed and prepared with dimethyl sulfoxide (DMSO). Another study by Vetter (2000) used high-performance Liquid Chromatography (HPLC) with a mixture of solvents, water/methanol (80:20) and an alternative option of water/acetonitrile. The current study, which gave a high yield of linamarin, is by Samanthi *et al.* (2020); in this study, linamarin was extracted and purified from cassava root peels by the cryocooled method. The maximum yield reported was 18 g from 1 kg of fresh cassava root peels. There is no single study that investigated the properties of linamarin in different solvents. The potential of linamarin is limited by its difficulties in isolation from the natural source (Idibie *et al.*, 2007).

1.2 Statement of the Problem

Although cassava is among the major staple food crops, its effective utilization is limited by factors like susceptibility for rotting if stored for a long time, toxicity for the bitter variety, and a high cyanogenic glucoside level. The problems of cyanide content remain as potential toxic effects, leading to sickness and sometimes death (Oluwole *et al.*, 2000). Several processing techniques have been reported; these include: soaking, fermentation, grating, soaking or seeping, pounding, roasting, pressing, peeling, boiling, steaming and slicing (Cooke & Maduagwu, 1978; Nambisan, 1994). However, these traditional techniques are time-consuming, complicated, and highly contribute to the loss of nutritive value of cassava (especially protein, carbohydrates, vitamins and minerals). The best alternatives for cassava processing are of great importance; the methods should address the degradation of linamarin, which is abundant in cassava, to release its lethal nitrile group as free cyanide. More studies are currently focusing on using biotechnology using the enzymatic catalysis to hydrolyze the cyanogenic through glycosidic bond and finally make cyanide a free molecule. Using endogenous linamarase for hydrolysis of the cyanogenic compound will be of great importance as it simplifies the complicated processing procedures and retains the nutritive value while adding the protein content in cassava products. Biophysical characterization and other application of linamarase from cassava are limited due to the unavailability of crystallized structure in the Protein Data Bank (PDB). On the other hand, understanding the properties of linamarin substrate in different solvents is very important since it is always extracted and purified by different methods and different solvents.

Therefore, this work focuses on understanding the molecular structure and the interaction mechanisms of the linamarase–linamarin complex. In addition, the work identifies specific residues responsible for the enzyme’s reaction mechanism. Besides linamarase, the study also focuses on understanding the conformational properties of linamarin in different solvents.

1.3 Rationale of the Study

Understanding the structure of linamarase is an essential step in ensuring the monitoring and control the cassava processing. Since the application of linamarase enzymes for hydrolysis of cyanogenic is an effective method that ensures the nutritive value of cassava products are retained while also adding the protein content of cassava product. Its application is of great importance. Investigating linamarase at the molecular level will shed light on biochemists and chemists in understanding the enzyme's reaction mechanism and stability at different conditions and specific interacting residues. Biotechnology specific to linamarase in cyanogenic removal will expand the cassava market through industrialization and commercialization.

1.4 Research Objectives

1.4.1 General Objective

Structural determination of linamarase towards understanding its interaction mechanism and the effect of solvents on physical properties of linamarin.

1.4.2 Specific Objectives

- (i) To develop and validate the linamarase structure by using the homology modelling approach.
- (ii) To assess linamarase–linamarin complex interactions mechanism through molecular dynamics calculations.
- (iii) To investigate the stability and conformational of apo and holoprotein.
- (iv) To assess solvent effects on the physical and conformational stability of linamarin.

1.5 Research Questions

- (i) What is the optimal structure of linamarase obtained through modelling?
- (ii) How does linamarase–linamarin interact?

- (iii) What are the conformational stabilities of apo and holoprotein?
- (iv) What are the physical and conformational stabilities of linamarin in different solvents?

1.6 Significance of the Study

This study is deemed to come out with the best cassava treatment, which will effectively hydrolyse cyanogenic, maintain and retain the cassava's nutritive values, reducing the risks of human intoxication caused by cyanogenic consumption. In addition, this research will positively contribute to food processing technology and provide the solution to issues related to the adequacy of dietary energy supply and calories derived from traditional foods. It will help avoid contamination with biotoxins/microbial-derived toxins such as mycotoxins, cyanotoxins or chemical toxins such as agrochemical residues, inorganic species such as heavy metals contaminate during processing using the traditional methods.

This study will bring a significant technological, economic and environmental benefit since it is expected to be environmentally friendly and cost-effective as it will facilitate cyanide release by enzymatic hydrolysis. Apart from retaining cassava safety, it will unlock the potential of green technology applications, research, wealth, social, and the job will be created since cassava is among the most produced plants in Tanzania, Africa and globally; because it can tolerate and produce in harsh environments.

1.7 Delineation of the Study

A computational approach was used for this study to develop the linamarase structure and investigate the biophysical chemical properties through interaction with its substrate linamarin. For determining the linamarase structure, homology modelling was used. The modelled structure was subjected to molecular dynamics to determine conformational stability. The interaction of linamarase–linamarin was determined by molecular docking to identify the key residues participating in the interaction and maintain the complex's stability. The stability of the complex as well was determined by molecular dynamics. It is expected that the proposed linamarase structure will be an excellent starting point in designing and upscaling the use of linamarase in cassava processing. Moreover, linamarin's conformational stability was determined by molecular dynamics simulation. The properties of linamarin in different solvents verified the presence of specific groups, which contributes to their interactions.

CHAPTER TWO

LITERATURE REVIEW

2.1 Experimental Determination of Protein

The consistency of 3D structures determined has improved since the first protein, the myoglobin crystal structure, was solved in 1960; this was accomplished using experimental techniques such as X-ray crystallography and Nuclear Magnetic Resonance (NMR) spectroscopy (Muhammed & Aki-Yalcin, 2019). The experimental methods have limitations for determining some types of protein; for example, NMR works for small proteins while X-ray crystallography is suitable for molecules, which can crystallize. Therefore, like membrane protein, some protein is difficult to purify and crystallize; it is complicated to be determined by experimental methods. Moreover, these experimental methods are costly and time-consuming.

2.2 Structure Determination by Homology Modeling

It can be observed that the current number of experimentally determined proteins in the Protein Data Bank (PDB) to be 14 400 (Berman *et al.*, 2003). In contrast, the number of non-redundant sequences deposited in the National Center for Biotechnology Information (NCBI) is approximated to be 155 000 000, while for protein sequences is about 106 000 000 (Database resources of the National Center for Biotechnology Information [DNCBI], 2016). So this indicates that the protein sequences are significantly higher than the three-dimensional protein structure (3D) deposited in the PDB (the sequences are 736 times higher than the 3D protein). By 2006 the difference between annotated sequence available in NCBI and the solved structure (3D) in PDB was 120 times higher, but up to date, there is 736 times higher; this signifies an almost six times increase (Levitt, 2007). This growing gap between the sequence in NCBI and solved structure (3D) in PDB is a worrying condition that needs alternative methods to mitigate the situation. So homology modeling is a computational method used to predict the protein's structure using its amino acid sequence based on the template.

Homology modeling is considered to be a powerful method of predicting the structure of a protein. When the 3D structures produced by homology modeling were compared to those produced by de novo prediction or threading using the Root Mean Square Deviation (RMSD), the structures were more accurate (Werner *et al.*, 2012). Moreover, homology modeling is the best method in protein structure determination since it is cheap and fast with well-defined steps. Because homology modelling produces high-quality structures, it has influenced even the virtual screening and

docking methods in drug discovery (Cheng *et al.*, 2012). The steps of homology modeling will be discussed, the tools and servers, the application in drug discovery, and the challenges of homology modeling. The following are the most common and standard steps for homology modeling consider:

2.2.1 Identification of Templates

This process starts by using the target (query) sequence to search the template structure in the PDB, and this is done through the NCBI database using the Basic Local Alignment Search Tool (BLAST) (DNCBI, 2016) which it performs the pairwise sequence–sequence alignment. Another approach can be used in Uniprot (<http://www.uniprot.org/>) (Magrane & UniProt, 2011) profile–profile alignment (Wang & Dunbrack, 2004) also Hidden Markov Models (HMMs) (Söding, 2005). The identified template candidates are screened with similarity criterion, and the more similar sequences are trusted to generate a 3D structure with high quality and accuracy. The minimum sequence similarity for generating a 3D structure by homology modeling is > 25%. From the general rule, 3D structure generated from sequence similarities of > 50% is trusted and accurate enough for drug discovery application; those with identities between 25 and 50% can be used for design mutagenesis experiments and assess target druggability while those between 10 and 25% are not trusted to be best (Hillis *et al.*, 2004). There are also other essential factors to be taken into consideration, apart from high sequence similarity. These factors include: the target and template should have high similarity in the phylogenetic tree since the target sequence with the template is likely to produce a 3D structure with high quality and accuracy (Peng, 2013). Other factors are the pH of the environment, the solvent used, and the presence of bound ligands; the resolution of the experimental structure available in PDB is considered in choosing the template.

2.2.2 Sequence Alignment

After identifying the template(s), it is important to perform the target template's sequence alignment or template. A single template or multiple templates may be chosen to be aligned with the query (target). Alignment of the sequence is critical in determining the final 3D structure with high quality (Dalton & Jackson, 2007). A simple error during the alignment of amino acid causes the α carbon shift, so only one gap of residue in the α helix part contributes to the rotation of other residues in the helix. The most popular alignment methods (O'Sullivan *et al.*, 2004) include: Clustal W (<http://www.genome.jp/tools-bin/clustalw>) (Larkin *et al.*, 2007), MUSCLE (<https://www.ebi.ac.uk/Tools/msa/muscle/>) (Edgar, 2004) and ESPript (Robert & Gouet, 2014) for visualization.

2.2.3 Model Building

The 3D structure model can be generated by target sequence based on the selected templates. The model can be built by the following methods:

(i) Rigid-body Assembly

Works by breaking down the protein structure into different parts, the essential parts being the conserved core regions, side chains and loops. This approach uses the template protein structure to assemble these rigid bodies to build the protein model (Katebi *et al.*, 2010); the tools applicable here include SWISS-MODEL (Kim *et al.*, 2012).

(ii) Segment Matching

For this method, the atomic position of the template structure is clustered and used in determining the positions of atoms of the model. The segments of the known structure from the database are selected for matching the segments by considering the sequence identity, geometry and energy. This can be done by SEGMOD/ENCAD (Levitt, 1992).

(iii) The spatial Restraint Method

Depends on the template structure restraints to build the model. When developing the alignment, the restraints are produced to the target structure. The restraints developed depend on stereochemical restraints on the bond angle, bond length, dihedral angles, van der Waals contact distances; this can be done by MODELLER (Sali & Blundell, 1993).

(iv) The artificial Evolution Method

Combines the stepwise template evolutionary mutations and rigid-body assembly until the target sequence is the same as the template sequence; this can be done by NEST (An integrated model-building application that takes as input a sequence alignment of a target to one or multiple template PDB files and produces a model) (Petrey *et al.*, 2003).

2.2.4 Loop Modeling

Loops are the empty gaps that remain in the sequences of the homology proteins, and this is because some parts that are not conserved during evolution loops are often available even without deletions or insertions in the query or template. They are used to distinguish the protein specificity and their function. The quality of the model generated depends on the loop modelling. Since loops

greatly influence structural variability compared to strands and helices (Kmieciak *et al.*, 2016). Loops can be developed by either database search or conformational search approach. In the database search, the method searches in the PDB for the known protein to determine the segments. The conformational search is determined by the scoring function (Werner *et al.*, 2012). The *de novo* methods have been proposed to deal with the challenges of loop modelling. They predict the conformations of the loop by investigating the conformational space developed. The *de novo* methods involve Monte Carlo simulations, molecular, simulated annealing and genetic algorithms. Some servers can be used this include arch PRED (Howlin, 1993) and Congen (Vyas *et al.*, 2012).

2.2.5 Sidechain Modelling

The sidechain modeling is done by directing the side chains to the coordinates of the backbone of a particular protein obtained from *ab initio* modeling. The side chains are available to protein in very rarely several lower energy sources, also known as rotamers. Tools for side-chain modeling are RAMP (Krivov *et al.*, 2009) and SCWRL (Zhu *et al.*, 2008).

2.2.6 Model Optimization

It starts with energy minimization using the molecular mechanics force field (Krieger *et al.*, 2004). This step eliminates the significant error while minor errors are introduced, which later accumulate. So, to obtain the best energy minimized structure, the atom positions are restrained by a few hundred steps using force fields such as quantum force fields (Krieger *et al.*, 2002). The model can be optimized using molecular dynamic and Monte Carlo (Hooft *et al.*, 1996).

2.2.7 Model Validation

Since the developed models are expected to be applicable in different aspects/ areas, validation is crucial. The model's accuracy depends on several factors like sequence similarity, environmental and the template's quality. For proper model validation, it is crucial to perform stereochemical analysis to determine the torsion angle, bond length, and rotational angle. The following are popular tools for determining the stereochemistry of the modelled structure: PROCHECK (<https://www.ebi.ac.uk/thorntonsrv/software/PROCHECK/>) (Chen *et al.*, 2010), WHATCHECK (<https://swift.cmbi.umcn.nl/gv/whatcheck/>) (Laskowski *et al.*, 1993), and Molprobity (<http://molprobity.biochem.duke.edu/>) (Chen *et al.*, 2010). Another essential tool is the Ramachandran plot (<http://mordred.bioc.cam.ac.uk/~rapper/rampage.php>) to determine the protein structure's quality. It allocates the residues that are stereochemically optimized to be allowed and accepted regions while those

with problems are out of acceptable regions (Patel *et al.*, 2019). Other tools are applicable in determining the unique properties of the modelled structure based on 3D conformations; these include: PROSA II (<https://www.came.sbg.ac.at/prosa.php>) (Cavasotto & Phatak, 2009) and VERIFY3D (<http://servicesn.mbi.ucla.edu/Verify3d/>) (Wiederstein & Sippl, 2007). These tools consider the environmental parameters of the constructed model concerning the expected environmental conditions.

2.2.8 Application of Homology Modelling in Different Aspects

- (i) Used in structure-based drug design
- (ii) Analysis of mutations
- (iii) Insight into binding mechanisms
- (iv) Identification of the active sites
- (v) Looking for ligands and designing novel ligands
- (vi) Modelling of substrate specificity
- (vii) Protein-protein docking simulations

2.2.9 Challenges of Homology Modeling

There are countless numbers of homology modeling methods employed in building models. Many studies have verified that there is no single method or server that is superior to all others. So selecting a method depends on some factors like what is the intended application of the built model. Even though homology modeling is a powerful method in determining the 3D structure, the method faces the following challenges:

- (i) Since homology modeling builds models using sequence similarity, the template is unliganded. The model is also obtained without ligand information; therefore, there is a need to introduce complete automated homology modeling to accommodate the challenge (Gupta *et al.*, 2014).
- (ii) Another challenge, sometimes homology modeling, leaves unresolved questions in the built model; this can be solved using the coordinates of experimentally determined templates conceivable for targets (Bishop *et al.*, 2008).

- (iii) Homology modeling also faces loops and insertion challenges complicated to model (Geraldene & Mahmoud, 2017). The best models are obtained by optimizing the loop region and sidechain. One can use molecular dynamics to optimize through refinement of the structure. Using multiple sequence alignment is advantageous to solve low sequence similarity levels between target and template. However, using multiple sequences should be taken with great care by including templates from the analogous phylogenetic tree (Shen & Sali, 2006). After the homology modeling process, several models generated have different parameters; this allows researchers to choose their best.

2.3 Molecular Docking

The science of computer-aided drug design and discovery (CADD) have undoubtedly come to rescue a very long and tedious process of hunting the lead molecule. The macromolecules such as protein, RNA, and DNA are involved in biological processes; their interaction with small molecules (ligands) is crucial to regulate biological function. Molecular docking is a method, which helps to predict the favourable intermolecular orientation in the interaction between the protein (receptor) and ligand forming a stable complex. The docking method requires a protein from a high-resolution x-ray, NMR or a structured procedure by homology modeling with the pre-determined binding site. Therefore, the docking method works by fitting the ligand to the active site to anticipate its activity and affinity to the receptor Fig. 5 by optimizing different parameters like hydrophobic steric and estimating the binding free energy.

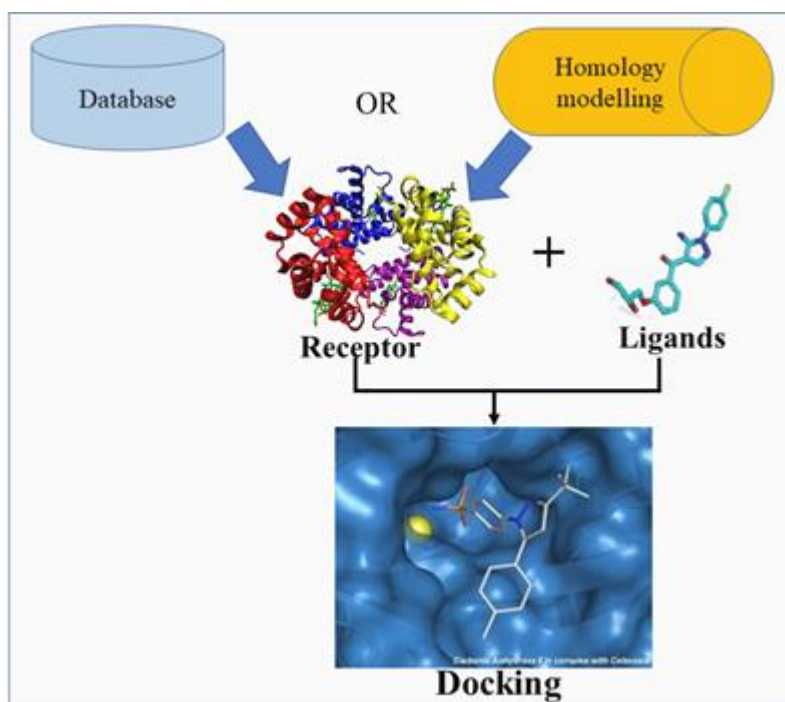


Figure 5: The flow chart of molecular docking

During the interaction between molecules, there are various binding modes generated. So, there are needs for algorithms that enable us to fish out the fruitful conformations. The algorithms are classified according to the degree of freedom they ignore.

2.3.1 Docking Algorithm

The first most straightforward algorithm treats molecules as two rigid bodies considering only six degrees of freedom (three rotational and three translational), good examples are: DOCK (Kuntz *et al.*, 1982), others which are matching algorithms (MA) include: LibDock (Diller & Merz, 2001), LIDAEUS (Wu *et al.*, 2003), PhDOCK (McCarthy *et al.*, 2003) and Q-fit (Jackson, 2002). The MA is a good algorithm because of its high speed. However, it has a list of limitations such as limited flexibility and needs prior defined receptor geometry.

The second algorithm is incremental construction (IC); it works by ligand fragmentation into different segments from the rotatable bonds. The fragment with maximum interaction with a receptor is known as an anchor. When the anchor is identified, other fragments are added by considering forming a hydrogen bond. When a specific fragment is added, the poses that indicate having the least energies are taken for the next iteration, making this algorithm robust and fast (Rarey *et al.*, 1996). This algorithm has been used in programs like FlexX (Rarey *et al.*, 1996), DOCK 4.0 (Ewing *et al.*, 2001), SLIDE (Fradera *et al.*, 2004) and FLOG (Miller *et al.*, 1994). This algorithm's major challenge/limitation is that, it works well with small-size ligands and not large ligands because large-sized ligands generate many fragments that cause problems.

The third algorithm is Monte Carlo (MC), which modifies the ligand by translation rotation several parameters are considered at a time. Molecular mechanics is used in evaluating the conformation at the binding site. So it can be accepted or rejected for the next iteration based on the Boltzmann probability constant. The conformation to be accepted or rejected depends on the change of energy for temperature. The following programs uses MC algorithms: FDS (Taylor *et al.*, 2003), Dockvision 1.0.3 (Hart & Read, 1992), GlamDock (Tietze & Apostolakis, 2007), ICM (Totrov & Abagyan, 1997), PRODOCK (Trosset & Scheraga, 1999), ROSETTSLIGAND (Meiler & Baker, 2006) and AutoDock (Goodsell & Olson, 1990). The major challenge is the uncertainty of the convergence; however, this can be improved by performing independent experimental runs.

The fourth is the genetic algorithm (GA); it uses an almost similar principle of MC in the search for global minima (Morris *et al.*, 1998). They use Darwin's theory, considering ligands' population; it alters the ligands by mutation or crossover. It starts by creating a populated list of ligands after assessment then the fittest ligands are selected from the previous step. The GA is applicable in a

program like DAR–WIN (Taylor & Burnett, 2000), DAVALI (Clark & Ajay, 1995), FITTED (Corbeil *et al.*, 2007), GAMBLER (Charifson *et al.*, 1999), PSI–DOCK (Pei *et al.*, 2006) and GOLD 3.1 (Verdonk *et al.*, 2005). Like MC, GA also have the challenge of the uncertainty of convergence.

The first algorithm is the hierarchical method; this considers the ligands conformations with low energy are aligned after pre-computational. A list of populated ligands conformations are combined and arranged in a hierarchical order, and this allows conformations with similar properties to be positioned closer in a hierarchical order. This hierarchical algorithm uses GLIDE software (Friesner *et al.*, 2004; Halgren *et al.*, 2004).

2.3.2 Scoring Function

Searching for the best conformations that best fit the receptor, a lot of sampling is performed from a vast degree of freedom. When one is interested in a single ligand that binds well to the receptor, ranking and evaluating the predicted ligand's conformation is paramount. Therefore, the scoring function implements the algorithms by predicting the orientation, representing the actual protein–ligand complex. On the other hand, if multiple ligands are considered, the scoring function should rank the ligands pose from the least to the highest binding energy. The scoring function uses a lot of assumption and simplification in estimating the binding free energy of protein–ligand complexes, which are very close as possible to the actual binding energy within a short time. The binding free energy (ΔG) is obtained from Gibbs Helmholtz (Equation 1) as:

$$\Delta G = \Delta H - T\Delta S \quad (1)$$

Where ΔG is the binding free energy, ΔH is the enthalpy, T is the temperature in kelvin, and ΔS is the entropy. The binding constant (K_i) is related to ΔG in the following ways:

$$\Delta G = -RT \ln K_i \quad (2)$$

Generally, the scoring function is classified into three main categories: force field, knowledge-based and empirical.

Force field functions are based on physical properties of atomic interaction such as electrostatic interaction, van der Waals interactions, bond angles, bond length and torsions (Huang *et al.*, 2006). This force field depends on ab initio quantum mechanics and experimental data, and it can be verified according to the Equation 3:

$$E = \sum_i \sum_j \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} + \frac{q_i q_j}{\epsilon(r_{ij}) r_{ij}} \right) \quad (3)$$

Whereby r_{ij} represent the distance between protein atom (i) and that of ligand (j) A_{ij} and B_{ij} stands for van der Waal parameters, q_i and q_j are atomic charges while $\epsilon(r_{ij})$ is the distance-dependent dielectric constant. The program DOCK (Meng *et al.*, 1992) is the most common example that uses the FF scoring function, but the main challenge is not considering the effect of solvent. To improve the previous program schoichet group (Shoichet *et al.*, 1999) include water in the interaction between protein and ligand using an implicit solvent model. It uses the Poisson–Boltzman to model the proteins electrostatic potential DelPhi (Nicholls & Honig, 1991) to estimate the electrostatic interaction between ligand and receptor through precomputed receptor potential HYDREN (Rashin, 1990) is used to calculate the ligand desolvation. The following scoring function is considered in this category, such as HADDOCK Score (van Dijk *et al.*, 2006), ICM SF (Totrov & Abagyan, 1997), Gold Score (Jones & Jorgensen, 1997) and QXP SF (McMartin & Bohacek, 1997).

2.3.3 Empirical Scoring Function

The basic concept of this scoring function is based on the data from X-ray structural information or experimentally determined binding energies. It sums up the individual terms. The oldest empirical scoring function the binding free energies was LUDI (Böhm, 1992).

$$\Delta G_{\text{bind}} = \Delta G_0 + \Delta G_{\text{hb}} \sum_{\text{h-bonds}} f(\Delta R, \Delta \alpha) + \Delta G_{\text{ionic}} \sum_{\text{ionic int.}} f(\Delta R, \Delta \alpha) + \Delta G_{\text{lipo}} |A_{\text{lipo}}| + \Delta G_{\text{rot}} \text{NROT}. \quad (4)$$

Where ΔG_0 stands for independent binding energy and ΔG_{hb} represents the contribution of binding energy obtained from hydrogen bonds. The ΔG_{lip} is the contribution from lipophilic interaction, A_{lipo} is the protein and ligand, ΔG_{rot} is binding energy loss caused by freezing of the internal degrees of freedom of the ligand. The NROT is for rotatable bonds while $f(\Delta R, \Delta \alpha)$ represents the penalty, accounting for the significant deviations from salt bridge geometry and ideal hydrogen bond.

So the different scoring functions included here include the following: SCORE2 (Böhm, 1998), RankScore (Moitessier *et al.*, 2006), HINT (Cozzini *et al.*, 2002), GlideScore (Friesner *et al.*, 2004) and LigScore (Krammer *et al.*, 2005).

2.3.4 Knowledge-based Scoring Function

This function uses information from the experimentally determined atomic structures; the interaction between ligand and protein provides interatomic contact frequency information analysed by this function. The function can be calculated by the Equation 5:

$$w(r) = -k_B T \ln [g(r)], g(r) = (r)\rho(r)/\rho^*(r) \quad (5)$$

Where T is the absolute temperature of the system, K_b is the Boltzmann constant, $\rho(r)$ is the number of density for a protein–ligand atom at distance r , $\rho^*(r)$ is the pair density, $g(r)$ is the pair distribution function. Examples of the function include PMF (Muegge, 2006; Muegge & Martin, 1999), SMOG (Ishchenko & Shakhnovich, 2002), BLEEP (Mitchell *et al.*, 1999), IT Score/SE (Huang & Zou, 2010), etc.

2.4 Molecular Dynamics

Molecular dynamics (MD) simulation is the science that tries to understand the properties of macromolecules by mimicking the behavior of atoms in real life and assuming given potential energy. In a molecular dynamics simulation, particles are allowed to interact at a specific time, resulting in a trajectory that contains details about the studied system's dynamical behavior. This section of the thesis starts by discussing the crucial components of the dynamic behavior of systems. In order to perform an MD simulation of a specific system, the following information is required at the microscopic level: Atomic positions and velocities, which assumes a particular potential energy function; consider Fig. 6. The energy function allows determining the force which each atom experiences at a given position to other atoms. Generally, the microscopic behaviour of a system is calculated by the laws of classical mechanics. Newton's laws tell how the forces affect the motions of the atoms.

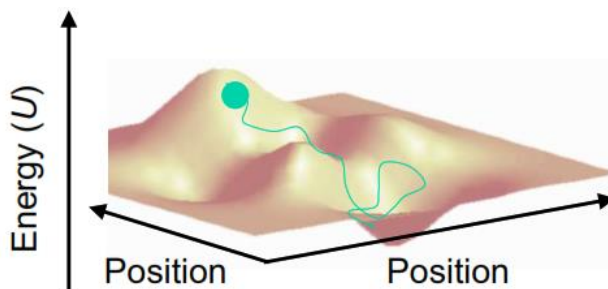


Figure 6: Relationship between energy function and position of atoms in a system

The force is measured as the derivative of the interacting potential, $V(R)$, in classical mechanics, and the force acting on each particle in the system is calculated using Equation 6.

$$F_i = m_i a_i = m_i \frac{d^2 r_i}{dt^2} = -\nabla V(R) \quad (6)$$

F_i represents all forces that act on i^{th} atoms in the system, making the vectorial summation; also, m_i is the mass of the atom, a_i is its acceleration. The positions of each atom can then be calculated using knowledge of the atomic forces and masses over a series of concise time steps (femto seconds about 1 to 10 fs). The initial positions and velocity must be known, typically defined in an input

data file, and the forces must be measured using potential energy. The initial coordinates at time $t = 0$ are derived from experimental data, but the velocities are unknown and assigned. A Maxwell–Boltzmann distribution can assign velocity at a given temperature, which is usually 300 K for many biological systems.

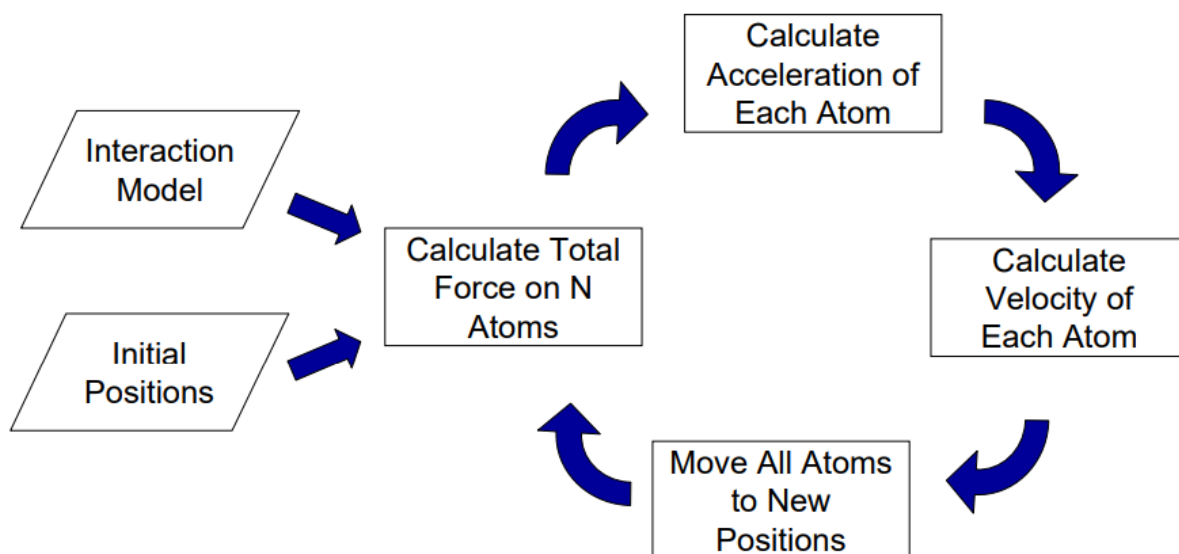


Figure 7: Schematic representation of molecular dynamics process

2.4.1 Energy Minimization

The initial structure used during the MD simulation is obtained from neutron, X–ray and electron diffraction, NMR, infrared, Raman and neutron spectroscopy. These molecules are unstable and contain many clashes, so energy minimization is critical to overcoming these challenges. Energy minimization is the method of arranging a series of atoms in space so that the net inter–atomic force on each atom is acceptable close to zero and the location on the potential energy surface (PES) is a stationary point, according to some theoretical model of chemical bonding.

2.4.2 Force Fields

A force field is a mathematical term that describes the relationship between a system's energy and the positions of its particles. It comprises an analytical form of the interatomic potential energy, $U(r_1, r_2, \dots, r_N)$, and several parameters that go into it. The parameters are usually obtained by fitting experimental data such as neutron, X–ray, electron diffraction, NMR, ultraviolet, Raman, and neutron spectroscopy to ab initio or semi–empirical quantum mechanical calculations. A typical FF expression, on the other hand, may look like this in Equation 7:

$$U = \sum_{\text{bonds}} \frac{1}{2} k_b (r - r_0)^2 + \sum_{\text{angles}} \frac{1}{2} k_a (\theta - \theta_0)^2 + \sum_{\text{torsions}} \frac{V_n}{2} [1 + \cos (n\phi - \delta)] \\ + \sum_{\text{improper}} V_{\text{imp}} + \sum_{\text{LJ}} 4\epsilon_{ij} \left(\frac{\sigma_{ij}^{12}}{r_{ij}^{12}} - \frac{\sigma_{ij}^6}{r_{ij}^6} \right) + \sum_{\text{elec}} \frac{q_i q_j}{r_{ij}} \quad (7)$$

The first four terms describe intramolecular or local contributions to the total energy (bond stretching, angle bending, and dihedral and improper torsions). In comparison, the last two terms describe repulsively and van der Waals interactions (in this case through a 12–6 Lennard–Jones potential) and Coulombic interactions.

A force field typically includes functional terms like Lennard–Jones, Coulomb, harmonic vibration, and model parameters that characterize the function terms. Different force fields are available for molecular dynamic simulation, for example, CHARMM27 (MacKerell *et al.*, 2000), GROMOS (Oostenbrink *et al.*, 2004), AMBER (Lindorff–Larsen *et al.*, 2010; Wang *et al.*, 2004) and OPLS (optimized potential for the liquid system) (Jorgensen *et al.*, 1996; Jorgensen & Tirado–Rives, 1988). So, choosing the force field to be used during MD simulation depends on the system. This study OPLS–AA force field has been used to study linamarase and linamarase–linamarin complex interaction simulation. The potential function of OPLS is described in Equation 7.

2.4.3 Integrating the Equation of Motion

After determining the forces of every atom in the system, using the force field, the position and velocity of each particle can be computed, and the time step Δt is minimal, about 1fs. So, extending the simulation time becomes complicated as it affects energy conservation. For integrating the equations of motion, some numerical algorithms have been developed, which include the following; Verlet algorithm (Verlet, 1967), Leap–frog algorithm (Van Gunsteren & Berendsen, 1988), Velocity Verlet (Swope *et al.*, 1982), Beeman’s algorithm (Beeman, 1976) and Symplectic reversible integrators (Legoll & Monneau, 2002; Martyna *et al.*, 1999). Verlet algorithm is frequently used in integrating the equation of motion (Verlet, 1967). The Verlet algorithm is derived by the Taylor expansion of the particle's coordinate at time $t \pm \Delta t$, where Δt is the time step.

$$\vec{r}_i(t \pm \Delta t) = \vec{r}_i(t) \pm \vec{v}_i(t)\Delta t + \frac{\vec{a}_i(t)}{2!}\Delta t^2 \pm \frac{\ddot{\vec{r}}_i}{3!}\Delta t^3 + O(\Delta t^4) \quad (8)$$

This sum up to the following Equation:

$$\vec{r}_i(t + \Delta t) = 2\vec{r}_i(t) - \vec{r}_i(t - \Delta t) + \vec{a}_i(t)\Delta t^2 + O(\Delta t^4) \quad \vec{r}_i(t + \Delta t) \approx 2\vec{r}_i(t) - \vec{r}_i(t - \Delta t) + \vec{a}_i(t)\Delta t^2 \quad (9)$$

In classical molecular dynamics, Equation 9 can be used to propagate the positions of each atom one at a time. The estimate of a new position in Equation 9 contains an error of the order of Δt^4 . However, the velocity terms are not explicitly included in this Equation. Equation 10 is used to calculate the necessary velocity from the trajectory:

$$v_i(t) = \frac{r_i(t+\Delta t) - r_i(t-\Delta t)}{2\Delta t}. \quad (10)$$

2.4.4 Molecular Mechanics Poisson–Boltzmann (Generalized Born) Surface Area (MM–PB(GB)SA)

Studying protein–ligand interaction is very important to quantify the binding free energy, so several computational approaches are proposed to estimate these energies. These approaches include thermodynamic integration (TI), free energy perturbation (FEP), linear interaction energies (LIE), molecular mechanics Generalized Born Surface area (MM–GBSA) and molecular mechanics Poisson –Boltzmann surface area (MM–PBSA) (Kollman, 1993; Meirovitch, 2007; Parenti & Rastelli, 2012; Ytreberg *et al.*, 2006). Among these computational approaches, TI and FEP are the most computationally expensive, most rigorous, they transform a molecule from initial to the final state over a reversible path, so they are involved in small molecule perturbations and are inefficient in protein–ligand complexes. So, the use of LIE or MM–PBSA is the best alternative as they estimate the free energy of initial and final states by using an ensemble structure, making these approaches highly efficient (Foloppe & Hubbard, 2006; Homeyer & Gohlke, 2012).

The MM–PBSA is the powerful computational approach that combines molecular dynamics; it considers the entropic contribution and conformational fluctuation to the binding free energy (Homeyer & Gohlke, 2012). The MM–PBSA computational approach depends on three energy terms to determine the binding free energy in a vacuum, which consist the bond, angle, torsion and non–bonded (that is for van der Waals and electrostatic interactions), another term is desolvation of various species (include polar and non–polar solvation energies). The third term is related to the entropy of configuration (complex formation in the gas phase).

The MM–PBSA is used to determine the binding free energy and decomposes the total binding energy to various components (Meliciani *et al.*, 2009; Venken *et al.*, 2011; Wang & Kollman, 2001). The general theory is describing the protein–ligand complex free energy insolvent and the isolated items. The MM–PBSA can also be used to describe protein–protein or protein–DNA complex interaction, and the description is given by:

$$\Delta G_{\text{binding}} = G_{\text{complex}} - (G_{\text{protein}} + G_{\text{ligand}}) \quad (11)$$

Where the free energy of each component is considered, Equation (12) is given:

$$G_x = \langle E_{MM} \rangle - TS + \langle G_{solvation} \rangle \quad (12)$$

Whereby x can be either protein, ligand, or complex $\langle E_{MM} \rangle$ describe the molecular mechanic's potential energy in the vacuum. The TS is an entropic contribution in a vacuum where T and S represent temperature and entropy, respectively, $\langle G_{solvation} \rangle$ is the free energy of solvation.

Describing more the molecular mechanic's potential energy, it includes the bonded and non-bonded energy terms, which the following Equation can calculate:

$$E_{MM} = E_{bonded} + E_{nonbonded} = E_{bonded} + (E_{vdW} + E_{elec}) \quad (13)$$

E_{bonded} describes the angle bond, dihedral and improper interactions, while the nonbonded, E_{elec} is for electrostatic and the van der Waals interaction. Another term from Equation (12) is the free energy of solvation which express the energy involved in transferring a solute from vacuum to solvent thus can be expressed by Equation (14):

$$G_{solvation} = G_{polar} + G_{nonpolar} \quad (14)$$

Whereby G_{polar} is electrostatic, and $G_{nonpolar}$ is non-electrostatic contributions. The polar solvation energy is detailed described by solving Poisson-Boltzmann (PB) Equation:

$$\nabla \cdot [\epsilon(r) \nabla \cdot \phi(r)] - \epsilon(r) \kappa(r)^2 \sinh [\phi(r)] + \frac{4\pi \rho^f(r)}{kT} = 0 \quad (15)$$

Whereby $\phi(r)$ for electrostatic potential, $\epsilon(r)$ is dielectric constant, $\rho^f(r)$ is the fixed charge density, and k^2 is associated with reciprocal of Debye length thus is much related to the ionic strength of the solution. The nonpolar solvation energy consists the terms like repulsive and attractive forces which exist between the solute and solvent (non-electrostatic terms), which are brought by the cavity formation and van der Waals interactions (Levy *et al.*, 2003; Tan *et al.*, 2007; Wagoner & Baker, 2006) these can be expressed by Equation (16):

$$G_{nonpolar} = G_{cavity} + G_{vdW} \quad (16)$$

Where G_{cavity} stands for work done by solute in creating the cavity, G_{vdW} is the solvent and solute van der Waals energy. Various models can be used to determine these energies; some of them are SASA-Only Nonpolar Model; the model highly depends on the energy contribution by nonpolar. It linearly depends on the nonpolar solvation energy; thus, it can be calculated as follows (Fogolari *et al.*, 2003; Rastelli *et al.*, 2010; Wang & Kollman, 2000).

$$G_{\text{nonpolar}} = \gamma A + b \quad (17)$$

Whereby A is SASA, γ is a coefficient which has relation to the surface tension of the solvent, and b is a fitting parameter.

2.4.5 Limitation of Molecular Mechanics Poisson –Boltzmann Surface Area

The MM–PBSA method is limited due to poor precision; despite its popularity in estimating free energy of ligands bound to their macromolecules, it has many standard deviation errors (Genheden & Ryde, 2015). When the method has poor precision, it becomes useless, and it is challenging to give results that can be trusted when one compares different approaches (Genheden & Ryde, 2015). So, to overcome such challenges, it is advised to perform many independent simulations. A study by Genheden and Ryde (2015) ran many independent in silicon experiments, an equilibrium of 100 ps and 100–200 ps, the time between 20–50 ns is sufficient to reduce the error.

2.5 Some Successful Examples and Applications of Computational Approaches in Structural Determination as well as Protein–ligand Interaction

2.5.1 Application of Homology Modelling and Molecular Docking

The previous sections have discussed different computational methods and their limitations in structural determination. For this section, the successful application of homology modeling, molecular docking, MD simulation and their related methods in structural determination and interacting residues to ligands and the impact in drugs and related fields are discussed. As described in section 2.2, homology modeling is a compelling method in determining the 3D model of different protein structure. This method is the best alternative to cover the gap between the essential data of protein sequences and the 3D protein determined and those deposited in PDB. The most critical challenge is obtaining the membrane protein, which is very complicated in extraction and purification, so their experimental structural determination is also complicated. Homology modeling is an appropriate method to overcome such challenges (Almeida *et al.*, 2017).

A study by Yin *et al.* (2019) was used to identify and molecular characterise psychrophilic GH1 β –glucosidase (BglG). The homology modeling was performed by using the SWISS–MODEL server, the template of β –glucosidase (PDB: 3ahx) of *Clostridium cellulovorans*, which have the highest identity (40%) and share similarity (59%). From the same study, four templates were used to perform multiple sequence alignment. The templates from *Bacillus polymyxa* (PDB: 1e4i, identity 39%, and similarity 57%) (González *et al.*, 1998), *Paenibacillus polymyxa* (PDB: 1uyq, identity 39%, and similarity 57%), *C. cellulovorans* (PDB: 3ahx, identity 40%, and similarity 59%)

(Jeng *et al.*, 2011) and *B. polymyxa* (PDB: 1tr1, identity 39%, and similarity 57%) (Lequerica *et al.*, 1998). Modeller 9.18 (a python editing software) was used to model three models, which later were subjected to PROCHECK and PROSA to determine the best generated model. After assessing the generated models, the first model with the smallest Molpdf and DOPE score close to 1 was identified as the best.

More analysis indicates 91.1% amino acids in the core area and only 0.2% in the disallowed area. Furthermore, the modelled structure from BglG (Yin *et al.*, 2019) was subjected to molecular docking using pNPG ligand, which was obtained from the PubChem database, Autodock 4.2.6 software was used. Five hydrogen bonds were observed to be formed between the pNPG ligand and surrounding amino acids. Three among the bonds were between E163 and pNPG ligands with 2.0, 2.8 and 2.0 Å, respectively, while E 373 was 2.4 Å hydrogen-bonded to pNPG. So, it was proved that E163 and E373 are catalytic residues in the active site of BglG (Stepper *et al.*, 2013).

Another study by Haghighi *et al.* (2020) used homology modeling and molecular docking to study the Glutamate Dehydrogenase (GDH) from *Cyanobacterium synechocystis* PCC 6803 the sequences of GDH (with accession number ALJ68932, 428aa) obtained from NCBI, to obtain the template, a BLAST search was performed against the PDB database. The BLAST results identified three templates 1EUZ with 49% (chain A) to 1GTM (chain A) with 49%, and 1V9L (chain A) with 42%. The target and template sequences were Clustal Omega (EMBL–EBI) (McWilliam *et al.*, 2013). Modeler 9v19 (Jianmin *et al.*, 2001) was used to generate 100 homology models; the best model was selected based on the lowest DOPE score and submitted for loop refinement. The ProSA web (Wiederstein & Sippl, 2007), RAMPAGE (Lovell *et al.*, 2003), TM-score (Zhang & Skolnick, 2005) and PROCHECK (Laskowski *et al.*, 1993) were used to check the quality of the refined model. The final refined model was minimized in chimera 1.12 (Pettersen *et al.*, 2004) by steepest descent for about 100 steps with 0.02 step size. The model was then subjected to molecular docking, which provided the three best poses. The docking model with ligands suggested the residues Lys 92, Lys 104, Gly 173, Arg 184, Asn 185, Ala 220, Val 322, Asn 324 and Ser 356.

More success stories are from researchers at former Ciba–Geigy (Cohen, 2007); they are the first group to develop renin inhibitor drug for hypertension, the structure was solved by homology modeling between peptide inhibitor (CGP38560) and renin (Cavasotto & Phatak, 2009). Moreover, a study by Kohring *et al.* (2008) developed a Plasmodium falciparum farnesyl transferase model that was used to optimize an antimalarial drug to specifically selective benzophenone-based inhibitors. A study by Noronha *et al.* (2007) used the homology model of

SAR to identify the best pharmacokinetics and inhibition of 1 Mm benzotriazine Src-kinase, which have better properties than other known benzotriazine.

Despite the popularity of homology, modeling is structural determination, and in drug discovery (i.e virtual screening, docking and MD simulations), there are many prejudgments in research and academic media. Most researchers believe it can hardly lead to a real drug; it is also perceived as a speculative theory with limited utility than producing “pretty pictures”.

The accuracy of homology modeling is highly dependent on the sequence similarity between the target and the templates. Models with a sequence identity of more than 50% are usually accurate, with only minor side-chain and loop positioning errors. This typical error is comparable to the typical resolution of an NMR-solved structure. More significant errors can be expected if the identity range is between 30 and 50%. Serious errors occur below 30% identity, which can result in a misfolded protein model. Critical errors can also occur in regions of the protein that have little sequence identity with the templates, despite having high sequence identity with the rest of the protein.

2.5.2 Applications of Molecular Dynamics

Structural determination by homology modeling is always followed by estimating binding affinity by molecular docking. However, molecular docking has the shortcoming of limited structural flexibility and inclusion of water molecules. Therefore, to accommodate such challenges, MD simulation is used, which allows total protein flexibility.

The study by Mazlan and Khairudin (2016) was performed on β -glucosidase (BglB) (EC 3.2.1.21) from *Paenibacillus polymyxa* (Esen & Blanchard, 2000) with PDB code 2O9T, which is capable of hydrolysing glycosidic bond specifically to cellobiose to glucose. In addition, the study continued molecular docking simulation, which identified potential residues as Glu 356 nucleophile and Glu 167 (acid/base catalyst) (Isorna *et al.*, 2007).

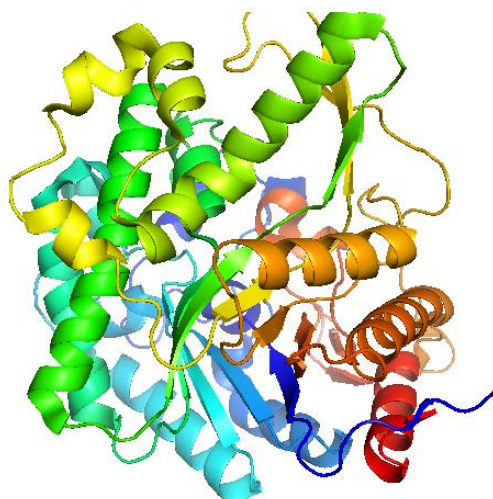


Figure 8: The structure of beta-glucosidase B with protein data base code 2O9T

These residues are found in the active pocket and are directly involved in binding with the substrate. Molecular docking helped to understand how substrate-bound into the active site and the crucial residues involved in the hydrolysis of cellobiose; MD simulation was used to understand the conformational change of BglB and its complexes. The trajectories were used to give important information on residues and substrate optimized conformations.

Four properties were considered to analyze the MD parameters, which include; Root Mean Squared Deviation (RMSD), Radius of Gyration (R_g), Root Mean Square Fluctuation (RMSF) and hydrogen bond analysis. Since using MD, it was interesting to measure the stability of protein and complexes over time. It was observed that BglB and complexes remained steadily and increased as they approached the end of the simulation (5 ns), and RMSD was ranging around 0.21 nm, which indicated a relatively rigid structure of BglB and the complexes. The substrate was observed to form with residues: ASN 407, Phe 418, Trp 402, Tyr 298, Asn 296, Glu 356, Glu 167, Asn 166, His 122, His 181 and Glu 409. Specifically, it was observed that Glu 167 was found at anomeric carbon making hydrogen bond throughout the simulation with OH1 acting as a proton donor. Moreover, Glu 356, a nucleophile residue, proved to make hydrophobic interaction with OH2.

The RMSF was used for determining the displacement of atoms around their average position for a specific time. It is based on the atoms of backbone residues; the regions with higher or lower flexibility were determined. Some residues on the dynamic behaviour of residues are determined. The most rigid amino acids were reported to move around 0.05 nm, while flexible residues moved around 0.34 to 0.40 nm. So, MD helped determine RMSF fluctuations of BglB, which have a smaller value than the complexes. The residues at 26–29, 2722–276, 43–53, 364–367 and 306–325. These residues were located on the surface and exposed on the loops or N and C terminal. Gln22, Try298, Glu167, Glu356, and Trp402 had small RMSF located at the binding pocket. So,

MD helps to determine the regions with high and low fluctuations. So, the core part of the protein is more rigid while the loop part is dynamic.

The MD simulation also determined the compactness of BglB by using a radius of gyration (R_g) to determine the loose package of protein. The lower the R_g indicates, the more package and compactness of the structure. From this study, R_g was determined and observed to rise R_g value up to the end of simulation time R_g dropped to 2.07 nm, verifying that BglB becomes compact at the end of the simulation. Moreover, hydrogen bond analysis was performed during the MD simulation and observed that residue Gln22 of BglB formed a hydrogen bond with OH1 of the substrate (60–80%) during the simulation time. Residue His122 and OH3, the hydrogen bond corresponding to 58.2% and 16.2%, respectively.

Protein flexibility is one of the significant limitations of standard molecular docking tools, which (for the most part) allow only the protein's side chain to be flexible, and this has remained a challenge (Kim *et al.*, 2018; Kokh *et al.*, 2011; Sinko *et al.*, 2013). Classical MD simulation is a powerful method for structural based drug design that allows protein flexibility. The challenges of static docking can be solved by ensemble-based molecular docking which is currently used to include different snapshots (Amaro *et al.*, 2008; Schames *et al.*, 2004), as is detailed in the following subsections:

2.5.3 Relaxed Complex Scheme

Understanding molecular recognition in its entirety is a problem of great interest in computer-aided drug design and molecular sciences in general. The interactions of ligand molecules with their receptors are dynamic and complex. Techniques that address these issues most effectively must account for the conformational flexibility of both the ligand and the receptor accurately and efficiently. While the ability to investigate ligand flexibility is well established, computer-aided drug design methodologies have only recently begun to consider receptor flexibility when looking for and optimizing functional inhibitors.

It is generally agreed that ligands can bind to receptor conformation, which rarely occurs in a range of receptor's dynamics. The local motion of residues at the active site can drastically alter and affect ligands' specificity and binding affinity to their target. Therefore, one needs to sample these few dynamics and incorporate the resulting conformation (Amaro *et al.*, 2008; Carlson, 2002; Davis & Teague, 1999; Ma *et al.*, 2002; May & Zacharias, 2005).

Using RCS, the protein structure from either PDB or homology modeling is subjected to MD simulation for a particular time scale. From the MD simulation, different snapshots are extracted

at a specific predetermined time interval like at every 20 ps, and then these biomolecules of snapshots are subjected to molecular docking experiments (Amaro *et al.*, 2008; Lin *et al.*, 2002). The MD simulation can produce many trajectory snapshots to be submitted for molecular docking (about 104 snapshots). However, these trajectory snapshots can be clustered into groups with very closely related configurations by using the base of RMSD. Findings reports that clustering by considering RMSD results into about a 99% reduction of computational expenses for the RCS docking (Amaro *et al.*, 2008). This study has used RCS for identifying and analyzing the interaction between the modelled linamarase and linamarin substrate. A good example is a study by Amaro *et al.* (2008), who used RCS to kinetoplastid RNA editing ligase one and W191G mutant of cytochrome c peroxidase. The study did straightforward practice by differentiating average and standard docking experiments and inclusion of RCS after MD simulation and obtaining snapshots. These examples verify that considering ensembles of the receptor is likely to change the predictive capacity of RCS calculations.

2.5.4 Application of Free Energies Methods

Computer-aided drug design involves a vast diversity of techniques used mainly in medicinal chemistry. These techniques can either be ligand-based or receptor-based approaches. Using ligand-based information of the ligand molecule that binds to the biological target of interest is used to identify the novel compounds. The receptor-based approach uses information from the receptor, which may be obtained from x-ray crystallography, NMR experiments, or homology modeling. Since the receptor-based approach enables evaluating the interaction between receptor and either novel putative ligand or already known ligand, this approach presents a severe advantage compared to the former approach. However, for that case, homology modeling remains the gold standard approach in determining the binding of the ligand to the target receptor (Tuccinardi, 2009).

The major limitation of molecular docking is determining and estimating the protein-ligand binding affinities determined using the scoring function. The scoring functions work by evaluating and ranking various binding poses, then later identifying the best pose that is considered top-scored (Tuccinardi *et al.*, 2015). The challenges of molecular docking can be overwhelmed by choosing the procedures that accurately estimate the protein-ligand binding affinities. In addition, this approach helps determine the free energy of binding, which is involved in forming protein-ligand complexes from the docking experiment. So, the commonly used method is molecular mechanics-Poisson Boltzmann Surface Area (MM-PBSA) (Kuhn & Kollman, 2000; Wang & Kollman, 2000).

A study by Botta and co-workers (2007) used virtual screening for the identification of RET kinase drugs from a database of synthesized compounds (Tuccinardi *et al.*, 2007). The human RET kinase was obtained from homology modeling, then authors docked with almost 170 putative inhibitors of RET kinase using the models of kinase, their MM-PBSA results were analysed, and ten best compounds were fished out subjected to inhibition assays. The experimental finding highlighted that the compounds had maximum inhibitory activity at 10 μ M in a range of 10 to 71%.

The study conducted by Grazioso and co-workers (2009) focused on 16 known $\alpha 7$ nAChR agonists, which were observed to have a potential affinity to $\alpha 7$ -subtype. First, the ligands were docked to 3D models of $\alpha 7$ nAChR, the complexes obtained were submitted to MD simulation at 500 ps with no distance or position constraints, and finally, energy minimized. Then MM-PBSA were used for evaluation of the binding free energy. Furthermore, results suggest a high correlation between predicted binding free energy and experimental results (giving $R^2 = 0.81$) (Grazioso *et al.*, 2009).

CHAPTER THREE

MATERIALS AND METHODS

3.1 Homology Modeling of Linamarase

Since the 3D structure of linamarase is not yet determined so it is not available in PDB. The modeling started by retrieving the amino acid sequences of cassava linamarase from the online database NCBI (DNCBI, 2018). The amino acid sequences were used as search queries to obtain the potential templates for homology modeling the information linked to Research Collaborator for Structural Bioinformatics PDB (Burley *et al.*, 2018) explicitly using the BLAST algorithm (Altschul *et al.*, 1990). The template fished out based on sequence similarity protein conformation and X-ray resolution. The target sequence was aligned with selected template sequences by Clustal W (Thompson *et al.*, 1994) to determine their similarities. The obtained file from Clustal W (Thompson *et al.*, 1994) was submitted to ESPript 3 (Robert & Gouet, 2014) for more secondary structure display and analysis.

For homology modeling were performed by using SWISS-MODEL (Waterhouse *et al.*, 2018), MODELLER (Sali & Blundell, 1993) and I-TASSER web server (Roy *et al.*, 2010); this was based on maintaining the reliability of the generated models. Using SWISS-MODEL, the 3D model was quality screened using Global Model Quality Estimation (GMQE), which should qualify in the range of 0 to 1. More analysis of the nativeness and agreement between the model and the experimental value done by Quality Model Energy Analysis (QMEAN) (Benkert *et al.*, 2008).

Using the MODELLER program (Sali & Blundell, 1993), python scripting was employed to determine the template for modeling. The best models were determined by ranking them to DOPE (Shen & Sali, 2006). Since I-TASSER (Roy *et al.*, 2010) is an automated web server, the model was generated automatically. All models from the modeling softwares were submitted for the quality assessment using PROCHECK (Laskowski *et al.*, 1993) and the web server SAVES 6.0 (Colovos & Yeates, 1993). To improve the quality of the models from both modeling free softwares, Coot software (Emsley & Cowtan, 2004) were used to perform structural refinement, and the best template was used to provide the electron map (MTZ file from PDB).

3.2 Molecular Docking

3.2.1 Protein (Receptor) Preparation

The best-determined model from homology modeling was initially prepared for docking calculation using chimera (Pettersen *et al.*, 2004) whereby the hydrogen at pH 7.4 and Gasteiger charges (Hou *et al.*, 2013) was added. For optimal performance, during docking, the protein structures were converted from PDB to pdbqt format using the AMDock tool (Valdés-Tresanco *et al.*, 2020). To determine the effect of structural and conformational changes when modelled protein is complexed with ligand, relaxed complex scheme (RCS) to allow full flexibility during the molecular docking. The molecular dynamic simulation (a long run) of 200 ns was performed in the RCS, which allowed obtaining 50 receptor ensemble structures extracted after every 4 ns.

3.2.2 Ligand Preparation (Linamarin)

The ligand was retrieved from the PubChem database, the 3D structure obtained with (CID:11128) (Kim *et al.*, 2021) and was saved in PDB format. The ligand was further prepared in a chimera (Pettersen *et al.*, 2004) by energy minimization, adding hydrogen and Gasteiger charges (Hou *et al.*, 2013), then finally converted to pdbqt.

3.2.3 Docking Bench Making and Method Validation

Making the docking calculation effective and trusted, the protocol was validated by comparing the set of experimental binding free energy from Baum *et al.* (2009) against the obtained binding free energy from docking calculation. A four times docking calculation was performed to determine trusted binding free energy. The docking calculation was performed using the Autodock vina (Trott & Olson, 2010) the default setting was retained. The calculated and experimental binding were compared, and the correlation was determined to allow the docking experiment to proceed (a correlation value obtained was 0.69). Two approaches were used to determine the activities of the modelled protein. The first approach was blind docking, whereby the natural ligand (linamarin) explored almost all favourable binding poses of the modelled structure. Furthermore, CASTp (Tian *et al.*, 2018) used an online server to predict the binding sites.

The predicted active site and the most favourable binding poses of the modelled protein have been summarised in Table 9. The residues represent the active site of the modelled protein. The active binding site was predetermined from the ensemble structure, and the natural ligand was allowed to dock to all the extracted 50 snapshots. The binding affinity of each structural protein

conformation was determined. Autodock vina tool (Trott & Olson, 2010) was used for docking calculations.

3.3 Molecular Dynamic (MD) Simulation

To determine the structural dynamic and stability of the initially modelled protein apo linamarase and complex (holo-structure), classical MD simulation was employed using Gromacs version 2018 (Abraham *et al.*, 2015). The ligand topology was determined using the LigParGen server (Dodda *et al.*, 2017; Jorgensen & Tirado-Rives, 2005), while the topology of the protein receptor was obtained by using the OPLS-AA force field (Robertson *et al.*, 2015). The system was explicitly solvated using SPC/E (Mark & Nilsson, 2001) water model in a cubic box. The system was brought to neutral condition by adding 5 Na⁺; energy minimization was performed for 500 steps.

The steepest descent algorithm was used to remove any available constraints and crushes (Fletcher & Powell, 1963). After energy minimization, the system was subjected to two equilibration steps. The first equilibration was performed at NVT (Constant Number of Particles, Volume and Temperature), commonly known as NVT ensemble for 100 ps. During the equilibration Berendsen method (Berendsen *et al.*, 1984) was used in maintaining the temperature and pressure at 300 K and 1 bar, respectively. After equilibration, the production run was performed for 200 ns at NPT (Constant Number of Particles, Pressure and Temperature) ensemble using Parrinello_Rahman barostat (Parrinello & Rahman, 1981) for the temperature at 300 K, for pressure coupling at 1 bar and v-rescale (Bussi *et al.*, 2007). The Particle-Mesh Eward (PME) (Cheatham *et al.*, 1995; Ewald, 1921) method was applied to treat long-range electrostatic interactions and LINCS (Fadrná *et al.*, 2005) algorithm was used for constraining the covalent bonds, the time step of 2 fs was used for all calculations. During the whole simulation time, the energetics, pressure and density were monitored. After attaining the equilibrium, the MD trajectory of the apoprotein was used to obtain 50 conformations sampled after every 4 ns, then subjected to relaxed complex scheme (RCS) docking calculations. For determination of conformation stability of the apoprotein as well as complex, the following properties were used: Root mean square deviation (RMSD), a radius of gyration (R_g), root mean square fluctuations (RMSF) and solvent accessible surface area (SASA).

3.4 Determination of the Binding Free Energy by Molecular Mechanics Poisson – Boltzmann Surface Area Analysis

Because of the limitations of molecular docking in estimating the stability and binding affinity of the protein-ligand complex, the MM-PBSA calculations were applied to the complexes. The

binding free energy was determined by using the gMMPBSA tool from GROMACS (Kumari *et al.*, 2014); a total of 200 snapshots were sampled. This calculation (EMM) represents a total van der Waal and electrostatic interactions energies from the gas phase. The $\Delta G_{\text{solvation}}$ represents polar and non-polar in which polar part (ΔG_{PB}) is determined linearly, solving the Poisson–Boltzmann Equation. The solvent dielectric constant (ϵ) was set at 80 while that of the solute was at 6. Moreover, different solute dielectric constants i.e. 2,4, were also tested. The term γ represents the surface tension proportionality constant set at $0.0072 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$ and the nonpolar solvation-free energy for solute set at $0.00 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$.

3.5 Linamarin Conformation in a Vacuum and Different Solvents

The crystal structure of linamarin was obtained from Cambridge crystallographic Data centre. A single molecule of linamarin was solvated in different solvents, including SPC/E (Mark & Nilsson, 2001) water model, methanol (MeOH), dimethyl sulfoxide molecules (DMSO). The solvated systems were submitted to energy minimization using the steepest descent algorithm to remove any crushes and constraints from the starting coordinates. The systems were equilibrated at NPT ensemble at a temperature of 300 K and pressure of 1 bar for 500 ps and 0.1 fs time step NPT ensemble was used to carry out the production run, keeping the temperature at 300 K, pressure at 1 bar for 200 ns and using Parrinello–Rahman barostat (Parrinello & Rahman, 1981). To constrain the bond LINCS (Fadrná *et al.*, 2005) was used, and Particle mesh Ewald (PME) (Cheatham *et al.*, 1995; Ewald, 1921) were used for handling the long-range electrostatic interactions. The simulation was performed by GROMACS code (Abraham *et al.*, 2015) version 2018 with OPLS-AA force field (Robertson *et al.*, 2015). The same procedure was performed for linamarin in a vacuum, except that the equilibration and production were performed at the NVT ensemble.

CHAPTER FOUR

RESULTS AND DISCUSSION

4.1 Homology Modeling of Linamarase

The SWISS-MODEL was used after the template search, and the best-identified template was Rice (*Oryza sativa* L), bound with 2,4-dinitrophenyl-2-deoxyl-2-fluoroglucoside (DNP2FG) with the PDB ID 3PTK. The template shares sequence similarity with that of the query by 0.49 and identity of 53.59%. The model from SWISS-MODEL shows QMEAN and GMQE values of -1.04 and 0.76, respectively. For more analysis, the GMQE estimates the relationship between target-template and target-structure quality by combining these two properties; the score ranges from 0 to 1. When the value is higher, it indicates the model is reliable, but when it is low, the model is unreliable. The model generated from the SWISS-MODEL had a 0.76 value, which suggests the high quality of the generated structure (Ghaffari *et al.*, 2020). The QMEAN scoring function determines the quality of global and local modelled structure, and the value always ranges from -4 to 0. When the value is closer to zero, this indicates a good score between the experimental and modelled structures.

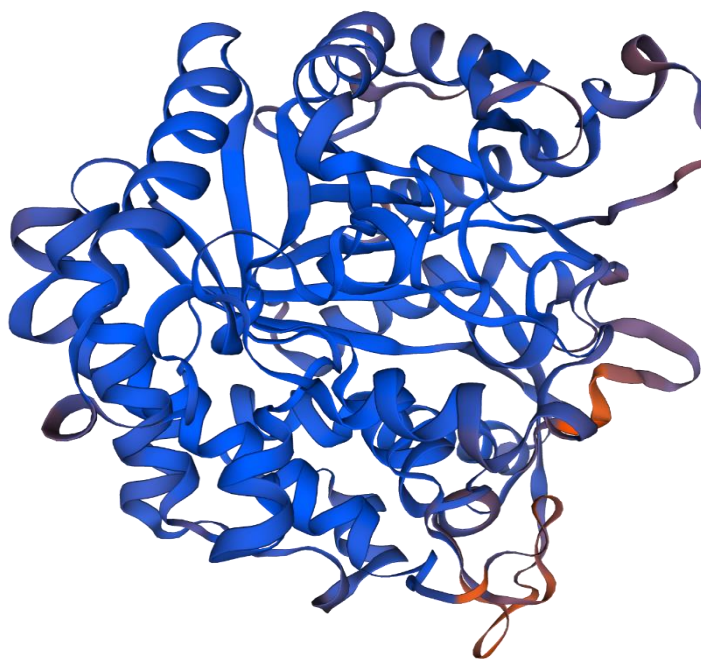


Figure 9: Protein structure generated by SWISS-MODEL

The analysis of the model's quality obtained from SWISS-MODEL was presented in Fig. 10.

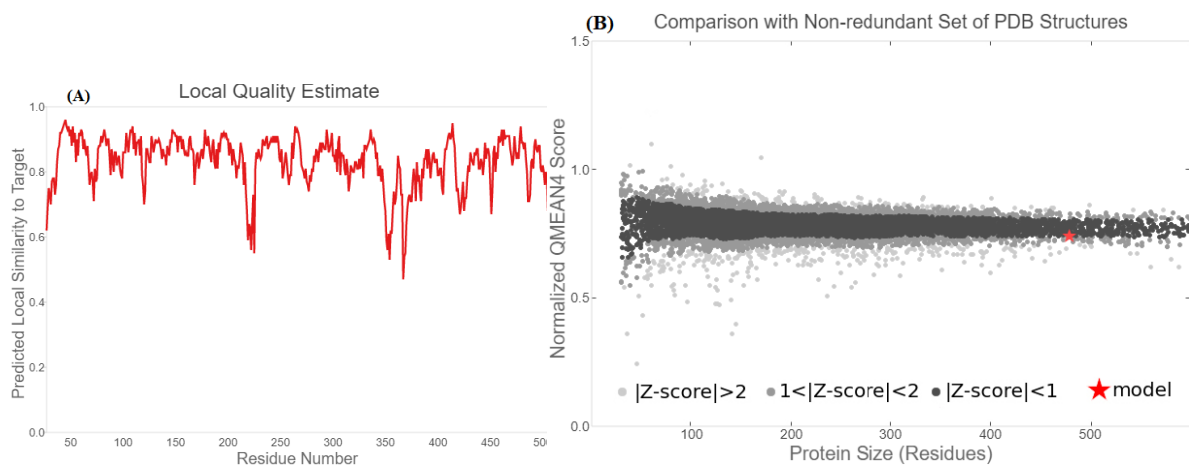


Figure 10: Structure validation of modelled Linamarase: (A) Local quality estimate of the residues of the predicted Linamarase model and (B) comparison of the predicted Linamarase structure with a nonredundant set of protein data base structures

The modelled structure was sequence aligned with the rice as a template, sorghum, and maize produced using Clustal W (Thompson *et al.*, 1994) after aligning the program and visualization software known as ESPript (Robert & Gouet, 2014) was used (Fig. 11).

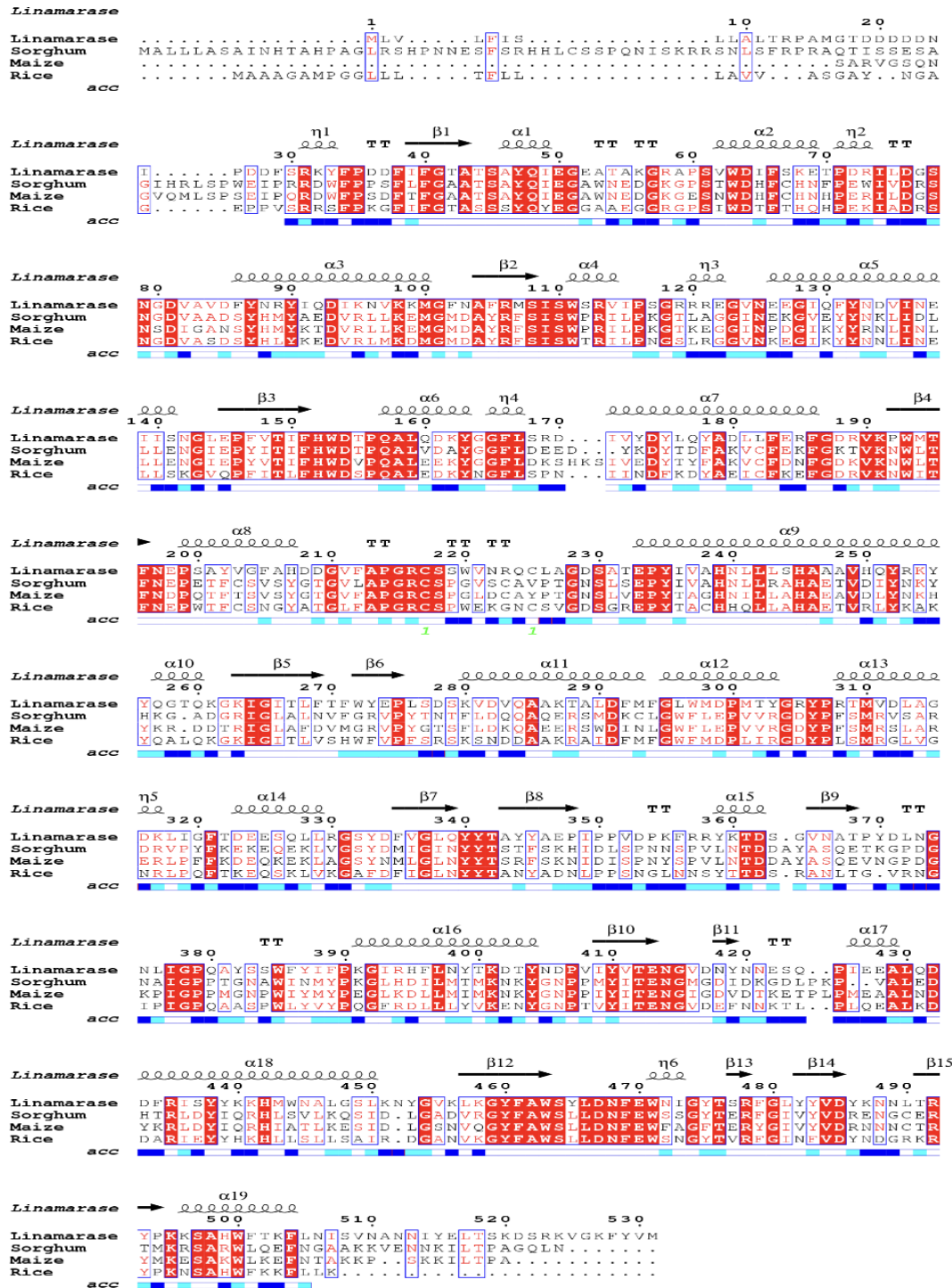


Figure 11: Multiple sequence alignment for modelled linamarase template from rice, sorghum and maize. The red highlighted represents identical; the secondary structures are presented on the top while the residues' surface accessibility to solvent is displayed at the bottom (the blue for surface accessible, cyan for partially accessible and white for buried residues)

The modelled structures generated by using three free software's, SWISS-MODEL, MODELLER and I-TASSER, were subjected to structural refinement by using the Coot software (Fig. 12) represent the Coot refinement process. The model structure after and even before refinement was submitted to SAVES 6.0 web server, which has multiple analysis, PROCHECK (Laskowski *et al.*,

1993) were used to determine the quality of the models and residues, which are outliers. The quality and accuracy of the models obtained from SWISS-MODEL, MODELLER, and I-TASSER were determined. From PROCHECK, their main and side chains are allowed to determine the stereochemical parameters. The Ramachandran analysis was summarized for models, which can be seen in Table 2 while stereochemical parameters summarized in Table 3, Table 4 and Table 5 for main parameters and Table 6, Table 7 and Table 8 for side-chain parameters. The SWISS-MODEL proved to generate the best model since it gave the following analysis: 89.6%, 9.6%, 0.7% and 0.0% of its residues in the most favoured, generously allowed and disallowed regions, respectively, which the best compared to parameters from other softwares (Paul *et al.*, 2021). The best model which was obtained from the SWISS-MODEL is presented in Fig. 13

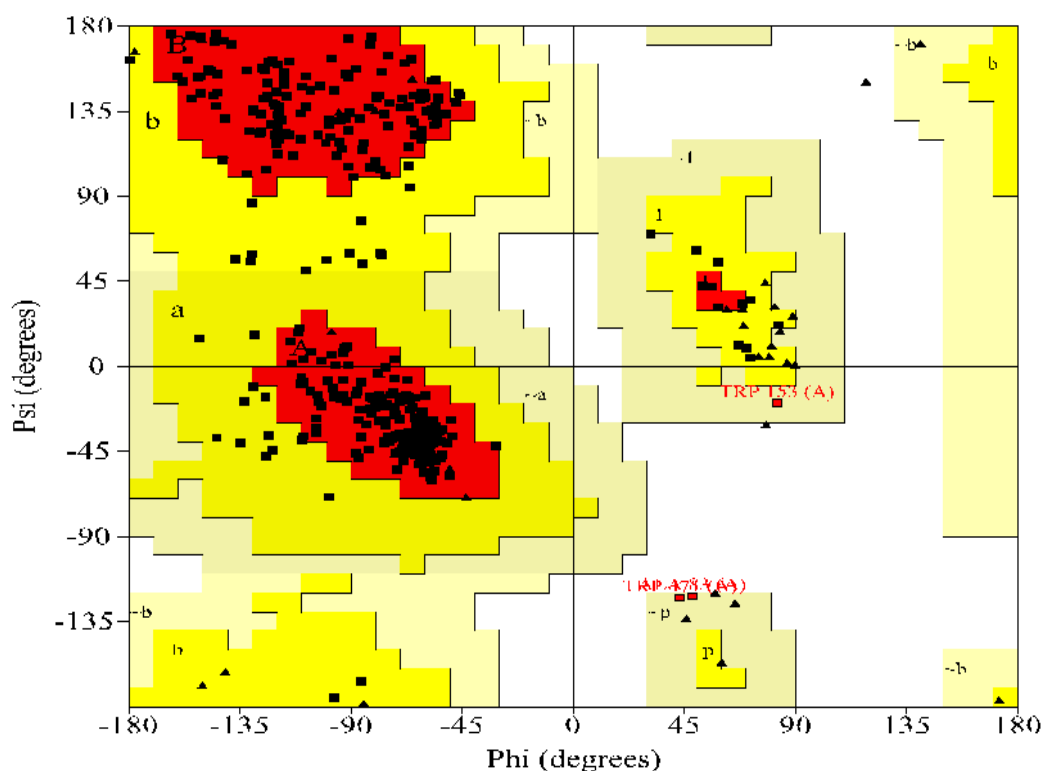


Figure 12: Structure validation using Ramachandran plot of modelled protein from SWISS-MODEL

The models from MODELLER at least provided good analysis, but I-TASSER had the poor model with many outliers, as shown in Fig. 13.

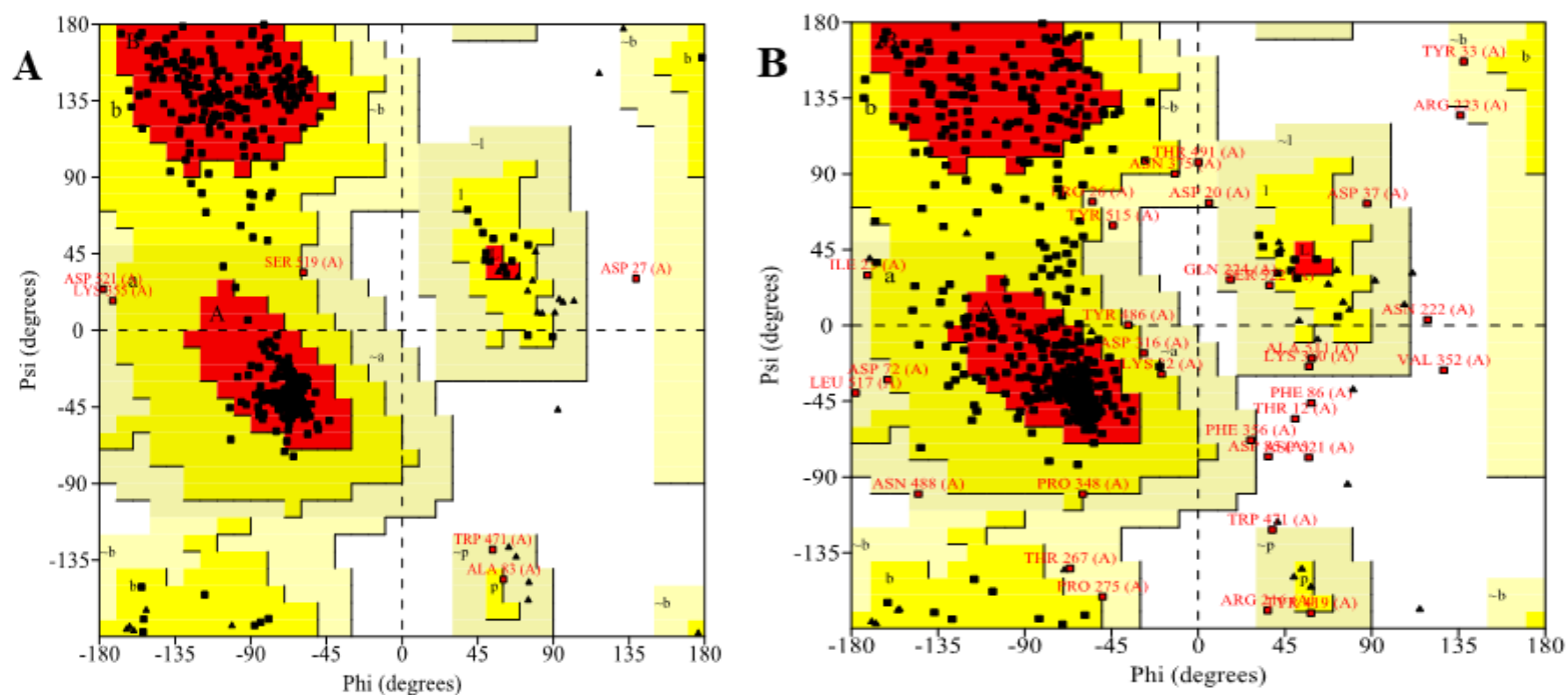


Figure 13: Structure validation using Ramachandran plot of modelled protein from MODELLER (A) and I-TASSER (B)

Table 2: Ramachandran plot statistics analysis of modelled linamarase from SWISS-MODEL, MODELLER and I-TASSER

Free Modeling software	SWISS-MODEL(%)	MODELLER(%)	I-TASSER(%)
Residues in most favoured regions [A,B,L]	89.6	89.0	73.2
Residues in additional allowed regions [a,b,l,p]	9.6	9.7	20.6
Residues in generously allowed regions [~a,~b,~l,~p]	0.7	1.1	4.7
Residues in disallowed regions	0.0	0.2	1.5

Table 3: Summary of main chain parameters for the structure generated by SWISS-MODEL

Stereochemical parameter	No of data pts	Parameter value	Comparison typical value	Value band width	No. of band widths from mean	Interpretation
%-tage residues in A, B, L	415	89.6	88.2	10	0.1	Inside
Omega angle st dev	473	9.0	6.0	3.0	0.1	WORSE
Bad contacts / 100 residues	1	0.2	1.0	10.0	-0.1	Inside
Zeta angle st dev	440	2.0	3.1	1.6	-0.7	Inside
H-bond energy st dev	322	0.7	0.7	0.2	0.1	Inside
Overall G-factor	476	-0.1	-0.2	0.3	0.3	Inside

St. Dev. Denotes the standard deviation of the score observed. The accuracy of the structure is depicted in the order BETTER>INSIDE>WORSE each parameter

Table 4: Summary of main chain parameters for the structure generated by MODELLER-MODEL

Stereochemical parameter	No of data pts	Parameter value	Comparison typical value	Value band width	No. of band widths from mean	Interpretation
%-tage residues in A, B, L	462	89.0	88.2	10	0.1	Inside
Omega angle st dev	525	5.5	6.0	3.0	-0.2	Inside
Bad contacts / 100 residues	18	3.4	1.0	10.0	0.2	Inside
Zeta angle st dev	491	1.7	3.1	1.6	-0.9	Inside
H-bond energy st dev	326	0.7	0.7	0.2	0.1	Inside
Overall G-factor	529	-0.1	-0.2	0.3	0.3	Inside

St. Dev. Denotes the standard deviation of the score observed. The accuracy of the structure is depicted in the order BETTER>INSIDE>WORSE each parameter

Table 5: Summary of main chain parameters for the structure generated by I-TASSER-MODEL

Stereochemical parameter	No of data pts	Parameter value	Comparison typical value	Value band width	No. of band widths from mean	Interpretation
%-tage residues in A, B, L	466	73.2	88.2	10	-1.5	WORSE
Omega angle st dev	528	17.7	6.0	3.0	3.9	WORSE
Bad contacts / 100 residues	1	0.2	1.0	10.0	-0.1	Inside
Zeta angle st dev	493	4.0	3.1	1.6	0.5	Inside
H-bond energy st dev	352	0.6	0.7	0.2	-0.3	Inside
Overall G-factor	531	-0.6	-0.2	0.3	-1.3	WORSE

St. Dev. Denotes the standard deviation of the score observed. The accuracy of the structure is depicted in the order BETTER>INSIDE>WORSE each parameter.

Table 6: Summary of side-chain parameters for the structure generated by SWISS-MODEL

Stereochemical parameter	No of data pts	Parameter value	Comparison typical value	Value band width	No. of band widths from mean	Interpretation
Chi-1 gauche minus st dev	55	10.9	13.6	6.5	-0.4	Inside
Chi-1 trans st dev	124	11.0	15.3	5.3	-0.8	Inside
Chi-1 gauche plus st dev	208	9.1	13.8	4.9	-1.0	Inside
Chi-1 pooled st dev	387	10.0	14.3	4.8	-0.9	Inside
Chi-2 trans st dev	111	12.8	17.7	5.0	-1.0	Inside

St. Dev. Denotes the standard deviation of the score observed. The accuracy of the structure is depicted in the order BETTER>INSIDE>WORSE each parameter

Table 7: Summary of side chain parameters for the structure generated by MODELLER-MODEL

Stereochemical parameter	No of data pts	Parameter value	Comparison typical value	Value band width	No. of band widths from mean	Interpretation
Chi-1 gauche minus st dev	68	6.0	13.6	6.5	-1.2	BETTER
Chi-1 trans st dev	146	9.1	15.3	5.3	-1.2	BETTER
Chi-1 gauche plus st dev	219	7.7	13.8	4.9	-1.2	BETTER
Chi-1 pooled st dev	433	8.0	14.3	4.8	-1.3	BETTER
Chi-2 trans st dev	103	10.3	17.7	5.0	-1.5	BETTER

St. Dev. Denotes the standard deviation of the score observed. The accuracy of the structure is depicted in the order BETTER>INSIDE>WORSE each parameter

Table 8: Summary of side chain parameters for the structure generated by I-TASSER-MODEL

Stereochemical parameter	No of data pts	Parameter value	Comparison typical value	Value band width	No. of band widths from mean	Interpretation
Chi-1 gauche minus st dev	61	21.4	13.6	6.5	1.2	WORSE
Chi-1 trans st dev	163	20.0	15.3	5.3	0.9	Inside
Chi-1 gauche plus st dev	211	20.3	13.8	4.9	1.3	WORSE
Chi-1 pooled st dev	435	20.4	14.3	4.8	1.3	WORSE
Chi-2 trans st dev	98	16.4	17.7	5.0	-0.3	Inside

St. Dev. Denotes the standard deviation of the score observed. The accuracy of the structure is depicted in the order BETTER>INSIDE>WORSE each parameter

4.2 Molecular Docking Benchmark and Validation

For the molecular docking results to be more trusted, one needs to validate the method. In this thesis, the experimental data were compared with experimental docking data described in more detail in Chapter three section 3.2.3. The docking calculations were repeated four times to get reliable results. The data of experimental and that of docking calculation were presented in Fig. 15 and Table 9 and Table 3. The data provided a good correlation of 0.69, which confirmed the reproducibility of the docking experiments.

Table 9: Summary of Experimental and calculated binding energies (kcal mol⁻¹)

PDB ID	ΔG_{exp}	ΔG_{calc}
2ZC9	-35	-38.28 \pm 0.47
2ZDV	-31	-36.30 \pm 0.35
2ZFF	-32	-34.83 \pm 0.62
2ZFP	-31	-31.17 \pm 0.47
2ZHQ	-37	-38.28 \pm 0.96
2ZI2	-32	-34.20 \pm 0.46

Baum *et al.* (2009)

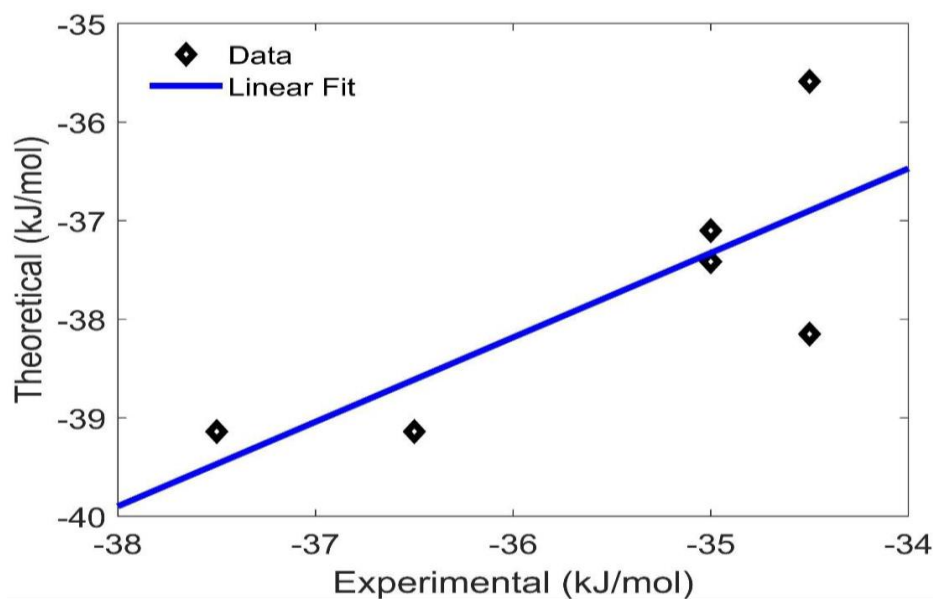


Figure 14: Correlation between theoretical and experimental binding-free energy (kJ/mol) ($r^2 = 0.69$)

The molecular docking was performed after the MD simulation, which allowed us to obtain different snapshots. The interaction of the ligand with the modelled as well as ensemble structures was investigated by molecular docking. It was observed that the binding free energy of the ligand to the modelled structure was -6.9 kcal/mol, which is equal to the inhibition constant of 8.39 μ M. The ensemble structures extracted from equilibrium MD simulation were used to investigate the effect of protein structure conformational changes when it interacts with ligands in binding-free

energy. As it can be seen clearly in Fig. 15 presented the variation in docking results to different ensemble structures. An ensemble structure with the best pose exhibited binding free energy of -7.2 kcal/mol, while that with poor pose exhibited -4.9 kcal/mol. These significant variations and differences in binding free energy among ensemble structure justify the importance of incorporating such conformations. The difference of 2.3 kcal/mol, which represents ≈ 500 -fold dissociation constant. From the docking results, one can conclude that the natural ligand strongly interacts with the linamarase enzyme (Paul *et al.*, 2021).

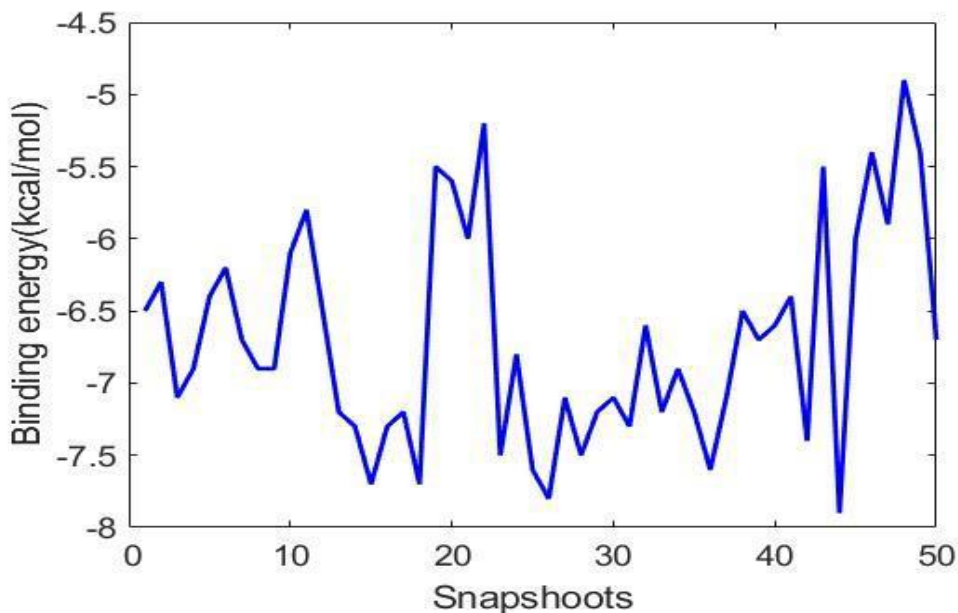


Figure 15: Binding free energy of different protein conformations

For identification of the active site and the necessary amino acids which are around and highly likely to be involved in interacting with ligand were identified by CASTp the (Tian *et al.*, 2018) the Table 10 represent different amino acids from modelled structure, ensemble structure also those observed to have more fluctuation.

Table 10: The amino acid residues at the active site of modelled, holo and ensemble structure

Residues at active site pocket	Holo structure (Residues interacting with ligand)	Ensemble structure (Residues interacting with ligand)	More fluctuating residues (Diff Apo–holo structures)
Gln48			Trh70
His152	His152		
Trp153	Trp153		
Asn197		Asn197	
Glu198			
Ser200			
Ala201	Ala201		
Val203			
Gly204			
Phe205	Phe205		
Asp208			
Asp209			
Val211			
Phe212		Phe212	
Asn222			
			Ser230
			Gly265
Phe269			Thr270
Phe271			
Tyr 273			Pro275
Ala289			
Phe292			
Met293			
Trp297			
			Met310
Tyr341	Tyr341		
Tyr344			
Ile349			
			Pro350
Tyr359			
Lys360			
Thr361			
Ser363			
Gly364			
Val365			
Asn366			
Ala367			
			Asn375
Ser384			
Trp385	Trp385	Trp385	
Glu 413	Glu413	Glu413	
			Pro425

Residues at active site pocket	Holo structure (Residues interacting with ligand)	Ensemble structure (Residues interacting with ligand)	More fluctuating residues (Diff Apo–holo structures)
			Leu450 Tyr460
Trp463	Trp463	Trp463	
Glu470	Glu470	Trp471	
Trp471			
Asn472		Asn472	
Ile473			
Phe479	Phe479	Phe479	
			Gly489

The binding interaction of linamarin with modelled and ensemble structure was investigated using protein plus and complemented by Discovery Studio visualizer. The analysis by visualization modelled structure and linamarin in protein plus shows the interaction of the following residues were shown in Fig. 16 (A) TYR341, GLN48, GLU13 and ASN197. For more analysis Discovery Studio Visualizer were used to identify the following residues interacting with linamarin; PHE479, TYR341, GLU470, TRP385, PHE212, PHE205, ALA201, TRP153, HIS152, GLU413 and TRP463, as can be seen in Fig. 16 (B).

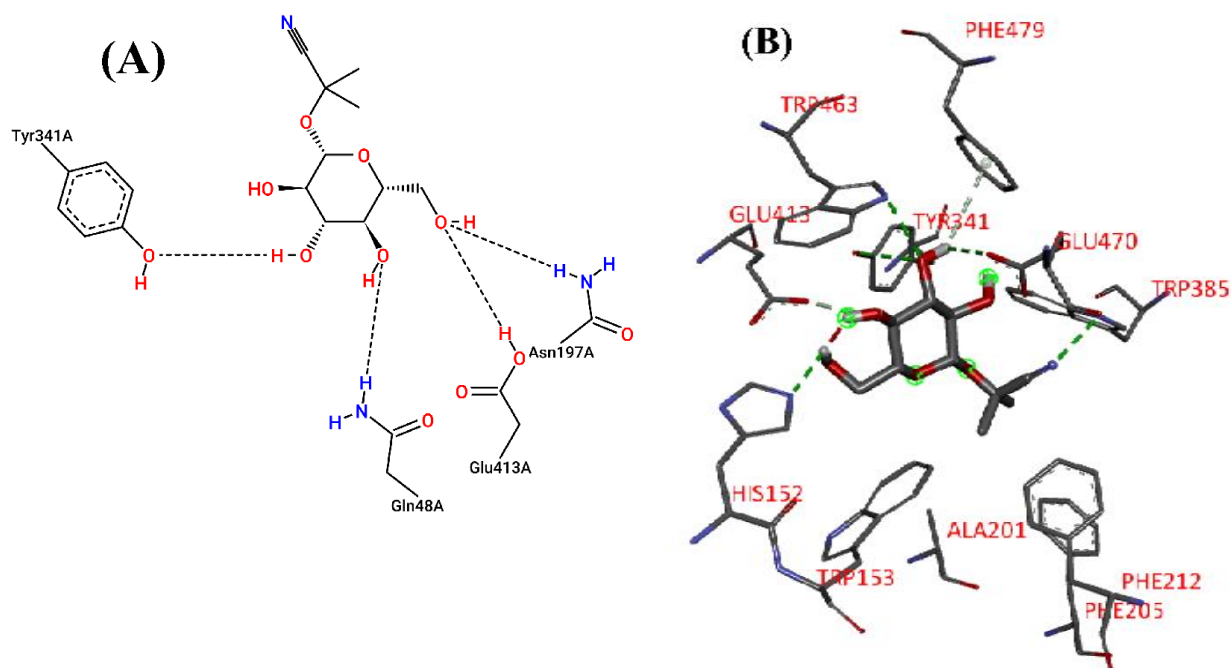


Figure 16: General interaction between linamarin and modelled structure using proteins plus (A) and Discovery Studio Visualizer (B)

To have more analysis, Discovery Studio Visualizer used to identify the hydrogen bond between linamarin and modelled structure as shown 3D in Fig.17 (A) and 2D in Fig. 17(B) (3.05, 3.38, 2.75, 2.60, 3.34) with GLU 470, TYR341, TRP463, HIS152 and TRP385, respectively.

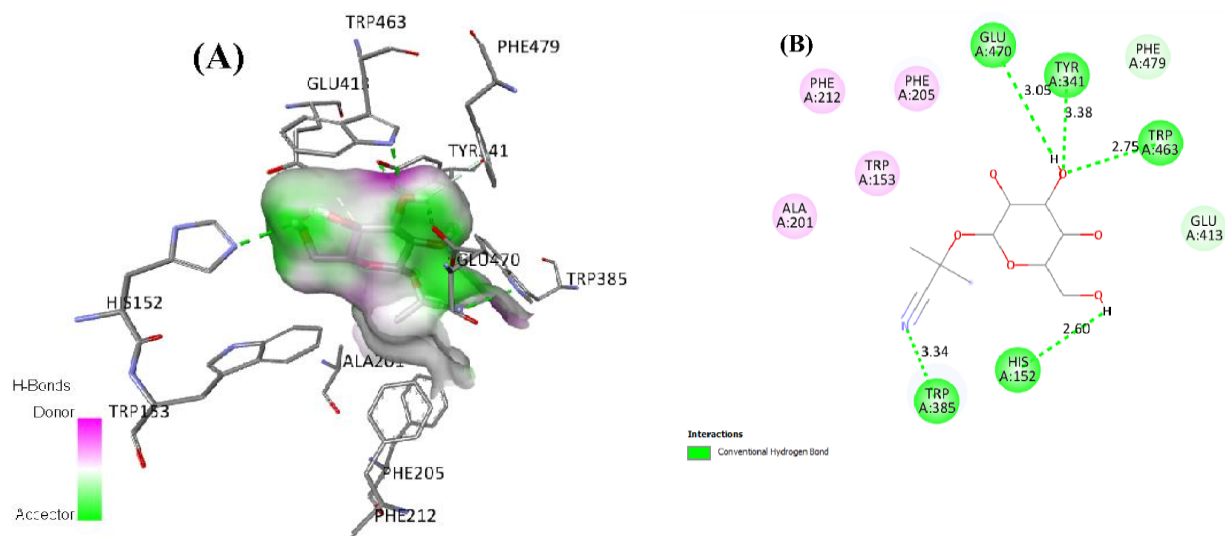


Figure 17: Linamarin interaction by a hydrogen bond with modelled structure (A) for 3D and (B) for 2D

The hydrophobic interaction was performed to identify the key residues and their respective distances.

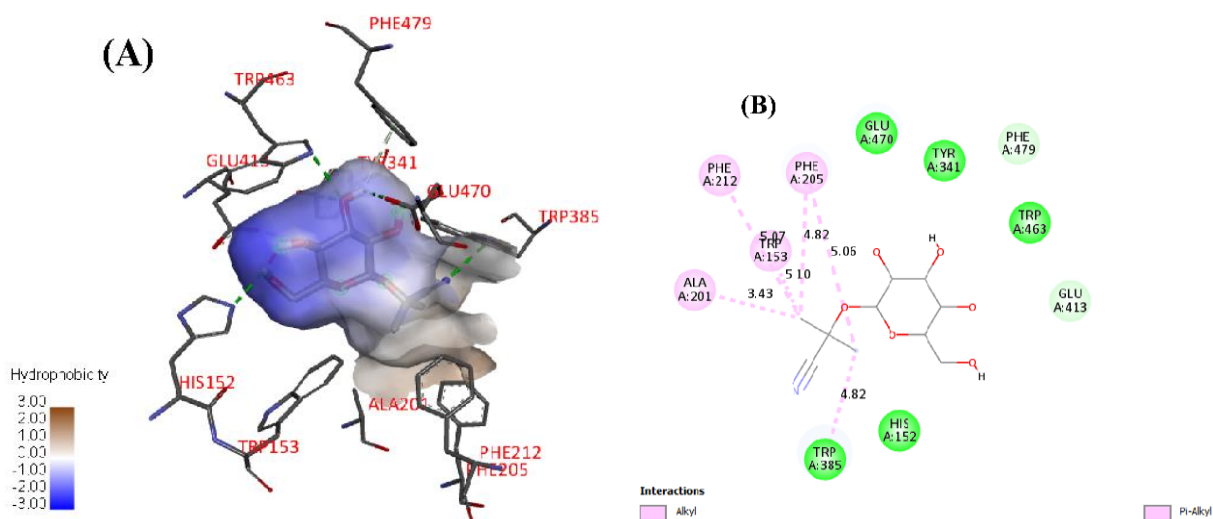


Figure 18: Hydrophobic interaction between linamarin and modelled structure 3D for (A) and 2D for (B)

The interaction between linamarin and ensemble structure was also generally determined using protein plus Fig. 19(A) and Discovery Studio Visualizer Fig. 19(B). The former identified GLU198, ASN197, GLU413, TRP385 and SER383, while the latter interacted with PHE479, TRP385, ASN472, TRP471, PHE212, ASN197, GLU413 and TRP463 residues.

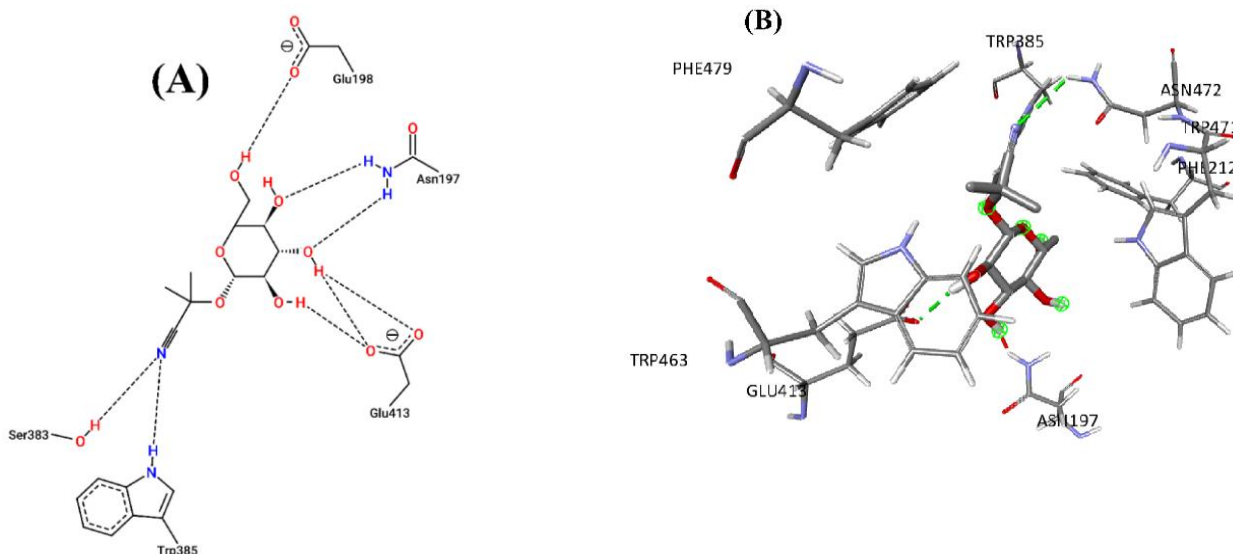


Figure 19: General interaction between linamarin and ensemble structure using proteins plus (A) and Discovery Studio Visualizer (B)

A more detailed analysis was performed using Discovery Studio Visualizer to determine the hydrogen bonding and hydrophobic interaction. It was observed that linamarin formed hydrogen bonding with TRP385, ASN472, GLU413 and unfavourable donor–donor ASN197 with the distance of (2.07, 2.96, 2.10 and 1.13 Å) respectively this can be observed in Fig. 20A for 3D and Fig. 20 for 2D (Paul *et al.*, 2021).

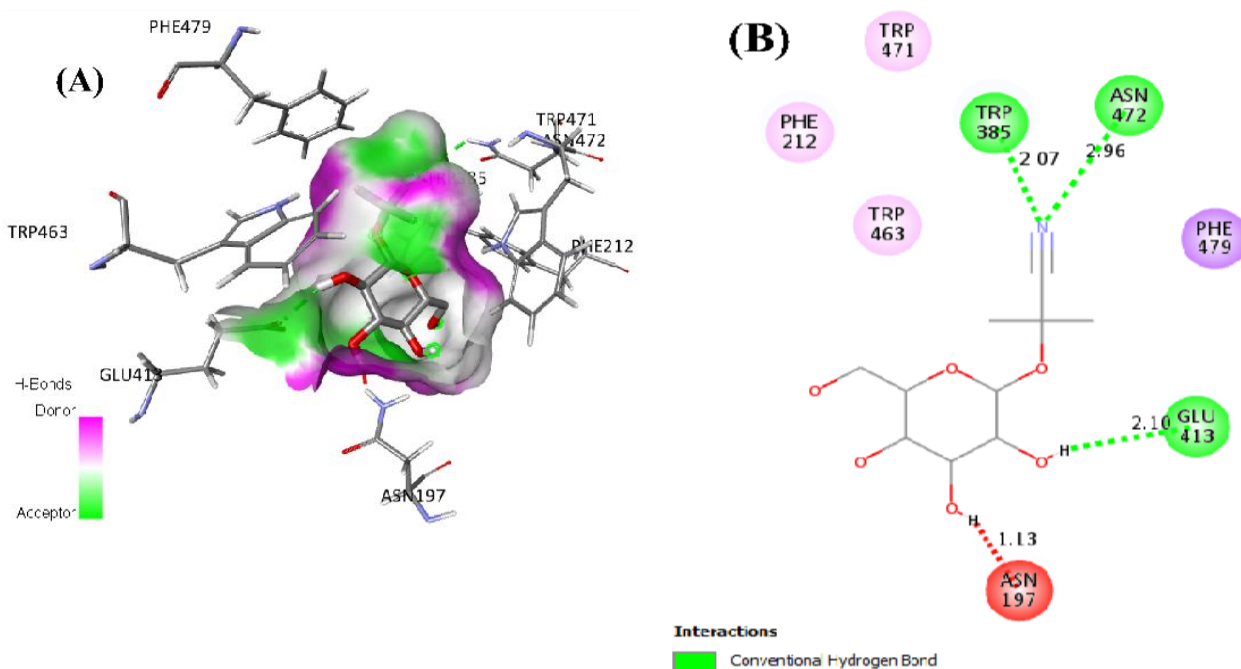


Figure 20: Linamarin interaction by a hydrogen bond with ensemble structure (A) for 3D and (B) for 2D

The analysis of hydrophobic interaction was performed the following bond observed: TRP463, PHE212, TRP471 and PHE479 with the distance of 5.52, 5.20, 5.07 and 3.58 Å, respectively, as it can be seen in Fig. 21 A for 3D and Fig. 21B for 2D.

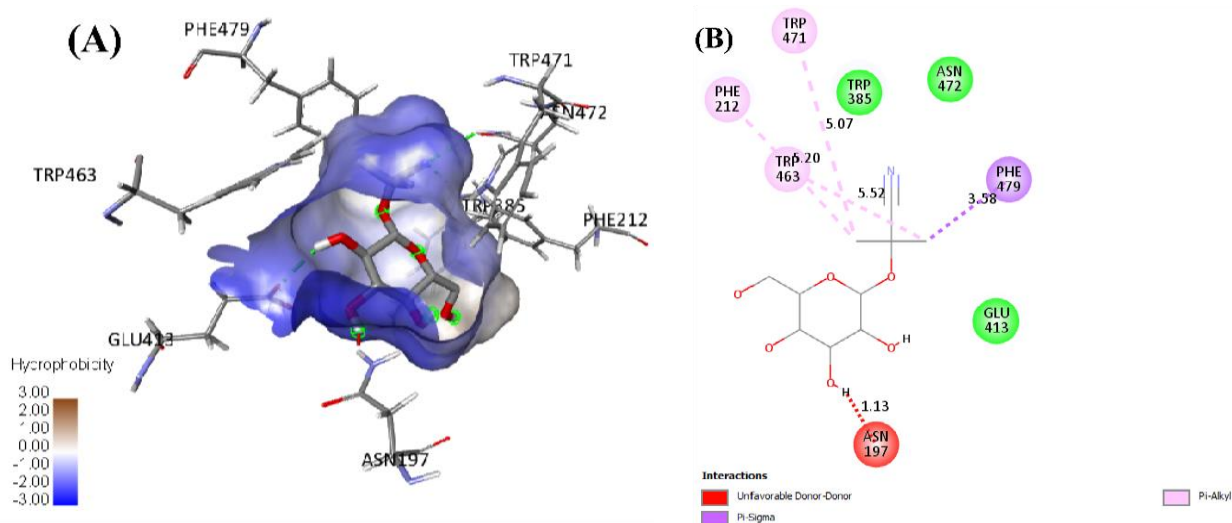


Figure 21: Hydrophobic interaction between linamarin and ensemble structure 3D for (A) and 2D for (B)

The Fig. 21 gives a detailed description of the binding modes of linamarin to the modelled and ensemble structure. The hydrogen bond appeared to be formed between the OH groups from the sugar moiety, while the hydrogen bond by the nitrogen (CN) was from the aglycone part of linamarin. Interestingly, the OH group from the sugar moiety of linamarin interacted with GLU413 residues from both structures (modelled and ensemble). On the other hand, CN from aglycone formed a hydrogen bond with two residues TRP385 and ASN472 but more interesting is that TRP385 was observed in modelled and ensemble structures. In both structures, CN was a hydrogen acceptor. Another critical property was the hydrophobic interaction observed for both structures; the main contribution for hydrophobic was from the aglycone part, particularly the methyl group (Paul *et al.*, 2021). It was observed that the methyl group interacts with common residue PHE212 that was observed from both structures.

These observations suggest that linamarin has a strong interaction with residues from both structures. To ascertain the stability of the protein–ligand complex, the interactions described above can not only be a factor to rely on. To have accurate information on the stability of ligand and fluctuation of the residues of the protein induced by ligand, MD simulation of both structures complexed with linamarin was performed.

4.3 Molecular Dynamics Simulation

The modelled structure and the two complexes were subjected to MD simulation to establish the stability and induced conformational changes over 200 ns. The RMSD provides essential insight into the structural stability of the modelled protein and the complexes. An average RMSD value of ≤ 3 Å indicates the system's stability (Wu *et al.*, 2007). The RMSD time-dependent and probability values of apo and the two complexes are shown in Fig. 21 a, b, complex 2, and apo structure equilibrated during the first 50 ns and remained stable with some fluctuations. Complex 1 equilibrated around 75 ns and had more fluctuations throughout the simulation. The apoprotein had a maximum RMSD value of 0.215 nm. However, as expected, it was interesting to note that complex 1 with the low binding-free energy of -6.9 kcal/mol showed a larger RMSD value with two maxima at 0.255 and 0.310 nm compared to complex 2 with the best binding-free energy of -7.2 kcal/mol, whose RMSD value shows the maxima at 0.19 nm. The lower RMSD value of complex 2 indicates the system's stability compared to complex 1. Based on the RMSD values for apoprotein and the two complexes 1 and 2, one can conclude that linamarin binding induces some conformational changes and fluctuation within 0.33 nm, suggesting stability of the complexes (Paul *et al.*, 2021).

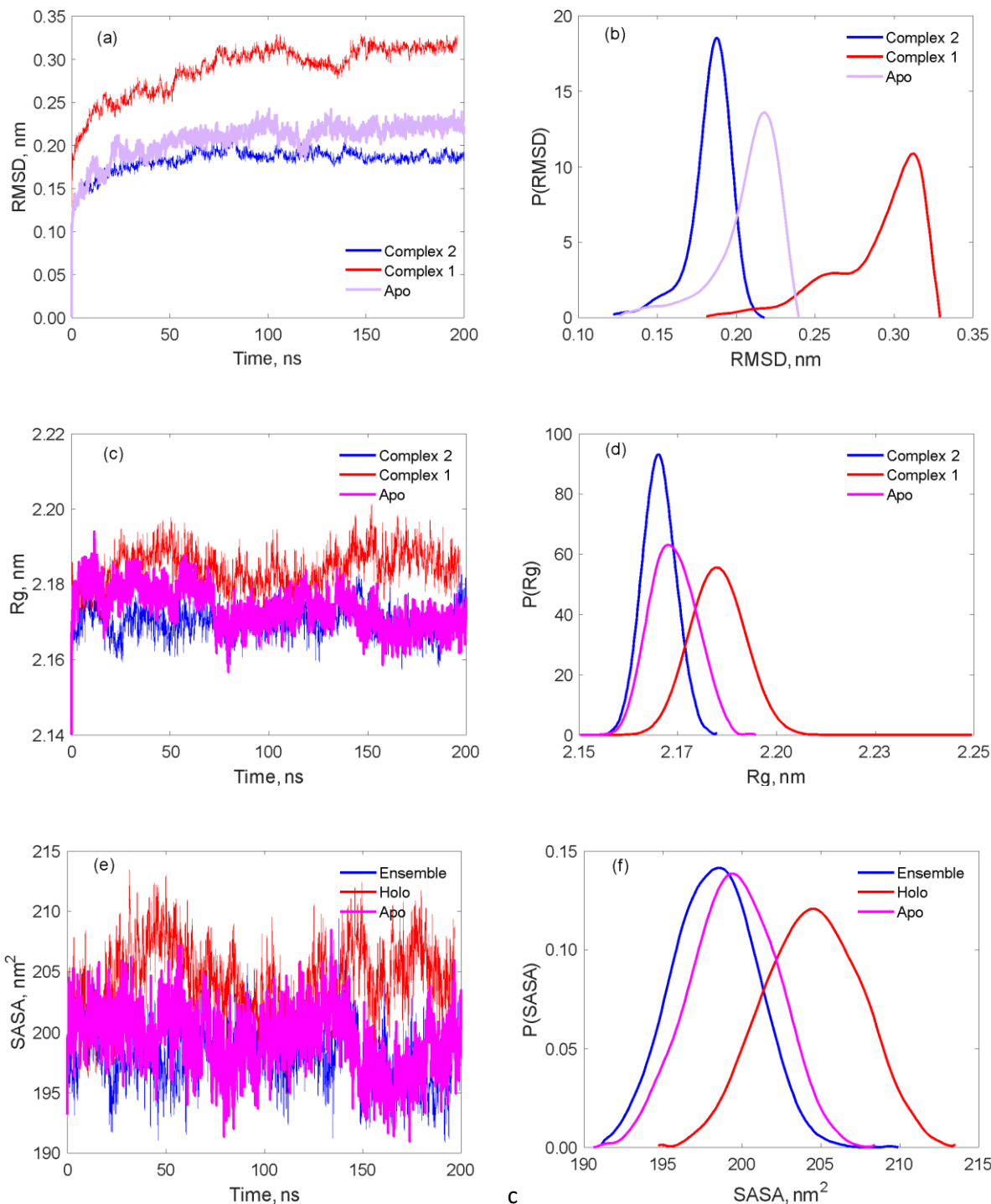


Figure 22: The probability values for RMSD, R_g and SASA of the apo protein and the complexes

The R_g was analysed to provide information on the protein compactness and size before and after ligand bound. As observed for the RMSD value, the R_g value follows the same trend; the R_g value for the apoprotein shows a maximum value at 2.17 nm; however, the complex 1 with binding-free energy of -6.9 kcal/mol indicates an increase in R_g value 2.18 nm while the complex 2 shows the maximum R_g value of 2.165 nm (Fig. 22 c and d). A similar observation is made for SASA, where

the complex with best binding–free energy showed a lower SASA value with a maximum at 196 nm² compared to complex 1 with the least binding energy whose maximum SASA value is 205 nm² (Fig. 22 e and f) (Paul *et al.*, 2021).

The induced conformational and plasticity of the protein amino acids was further investigated by computing the RMSF difference between the bound complex1 and free protein. Figure 23 shows that the binding of the protein–induced some noticeable residue fluctuations in some regions. The observed fluctuations are 64 – 84, 270 – 300 and 380 – 385 amino acids that form the active site. The observed fluctuations at the active site suggested that the ligand interacted with the residues and, to some extent, caused fluctuations. Other protein regions are observed to have small RMSF difference, suggesting that residues that were least involved in the interaction had little fluctuation (Paul *et al.*, 2021).

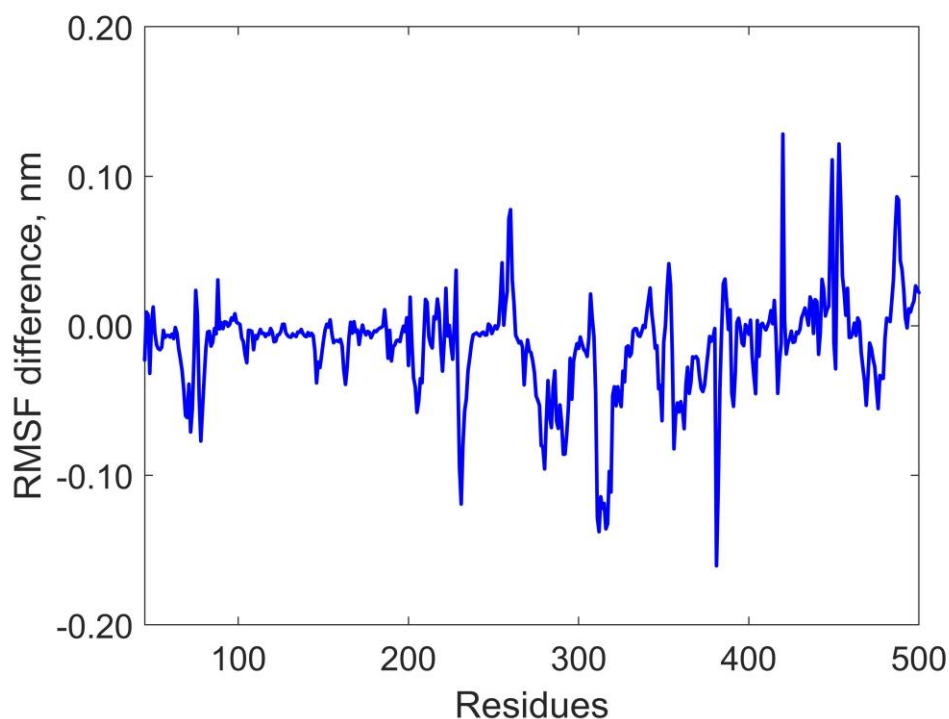


Figure 23: Root Mean Square Fluctuation difference between apo and complex 1

4.4 Molecular Mechanics Poisson–Boltzmann Surface Area Binding–free Energy Analysis

To better understand the amino acid contribution to the binding free energy of complex 1 and 2, a binding–free energy calculation based on MM/PBSA was performed. Table 11 clearly shows that complex 1 has lower binding–free energy compared to complex 2. However, the MM/PBSA results show higher binding–free energy values than docking results but have the same trend as

observed from docking calculations. The difference in results for the two complex structures is explained by the differences in vdW, electrostatic and polar energy contributions. In complex 1 the electrostatic energy terms strongly contributed to the interaction between linamarin and linamarase. However, to some extent, the vdW energy terms contributed to the interaction, while the polar energy terms were observed to oppose the interaction (Paul *et al.*, 2021).

Contrary to complex 1, in complex 2, the vdW energy terms highly contributed to the interaction, followed by electrostatic energy terms (Table 11). Furthermore, in complex 2 the opposing effect of polar energy terms is least observed compared to complex 1. Interestingly, in all complexes, the nonpolar (SASA) energy terms were observed to contribute equally (Table 11). The observed binding–free energy in complex 2 is nearly 2–folds higher than complex 1. Such difference is attributed to the structural changes in the two complexes, where the ensemble structure from equilibrium MD was observed to have a favourable interaction with linamarin. This observation is expected as the protein is not static but flexible with linamarin bound (Paul *et al.*, 2021).

Table 11: The binding–free energy (kJ/mol) for the two complexes calculated by the MM/PBSA method

Complexes	E_{vdW}	E_{elect}	E_{polar}	$E_{\text{non-polar}}$	$\Delta G_{\text{binding}}$
Complex 1	–78.489±25.149	–83.224±36.427	86.336±27.160	–14.026±3.896	–89.403±37.226
Complex 2	–131.361±12.676	–79.703±28.379	69.133±9.559	–14.659±0.594	–156.590±16.707

The observed differences in binding–free energy further motivated us to explore the effects of protein structural changes and residue contributions to binding–free energy. As shown in Fig. 24, linamarin adopted some orientation and/or conformational changes and interacted with different amino acids at the pocket. Interestingly, the interaction of linamarin with residues His152, Phe212 and Glu413 was observed in all complexes, although with different energetic contributions. Surprisingly, in complex 1, residue Glu413 is observed to have an opposing contribution effect to the binding–free energy by contributing nearly 5 kJ/mol, while in complex 2 it contributes the opposite energy i.e –5 kJ/mol. In addition, the contribution of His152 in complex 2 is nearly 2–folds higher than that in complex 1. Thus, one can conclude that a structural adaptation of the equilibrium MD structure has a role to stabilize the observed interaction in the linamarase linamarin complex (Paul *et al.*, 2021).

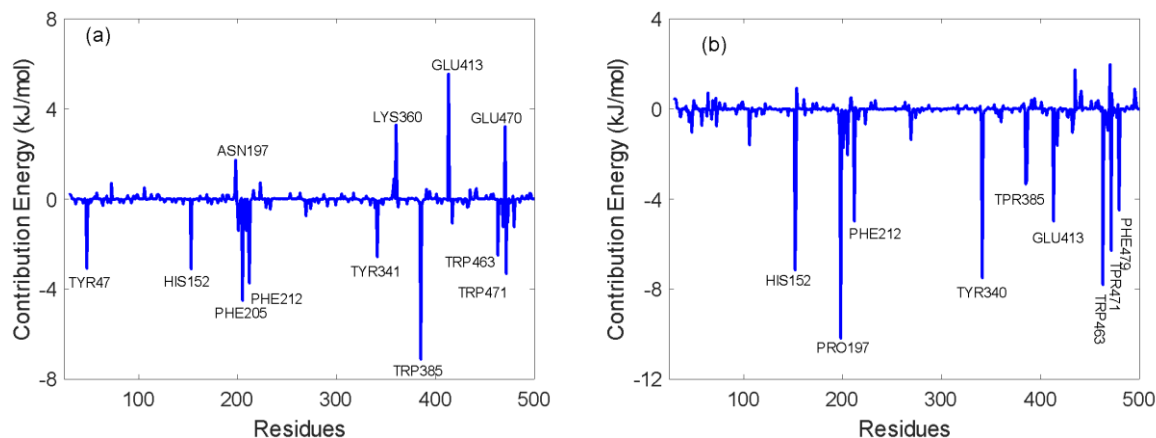


Figure 24: The energy contribution on protein–ligand binding (a) complex 1 and (b) complex 2

4.5 Stability and Conformational Preference of Linamarin in Different Solvents

This section presents and discusses in detail the effects of different solvents on the stability, conformational preferences and properties of linamarin.

4.5.1 Pair Distribution Function of Polar Groups Around Linamarin and Different Solvents

Gaining insight into the properties of linamarin in different solvents is of great significance in the pharmaceutical industry as solvents have drastic effects on hydration and solubility of molecules. In this thesis, to understand the influence of solvents on the properties of linamarin, first, the pair distribution function (PDF) was calculated to yield information on the hydration shell with vicinal solvents. The PDF further provided insights on the understanding of the number of hydration solvents as well as the possibility of hydrogen bond formation. As previously reported in literature, in this thesis the criterion for hydrogen bonding included the presence of donor atom (an atom with a hydrogen atom bonded to it) and acceptor (electronegative atom) within a cutoff distance of 3.5 Å and cutoff angle less than 30° between donor–acceptor atoms (D–H–A). The thesis begins by exploring the number of solvents interacting with hydrophilic atoms with their respective distances by calculating the radial distribution function (RDF) between the selected atoms from Fig. 25 and the solvents.

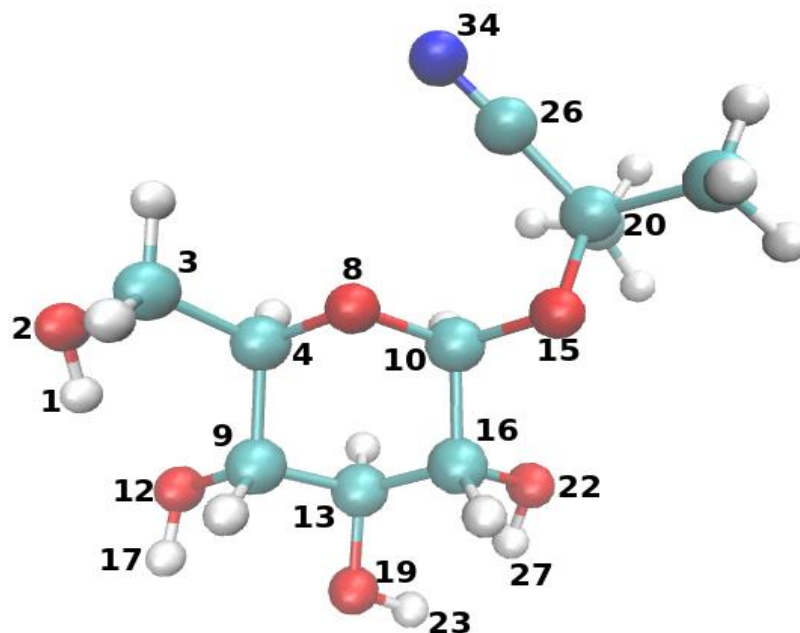


Figure 25: The chemical structure of linamarin used in this study, atoms are colour by colour: Oxygen in red, nitrogen in blue, carbon in light blue and hydrogen in white

Different maxima manifest the influence of solvents at different distances, the RDF in Fig. 26 shows a bimodal distribution of solvents at the vicinity of linamarin, suggesting the existence of two hydration shells with different size. Interestingly, MeOH was observed to interact more closely with linamarine than water and DMSO in all measured groups with a distance of ~ 0.18 nm. The lowering peak observed for MeOH in all groups suggests low density of MeOH surrounding linamarin as compared to the bulk solvents. The decrease in peak is attributed to hydrophobic groups which would tend to break the hydrogen bond networks and hence density decreases. On the other hand, DMSO and water show similar hydration effects to linamarin but with different density of solvent as depicted by the peaks maxima.

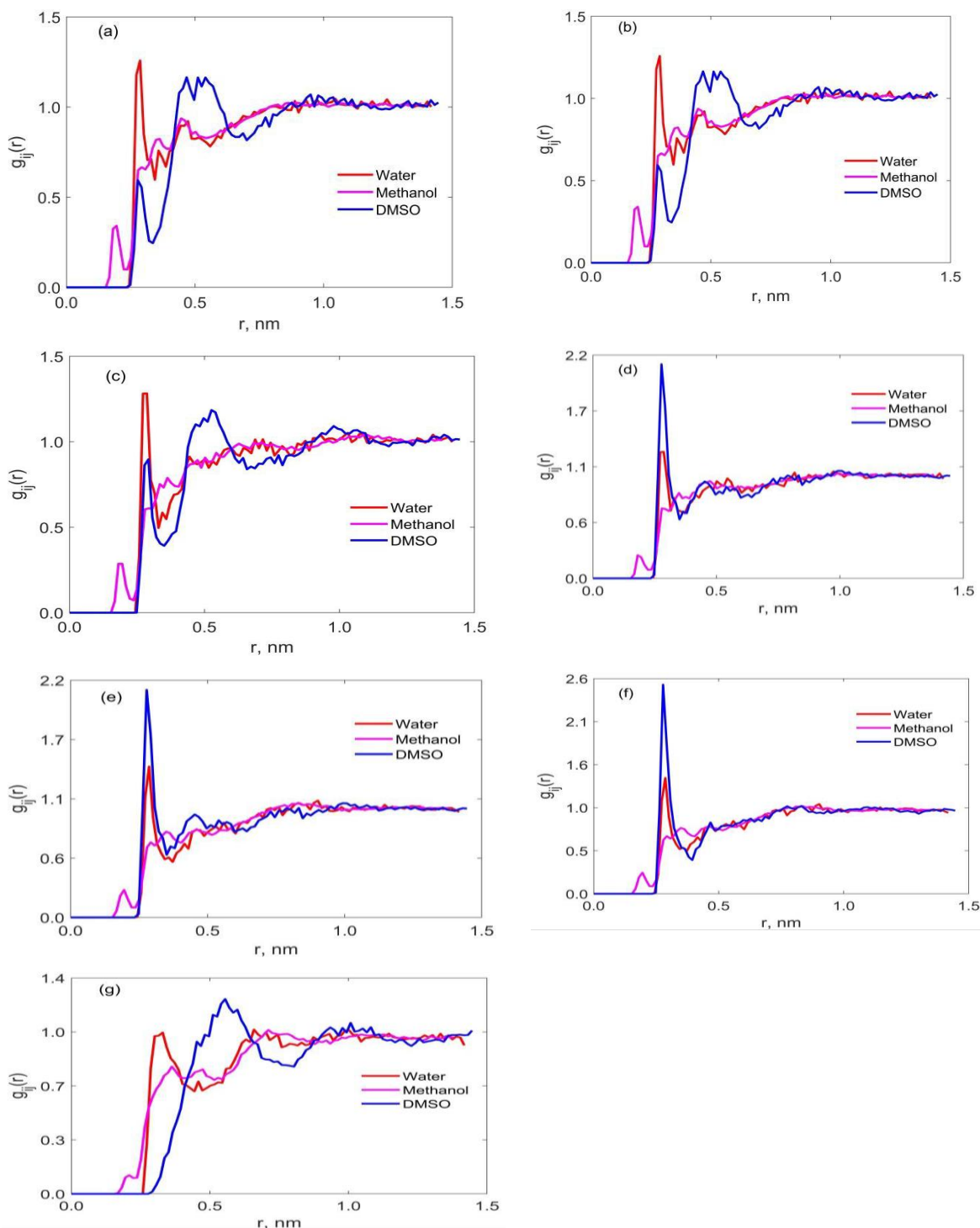


Figure 26: Pair distribution function of oxygen groups around linamarin molecule in water, MeOH and DMSO

From the above observation, it can be concluded that; linamarin as a small organic compound behaves differently in the solvents in question. The density of solvents around linamarin perturbed into two pronounced hydration layers. The distribution of solvent molecules across the linamarin region is varied. It was generally observed that the ether groups (O8 and O15) were dry, especially at the first layer depicted by the solvents. Both solvents in question were found populating with small density at the second layer of these groups. Another interesting observation was based on

the hydroxyl groups (O2, O12, O19 and O22) being linked with hydrogen atoms manifesting two distinct layers. The O19 and O22 groups depicted a very sharp first layer dominated by DMSO, followed by water and MeOH being least available, and this can be contributed by strong electronegativity potential of oxygen from DMSO to form a strong hydrogen bond with these group (DMSO being acceptor). The electronegative sulphur bonded to oxygen in DMSO creates a high potential for oxygen, which becomes a high potential for strong hydrogen bonding.

On the other hand, water has oxygen as a potential acceptor and two hydrogens that can be donated. So, it is probable to interact with these groups. Finally, MeOH was weakly available at this layer due to weak interaction with groups in questions. Furthermore, all solvents except for MeOH were observed at long distances surrounding N34.

Generally, the linamarin groups with hydrogen were observed to be more approached by both solvents but mainly DMSO, while those without hydrogen seem to be dry, then solvent molecules tend to bulk instead of approaching the linamarin molecules.

From the presentation and description from Fig .26, it enabled to determine the coordination numbers which indicates the solvents molecules found in the first coordination sphere, and this is given by using Equation 18:

$$n_i(r_{min}) = 4\pi\rho_j \int_0^{r_{min}} g_{ij}(r)r^2 dr \quad (18)$$

Whereby $n(r_{min})$ represents the quality of oxygen from solvents found within the first coordination peak (shell) of their oxygen from linamarin from RDF description. The ρ_j is the number of density of specific species while r_{min} refers to the pair distribution function at a minimum distance. The Table 12 helps to provide evidence that water molecules interacted with almost all groups of linamarin at an average distance of 0.3 nm like DMSO. Nevertheless, DMSO was observed to surround at closer proximity with high density at O19 and O22 only, but its average indicates higher distance with other groups. Thus, the hydroxyl groups (O19 and O22) significantly influence the solubility of linamarin, particularly in DMSO and water, while very little in MeOH.

Table 12: Summarizes the arrangement of the pair distribution functions and the first coordination numbers

Linamarin	r_{\min} , nm	$N(r_{\min})$
O2–OW	0.29	1.40
O8–OW	0.3	0.60
O12–OW	0.28	1.09
O15–OW	0.35	1.78
O19–OW	0.28	1.05
O22–OW	0.29	1.32
N34–OW	0.33	2.33
O2–OM	0.19	0.45
O8–OM	0.22	0.65
O12–OM	0.19	1.01
O15–OM	0.22	0.72
O19–OM	0.19	0.84
O22–OM	0.19	1.43
N34–OM	0.21	1.63
O2–OD	0.29	0.47
O8–OD	0.46	1.20
O12–OD	0.29	0.45
O15–OD	0.33	0.73
O19–OD	0.29	1.05
O22–OD	0.29	0.91
N34–OD	—	—

The abbreviation of OW, OM and OD corresponds to their respective water, methanol and DMSO, respectively

4.5.2 Distributions of Dihedral Angles

Different dihedral angles were investigated to understand better the origin of the observed conformational and orientational preferences of linamarin in a vacuum and different solvents; water, MeOH and DMSO. The definitions of different groups considered for the dihedral angle calculations are presented in Table 13.

Table 13: Naming of the dihedral angles of linamarin molecules

Abbreviations	Dihedrals	Distances
1–Dihedral	1–2–3–4	1–7 and 1–12
2–Dihedral	17–12–9–4	17–19 and 17–2
3–Dihedral	23–19–13–9	23–12 and 23–22
4–Dihedral	27–22–16–13	27–19 and 27–15
5–Dihedral	10–15–20–26	

The numbers were obtained from Fig. 25 linamarin structure numbered according to topology file parameter. Figure 27(a) shows the dihedral angle distribution in both solvents is characterized by the probability distribution of ϕ with maxima at $\phi = 40^\circ$ in vacuum and DMSO and 48° , 50° in MeOH and water, respectively suggesting a presence of gauche conformation. Analysis of the dihedral angle at Fig. 25 (b) revealed an orientation at $\phi = \pm 180$ which suggests the presence of

trans-conformation. As expected, visual inspection of the snapshot Fig. 25 (b) revealed that H17 pointed away from O2. Furthermore, the orientation preference of H23 towards O22 (Fig. 27(c)) was dominant and showed the value of $\phi = \pm 180^\circ$ suggesting the trans-conformation. Generally, as observed in Fig. 27, the orientation of different groups is solvent dependent and the effects of solvent are manifested by different maximums most populating at $\phi = \pm 180^\circ$. In both polar protic and aprotic solvents linamarin exists in trans-conformation. However, in the gas phase, linamarin presents a noticeable difference in conformation preferences compared to when in solvents.

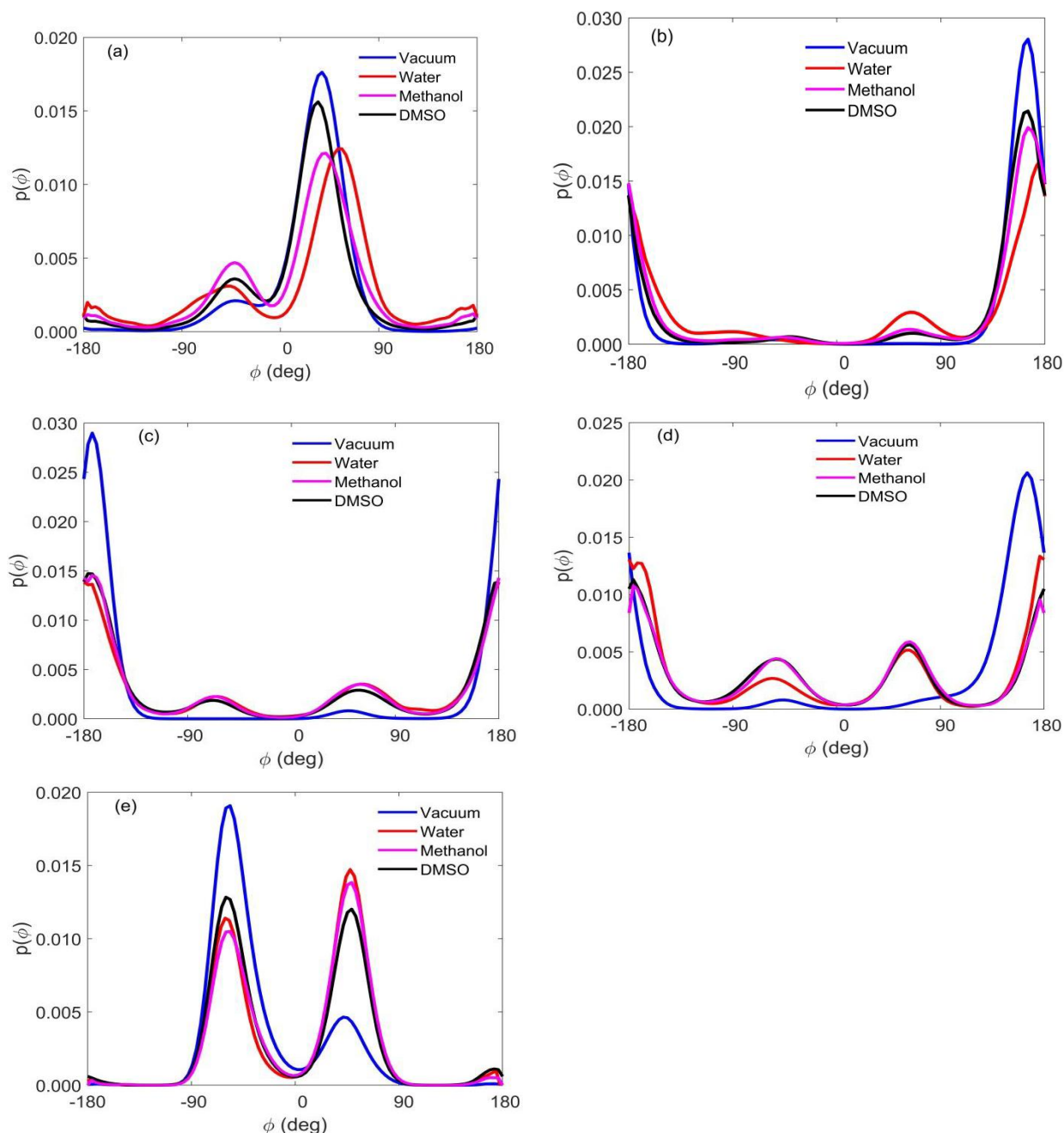


Figure 27: Distribution of dihedral angles of linamarin molecule in a vacuum and different solvents (a) 1-Dihedral, (b) 2-Dihedral, (c) 3-Dihedral, (d) 4-Dihedral and (e) 5-Dihedral

4.5.3 Distance and Angles Distribution

This section describes the angles distribution and the distance of atoms which were the critical points for considering such angles. For linamarin, the sugar moiety is surrounded by hydroxyl groups, in which hydrogen can either form intramolecular or intermolecular hydrogen bonding with their neighbouring oxygen or oxygen from solvents, respectively. The distance between specific groups was measured to understand the influence of solvents polarity. As observed for the dihedral angles, the influence of solvents was manifested by different maxima. For example, the distance measured between H1 and the neighbouring O8 & O12 presented in Fig. 28 (a) and (b) respectively both depicted two dominant maxima; this proves H1 points towards O8 and O12. The interaction observed aligns with the 1_dihedral angle discussed above; both solvents were observed to affect the interaction between these atoms in question.

The distance between H17 and the near oxygen atoms (O19 and O12) (Fig. 28 (c) and (d)), respectively. It can be observed that interaction between H17 and O19 is dominated by a single maximum at $d = 2.3$ nm, while interaction with O2 was found at a long distance with multiple maxima. At both distances, solvents affected the interaction between H17 and the neighbouring oxygen atoms. The interaction behaviour of H17 proved the existence of 2_dihedral angles.

Similarly, the distance measured for the rest of hydrogen atoms (H23, H27) with the neighbouring oxygen atoms (Fig. 28 e, f, g and h) were dominated by a single maximum in a vacuum, but when solute introduced into solvents, the probability was decreased, and the more than one maximum was introduced. The end to end distance was observed to decrease when linamarin encounters the solvents (Fig. 28 (i)). Although there is a general shift in the maximum probability density, the effect of solvents is still apparent for all the studied solvents. In general, the conformation orientation of linamarin is solvent dependent, as suggested by different maximums when linamarin was dissolved in solvents.

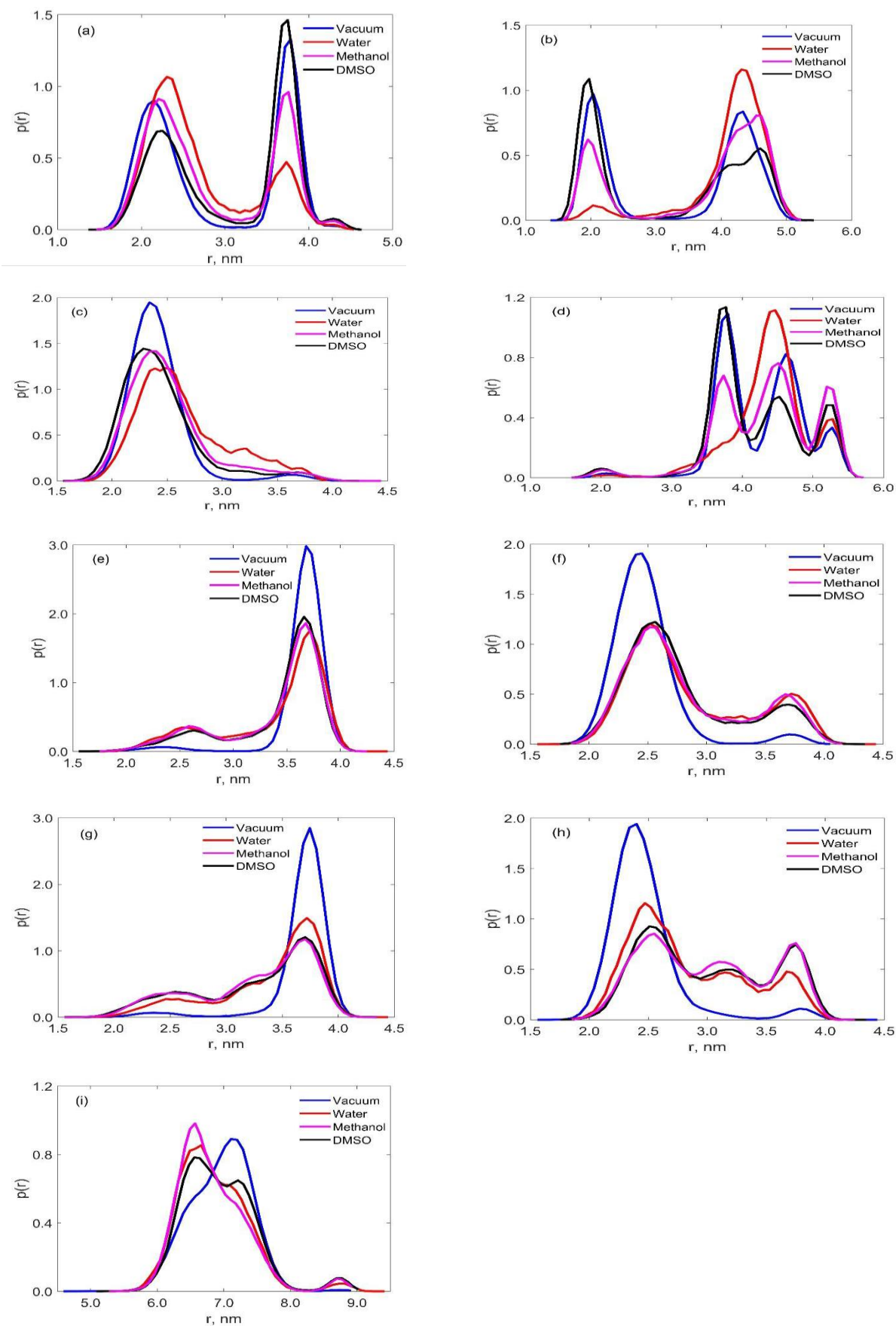


Figure 28: The probability distribution of the distance of the selected atoms and end to end for linamarin in a vacuum and different solvents

4.5.4 Hydrogen Bonds Analysis

Hydrogen bonds play an essential role in the biomolecular system by stabilizing the interactions. The existence of hydrogen bonds between linamarin and different vicinity solvent was investigated. The identified atoms which can participate in hydrogen bonding (HBs) from linamarin are O2, O8, O12, O15, O19, O22 and N34; these groups interact directly with the hydrogen of different solvents. Figure 29 shows the probability distribution of the total number of hydrogen bonds formed between different solvent molecules and single linamarin. To determine the total HBs between linamarin and different solvents (water, DMSO, MeOH); the two most crucial criterium were considered, the distance between donor–acceptor to be less than 3.5 Å while the angle between D–H–A should be equal to 30°.

It was interesting to see that oxygen of DMSO can only participate in hydrogen bonding as an acceptor exclusively, while oxygen of water and MeOH molecules can participate in both accepting and donating a proton. It was worth noticing that when the oxygen of both solvents is involved as a hydrogen acceptor, the number of HBs were closer to each other (Fig. 29 b). The maxima of the HBs were slightly different; water and MeOH was almost the same, while DMSO was slightly higher. The DMSO attained the maximum probability at $n_{hb} = 1.92$, water at $n_{hb} = 2.48$ and MeOH at $n_{hb} = 2.10$.

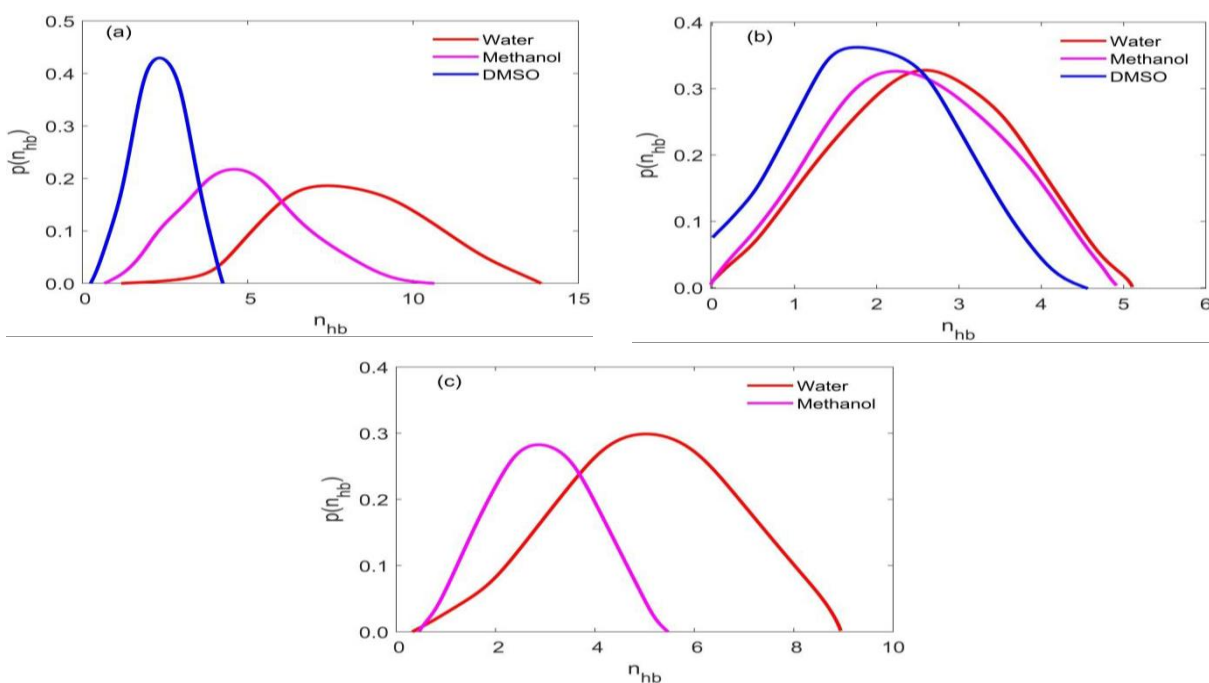


Figure 29: The probability distributions of the number of HBs between linamarin and different solvents

When the oxygen of solvent is considered donors, DMSO was excluded because its oxygen does not have hydrogen. There was a high distribution of the number of hydrogen bonds, and water had a high shift of HBs distribution to the right-hand side with a considerable dispersion ranging from n_{hb} 0.5 to around 9 with the maximum probability at 5.9 (Fig. 29 c). The distribution number of HBs for MeOH was observed to attain the maximum probability at 3. This distribution is attributed to the fact that water has a potential of two polarized hydrogen ions to be involved in HBs while MeOH has only one (Souda *et al.*, 2003).

It was also worth determining the total contribution of solvents towards hydrogen bonding with linamarin molecules. The general probability distribution of the total HBs without considering the solvents was an acceptor or donor. It was observed in Fig. 29 a and Fig. 30 water was the most probable with a maximum at around $n_{hb} = 8.5$, followed by MeOH with the maximum probability at $n_{hb} = 4.5$, while DMSO had the highest probability at $n_{hb} = 2.0$.

The results of a detailed analysis of hydrogen bonding between solvents and linamarin molecules were summarized in Table 14 and Fig. 31. It can be observed that the average number of HBs contributed by water was 8.5. Thus, the contribution observed is almost twice that contributed by MeOH; moreover, the contribution of MeOH was also almost twice that of DMSO. These differences are contributed by the fact that water and MeOH act as acceptor and donor; moreover, water can donate two hydrogen atoms while MeOH has only one hydrogen atom. On the other hand, DMSO can only provide oxygen which is an acceptor (Souda *et al.*, 2003).

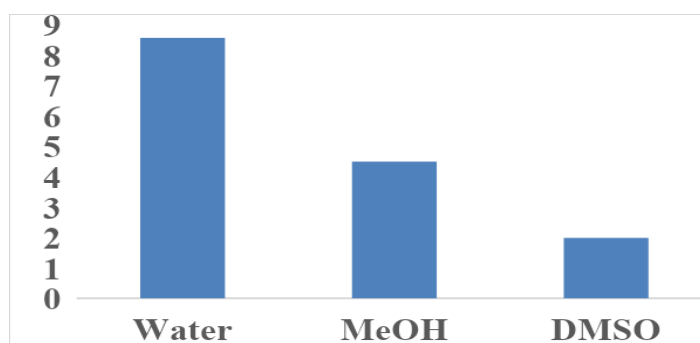


Figure 30: Contribution of hydrogen bonds between linamarin and different solvents

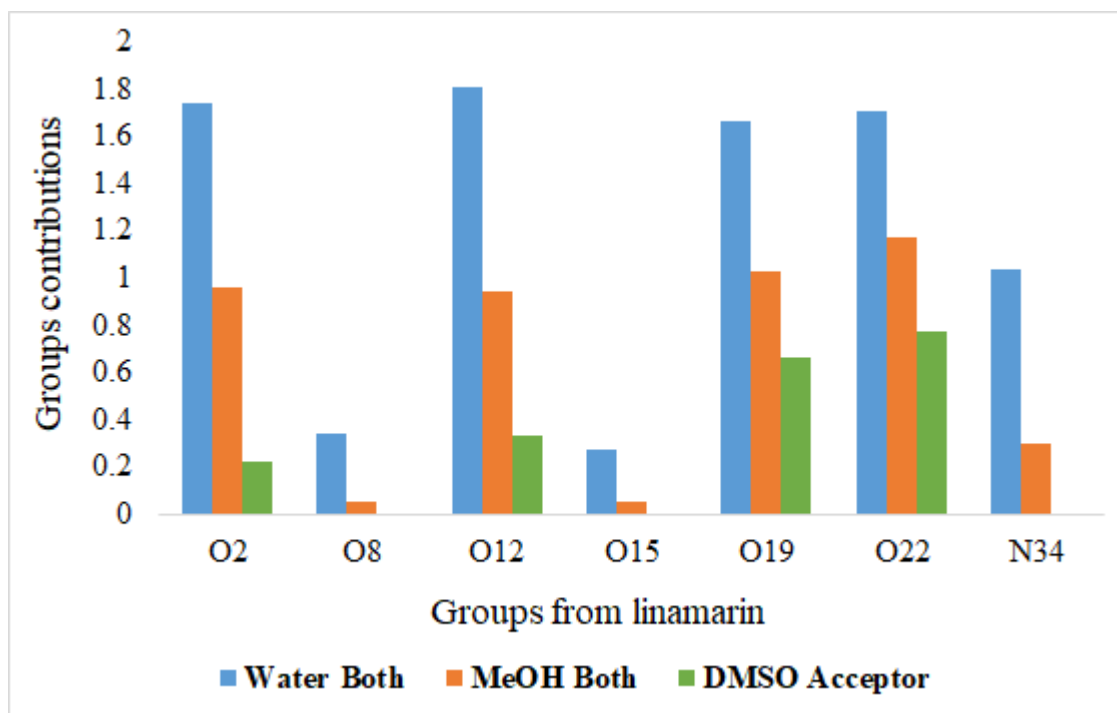


Figure 31: Contribution of different groups of linamarin during hydrogen bonding and different solvents

Table 14: The average number of HBs contributed by each group of linamarin in different solvents

HBs	Water			MeOH			DMSO Acceptor
	Acceptor	Donor	Both	Acceptor	Donor	Both	
O2	0.544	1.198	1.742	0.286	0.674	0.960	0.219
*O8	—	0.339	0.339	—	0.055	0.055	—
O12	0.639	1.170	1.809	0.421	0.520	0.941	0.328
*O15	—	0.271	0.271	—	0.051	0.051	—
O19	0.681	0.980	1.661	0.616	0.408	1.024	0.661
O22	0.709	1.000	1.709	0.702	0.470	1.172	0.776
*N34	—	1.037	1.037	—	0.294	0.294	—
SUM	2.573	5.995	8.568	2.025	2.472	4.497	1.984

Moreover, the contribution of specific groups from linamarin to hydrogen bond formation were analysed. The significant contributors from linamarin were identified from Fig. 25 with O2, O8, O12, O15, O19, O22 and N34 atoms. These atoms were involved in hydrogen bonding with both solvents. It was worth noticing that these groups participated as acceptors and donors (O2, O12, O19 and O22) and contributed to both solvents and groups who participated as acceptors only. More interestingly, all groups displayed higher participation in HBs when interacting with water molecules, followed by MeOH and DMSO being the least (Fig. 31 and Table 14).

Moreover, from Fig. 31 and Table 14, O12 has the highest contribution of all groups while O22 being the least. Generally, these differences of solvents participation are contributed by water being able to be acceptor and donor; however, it has two potential hydrogen atoms when acting as

a donor. On the other hand, MeOH also acts as an acceptor and donor but can donate only one hydrogen and finally, DMSO can form a single bond using its oxygen as an acceptor.

CHAPTER FIVE

CONCLUSION AND RECOMMENDATIONS

5.1 Conclusion

Homology modeling was performed to determine the (3-D) structure of cassava linamarase, which is not yet structure determined experimentally. The binding mechanism of linamarin to the modelled structure of linamarase was investigated by molecular docking. The structural stability of the modelled and protein–ligand complex was ascertained by molecular dynamic simulation. The MM/PBSA approach was used to rescore the docking results and detailed the amino acid contribution to the binding free energy.

The SWISS–MODEL was proved to be the best tool in providing a high-quality linamarase model from homology modelling. The MD simulation provided important information on the modelled protein, specifically on conformational stability. After MD simulation of modelled structure allowed us to perform a relaxed complex scheme to allow flexible docking. From MD simulation, the ensembles obtained were docked with the ligand (linamarin). From molecular docking of ensembles, it was observed that one of the ensembles revealed to have higher binding energy with linamarin compared to modelled structure, so this proves the importance of including ensembles in docking. The performed MD simulation of the complex observed to have strong interaction with the ligand from molecular docking. From MD simulation of complexes, both proved to be stable during simulation. From MD and MM/PBSA, it was revealed that both protein–ligand complexes had strong interaction via the following residues: Try47, Phe205, Phe212, Trp385, Trp463, Trp471, Trp153, Ala201, Phe205, Trp385, Trp463 and Phe479, which contributes by hydrophobic interaction with linamarin and retaining the stability of the complexes.

Moreover, residues: His152, Pro197, and Try340 have the highest contribution to binding free energy. The MD analysis provided more insight into the stability of the complexes. The RMSD of the complexes indicated some conformational changes and fluctuation within 0.33 nm, suggesting the complexes were stable.

Furthermore, the Rg provided information on the compactness and size before and after ligand introduced. The value of Rg value 2.165 nm suggests the complex retained the compactness. When the conformational plasticity of some amino acids modelled linamarase by RMSF investigated, it was observed that amino acids around 64–84, 270–300, and 380–385 found in the active site had

some fluctuation. The fluctuation of the listed amino acids suggests strong interaction with linamarin.

Conformational investigation of linamarin in a vacuum and different solvents provided better information on the properties of linamarin. Different polar regions of linamarin were identified to interact differently with solvents; this allowed us to determine the pair distribution function. Most of the polar atoms depicted two maxima, indicating a multi-layer solvation structure; the typical peak at 0.186 nm aligned with the pure water's intermolecular hydrogen bonding (0.188 nm). The polar atoms with hydrogen were observed to have a strong interaction with DMSO, while those that do not have hydrogen observed to be dry since solvents molecules prefer to bulk instead of approaching the groups of linamarin. The dihedral angle investigation was based on the polar region of linamarin to see how the hydrogen atom changes orientation when subjected to different solvents. It was observed that most of the dihedral angles were dominated by trans-conformation supported by the maxima at $\phi = \pm 180^\circ$. By looking at the dihedrals, it can be concluded that solvents have strongly influenced the conformational changes with no doubt. The distance and angles supported the finding; when the inter or intramolecular interaction was observed, it aligned in almost all observations. Finally, hydrogen analysis provided insight into how the polar region of linamarin, specifically the sugar moiety, has enormously contributed to interaction with the solvent. It can be concluded that water had a vital contribution to hydrogen bonding with linamarin. It was also proved that both solvents have almost equal capacity to work as a proton acceptor, but MeOH and water are the only solvents that work as proton donor and water being substantial donor compared to MeOH.

5.2 Recommendations

The computational approach has modelled the structure of linamarase and identified the conformational stability even when complexed with linamarin. The approach also identified the key residues which strongly interact with linamarin. The thesis further presents the conformational stability and properties of linamarin in a vacuum and different solvents. Despite the fascinating finding presented in this thesis, the following are the areas recommended for further study:

- (i) Experimental determination of linamarase structure either by x-ray crystallography, NMR or Cryo-electron microscopy (Cryo-EM).
- (ii) Since linamarin is a cyanogenic glucoside, its potential application as a drug is limited by cyanogenesis to untargeted cells, so computational and experimental study on drug carriers to deliver it to target cells should be done.

- (iii) A Quantum computational method is required on the reaction mechanism of linamarin, specifically on the release of toxic cyanide ions.
- (iv) An experimental study is required to extract and purify linamarase on a large enough scale to be applied directly during cassava processing, converting bound to free cyanide.

REFERENCES

- Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B., & Lindahl, E. (2015). GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, 1–2, 19–25. <https://doi.org/10.1016/j.softx.2015.06.001>
- Almeida, J. G., Preto, A. J., Koukos, P. I., Bonvin, A. M. J. J., & Moreira, I. S. (2017). Membrane proteins structures: A review on computational modeling tools. *Biochimica et Biophysica Acta: Biomembranes*, 1859(10), 2021–2039. <https://doi.org/10.1016/j.bbamem.2017.07.008>
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, 215(3), 403–410. [https://doi.org/10.1016/s0022-2836\(05\)80360-2](https://doi.org/10.1016/s0022-2836(05)80360-2)
- Amaro, R. E., Baron, R., & McCammon, J. A. (2008). An improved relaxed complex scheme for receptor flexibility in computer-aided drug design. *Journal of Computer-Aided Molecular Design*, 22(9), 693–705. <https://doi.org/10.1007/s10822-007-9159-2>
- Baum, B., Mohamed, M., Zayed, M., Gerlach, C., Heine, A., Hangauer, D., & Klebe, G. (2009). More than a Simple Lipophilic Contact: A Detailed Thermodynamic Analysis of Nonbasic Residues in the S1 Pocket of Thrombin. *Journal of Molecular Biology*, 390(1), 56–69. <https://doi.org/10.1016/j.jmb.2009.04.051>
- Beeman, D. (1976). Some multistep methods for use in molecular dynamics calculations. *Journal of Computational Physics*, 20(2), 130–139. [https://doi.org/10.1016/0021-9991\(76\)90059-0](https://doi.org/10.1016/0021-9991(76)90059-0)
- Benkert, P., Tosatto, S. C. E., & Schomburg, D. (2008). QMEAN: A comprehensive scoring function for model quality assessment. *Proteins: Structure, Function, and Bioinformatics*, 71(1), 261–277. <https://doi.org/10.1002/prot.21715>
- Berendsen, H. J. C., Postma, J. P. M., Van Gunsteren, W. F., DiNola, A., & Haak, J. R. (1984). Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics*, 81(8), 3684–3690. <https://doi.org/10.1063/1.448118>

- Berman, H., Henrick, K., & Nakamura, H. (2003). Announcing the worldwide Protein Data Bank. *Nature Structural Biology*, 10(12), 980. <https://doi.org/10.1038/nsb1203-980>
- Bhat, M. K., & Bhat, S. (1997). Cellulose degrading enzymes and their potential industrial applications. *Biotechnology Advances*, 15(3-4), 583-620. [https://doi.org/10.1016/s0734-9750\(97\)00006-2](https://doi.org/10.1016/s0734-9750(97)00006-2)
- Bhatia, Y., Mishra, S., & Bisaria, V. S. (2002). Microbial beta-glucosidases: Cloning, properties, and applications. *Critical Reviews in Biotechnology*, 22(4), 375-407. <https://doi.org/10.1080/07388550290789568>
- Bishop, A. O. T., Beer, T. A. P. D., & Joubert, F. (2008). Protein homology modelling and its use in South Africa. *South African Journal of Science*, 104, 2-6. http://www.scielo.org.za/scielo.php?script=sci_arttext&pid=S0038-23532008000100001&nrm=iso
- Böhm, H. J. (1992). LUDI: Rule-based automatic design of new substituents for enzyme inhibitor leads. *Journal of Computer-aided Molecular Design*, 6(6), 593-606. <https://doi.org/10.1007/BF00126217>
- Böhm, H. J. (1998). Prediction of binding constants of protein ligands: A fast method for the prioritization of hits obtained from de novo design or 3D database search programs. *Journal of Computer-aided Molecular Design*, 12(4), 309-323. <https://doi.org/10.1023/a:1007999920146>
- Bokanga, M., Ekanayake, I. J., Dixon, A. G., & Porto, M. C. M. (1994). *Genotype-environment interactions for cyanogenic potential in cassava. International Workshop on Cassava Safety*. <https://www.google.com>
- Burley, S. K., Berman, H. M., Bhikadiya, C., Bi, C., Chen, L., Di Costanzo, L., Christie, C., Dalenberg, K., Duarte, J. M., Dutta, S., Feng, Z., Ghosh, S., Goodsell, D. S., Green, R. K., Guranović, V., Guzenko, D., Hudson, B. P., Kalro, T., Liang, Y., Lowe, R.,... & Zardecki, C. (2018). RCSB Protein Data Bank: Biological macromolecular structures enabling research and education in fundamental biology, biomedicine, biotechnology and energy. *Nucleic Acids Research*, 47(1), 464-474. <https://doi.org/10.1093/nar/gky1004>
- Bussi, G., Donadio, D., & Parrinello, M. (2007). Canonical sampling through velocity rescaling. *The Journal of Chemical Physics*, 126(1), 014101. <https://doi.org/10.1063/1.2408420>

- Carlson, H. A. (2002). Protein flexibility and drug design: how to hit a moving target. *Current Opinion in Chemical Biology*, 6(4), 447–452. [https://doi.org/10.1016/s1367-5931\(02\)00341-1](https://doi.org/10.1016/s1367-5931(02)00341-1)
- Cavasotto, C. N., & Phatak, S. S. (2009). Homology modeling in drug discovery: Current trends and applications. *Drug Discovery Today*, 14(13–14), 676–683. <https://doi.org/10.1016/j.drudis.2009.04.006>
- Charifson, P. S., Corkery, J. J., Murcko, M. A., & Walters, W. P. (1999). Consensus scoring: A method for obtaining improved hit rates from docking databases of three-dimensional structures into proteins. *Journal of Medicinal Chemistry*, 42(25), 5100–5109. <https://doi.org/10.1021/jm990352k>
- Cheatham, T. E., Miller, J. L., Fox, T., Darden, T. A., & Kollman, P. A. (1995). Molecular Dynamics Simulations on Solvated Biomolecular Systems: The Particle Mesh Ewald Method Leads to Stable Trajectories of DNA, RNA, and Proteins. *Journal of the American Chemical Society*, 117(14), 4193–4194. <https://doi.org/10.1021/ja00119a045>
- Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., & Richardson, D. C. (2010). MolProbity: All-atom structure validation for macromolecular crystallography. *Acta Crystallographica. Section D, Biological Crystallography*, 66(Pt 1), 12–21. <https://doi.org/10.1107/s0907444909042073>
- Cheng, T., Li, Q., Zhou, Z., Wang, Y., & Bryant, S. H. (2012). Structure-based virtual screening for drug discovery: A problem: Centric review. *The American Association of Pharmaceutical Scientists Journal*, 14(1), 133–141. <https://doi.org/10.1208/s12248-012-9322-0>
- Chin, S. L., Lu, Q., Dane, E. L., Dominguez, L., McKnight, C. J., Straub, J. E., & Grinstaff, M. W. (2016). Combined Molecular Dynamics Simulations and Experimental Studies of the Structure and Dynamics of Poly-Amido-Saccharides. *Journal of the American Chemical Society*, 138(20), 6532–6540. <https://doi.org/10.1021/jacs.6b01837>
- Clark, K. P., & Ajay, R. S. (1995). Flexible ligand docking without parameter adjustment across four ligand-receptor complexes. *Journal of Computational Chemistry*, 16(10), 1210–1226. <https://doi.org/10.1002/jcc.540161004>

- Cohen, N. C. (2007). Structure-based drug design and the discovery of aliskiren (Tekturna): perseverance and creativity to overcome R & amp: D pipeline challenge. *Chemical Biology & Drug Design*, 70(6), 557–565. <https://doi.org/10.1111/j.1747-0285.2007.00599.x>
- Colovos, C., & Yeates, T. O. (1993). Verification of protein structures: patterns of non bonded atomic interactions. *Protein Science: A Publication of the Protein Society*, 2(9), 1511–1519. <https://doi.org/10.1002/pro.5560020916>
- Commission, S. P. T. (2000). *Exporting Fruit and Vegetables to New Zealand, A Guide for Pacific Island Producers*. Auckland, New Zealand. <https://www.google.com>
- Cooke, R. D., & Maduagwu, E. N. (1978). The effects of simple processing on the cyanide content of cassava chips. *International Journal of Food Science & Technology*, 13(4), 299–306. <https://doi.org/10.1111/j.1365-2621.1978.tb00807.x>
- Corbeil, C. R., Englebienne, P., & Moitessier, N. (2007). Docking Ligands into Flexible and Solvated Macromolecules: Development and Validation of FITTED 1.0. *Journal of Chemical Information and Modeling*, 47(2), 435–449. <https://doi.org/10.1021/ci6002637>
- Cozzini, P., Fornabaio, M., Marabotti, A., Abraham, D. J., Kellogg, G. E., & Mozzarelli, A. (2002). Simple, intuitive calculations of free energy of binding for protein–ligand complexes: Models without explicit constrained water. *Journal of Medicinal Chemistry*, 45(12), 2469–2483. <https://doi.org/10.1021/jm0200299>
- Cressey, P., & Reeve, J. (2019). Metabolism of cyanogenic glycosides: A review. *Food and Chemical Toxicology*, 125, 225–232. <https://doi.org/10.1016/j.fct.2019.01.002>
- Cutler, A. J., & Conn, E. E. (1981). The biosynthesis of cyanogenic glucosides in *Linum usitatissimum* (linen flax) in vitro. *Archives of Biochemistry and Biophysics*, 212(2), 468–474. [https://doi.org/10.1016/0003-9861\(81\)90389-1](https://doi.org/10.1016/0003-9861(81)90389-1)
- Dalton, J. A., & Jackson, R. M. (2007). An evaluation of automated homology modelling methods at low target template sequence similarity. *Bioinformatics*, 23(15), 1901–1908. <https://doi.org/10.1093/bioinformatics/btm262>
- Database resources of the National Center for Biotechnology Information. (2016). *Nucleic Acids Research*, 44(D1), D7–19. <https://doi.org/10.1093/nar/gkv1290>

- Database resources of the National Center for Biotechnology Information. (2018). *Nucleic Acids Research*, 46(D1), D8–d13. <https://doi.org/10.1093/nar/gkx1095>
- Davies, G., & Henrissat, B. (1995). Structures and mechanisms of glycosyl hydrolases. *Structure*, 3(9), 853–859.
- Davis, A. M., & Teague, S. J. (1999). Hydrogen Bonding, Hydrophobic Interactions, and Failure of the Rigid Receptor Hypothesis. *Angewandte Chemie*, 38(6), 736–749. [https://doi.org/10.1002/\(sici\)1521-3773\(19990315\)38:6<736:Aid-anie736>3.0.Co;2-r](https://doi.org/10.1002/(sici)1521-3773(19990315)38:6<736:Aid-anie736>3.0.Co;2-r)
- Delange, F., Ekpechi, L. O., & Rosling, H. (1994). *Cassava cyanogenesis and iodine deficiency disorders*. https://www.actahort.org/books/375/375_29.htm
- Diasolua, N. D., Kuo, Y. H., & Lambein, F. (2003). Amino acid profiles and protein quality of cooked cassava leaves or ‘saka-saka’. *Journal of the Science of Food and Agriculture*, 83(6), 529–534. <https://doi.org/10.1002/jsfa.1373>
- Diller, D. J., & Merz, K. M. (2001). High throughput docking for library design and library prioritization. *Proteins*, 43(2), 113–124. [https://doi.org/10.1002/1097-0134\(20010501\)43:2<113:aid-prot1023>3.0.co;2-t](https://doi.org/10.1002/1097-0134(20010501)43:2<113:aid-prot1023>3.0.co;2-t)
- Djazuli, M., & Bradbury, J. H. (1999). Cyanogen content of cassava roots and flour in Indonesia. *Food Chemistry*, 65(4), 523–525. [https://doi.org/10.1016/S0308-8146\(98\)00218-0](https://doi.org/10.1016/S0308-8146(98)00218-0)
- Dodda, L. S., Cabeza de Vaca, I., Tirado-Rives, J., & Jorgensen, W. L. (2017). LigParGen web server: An automatic OPLS-AA parameter generator for organic ligands. *Nucleic Acids Research*, 45(W1), W331–W336. <https://doi.org/10.1093/nar/gkx312>
- Dodda, L. S., Vilseck, J. Z., Tirado-Rives, J., & Jorgensen, W. L. (2017). 1.14* CM1A-LBCC: localized bond-charge corrected CM1A charges for condensed-phase simulations. *The Journal of Physical Chemistry B*, 121(15), 3864–3870.
- Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, 32(5), 1792–1797. <https://doi.org/10.1093/nar/gkh340>

- Emsley, P., & Cowtan, K. (2004). Coot: Model-building tools for molecular graphics. *Acta Crystallographica. Section D, Biological Crystallography*, 60(Pt 12 Pt 1), 2126–2132. <https://doi.org/10.1107/s0907444904019158>
- Esen, A., & Blanchard, D. J. (2000). A specific beta-glucosidase-aggregating factor is responsible for the beta-glucosidase null phenotype in maize. *Plant Physiology*, 122(2), 563–572. <https://doi.org/10.1104/pp.122.2.563>
- Ewald, P. P. (1921). Die Berechnung Optischer und Elektrostatischer Gitterpotentiale. *Annalen der Physik*, 369(3), 253–287. <http://dx.doi.org/10.1002/andp.19213690304>
- Ewing, T. J. A., Makino, S., Skillman, A. G., & Kuntz, I. D. (2001). DOCK 4.0: Search strategies for automated molecular docking of flexible molecule databases. *Journal of Computer-aided Molecular Design*, 15(5), 411–428. <https://doi.org/10.1023/A:1011115820450>
- Fadrná, E., Hladecková, K., & Koca, J. (2005). Long-range electrostatic interactions in molecular dynamics: An endothelin-1 case study. *Journal of Biomolecular Structure and Dynamics*, 23(2), 151–162. <https://doi.org/10.1080/07391102.2005.10531229>
- Falade, K. O., & Akingbala, J. O. (2010). Utilization of Cassava for Food. *Food Reviews International*, 27(1), 51–83. <https://doi.org/10.1080/87559129.2010.518296>
- Fletcher, R., & Powell, M. J. D. (1963). A Rapidly Convergent Descent Method for Minimization. *The Computer Journal*, 6(2), 163–168. <https://doi.org/10.1093/comjnl/6.2.163>
- Fogolari, F., Brigo, A., & Molinari, H. (2003). Protocol for MM/PBSA molecular dynamics simulations of proteins. *Biophysical Journal*, 85(1), 159–166. [https://doi.org/10.1016/S0006-3495\(03\)74462-2](https://doi.org/10.1016/S0006-3495(03)74462-2)
- Foloppe, N., & Hubbard, R. (2006). Towards predictive ligand design with free-energy based computational methods? *Current Medicinal Chemistry*, 13(29), 3583–3608. <https://doi.org/10.2174/092986706779026165>
- Fradera, X., Kaur, J., & Mestres, J. (2004). Unsupervised guided docking of covalently bound ligands. *Journal of Computer-aided Molecular Design*, 18(10), 635–650. <https://doi.org/10.1007/s10822-004-5291-4>
- 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., & Shenkin, P. S.

- (2004). Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *Journal of Medicinal Chemistry*, 47(7), 1739–1749. <https://doi.org/10.1021/jm0306430>
- García–Escudero, V., & Gargini, R. (2008). Autophagy induction as an efficient strategy to eradicate tumors. *Autophagy*, 4(7), 923–925. <https://doi.org/10.4161/auto.6714>
- Genheden, S., & Ryde, U. (2015). The MM/PBSA and MM/GBSA methods to estimate ligand–binding affinities. *Expert Opinion on Drug Discovery*, 10(5), 449–461. <https://doi.org/10.1517/17460441.2015.1032936>
- Geraldene, M., & Mahmoud, E. S. S. (2017). Homology Modeling in Drug Discovery: An Update on the Last Decade. *Letters in Drug Design & Discovery*, 14(9), 1099–1111. <https://doi.org/http://dx.doi.org/10.2174/1570180814666170110122027>
- Ghaffari, A. D., Dalimi, A., Ghaffarifar, F., & Pirestani, M. (2020). Structural predication and antigenic analysis of ROP16 protein utilizing immunoinformatics methods in order to identification of a vaccine against *Toxoplasma gondii*: An in silico approach. *Microbial Pathogenesis*, 142, 104079. <https://doi.org/https://doi.org/10.1016/j.micpath.2020.104079>
- Girald, W., Collin, A., & Izquierdo, M. (2011). Toxicity and delivery methods for the linamarase/linamarin/glucose oxidase system, when used against human glioma tumors implanted in the brain of nude rats. *Cancer Letters*, 313(1), 99–107. <https://doi.org/10.1016/j.canlet.2011.08.029>
- Goodsell, D. S., & Olson, A. J. (1990). Automated docking of substrates to proteins by simulated annealing. *Proteins*, 8(3), 195–202. <https://doi.org/10.1002/prot.340080302>
- Grazioso, G., Pomè, D. Y., Matera, C., Frigerio, F., Pucci, L., Gotti, C., Dallanoce, C., & De Amici, M. (2009). Design of novel $\alpha 7$ –subtype–preferring nicotinic acetylcholine receptor agonists: Application of docking and MM–PBSA computational approaches, synthetic and pharmacological studies. *Bioorganic & Medicinal Chemistry Letters*, 19(22), 6353–6357. <https://doi.org/10.1016/j.bmcl.2009.09.073>
- Gueguen, Y., Chemardin, P., Janbon, G., Arnaud, A., & Galzy, P. (1998). Investigation of the β –glucosidases potentialities of yeast strains and application to bound aromatic terpenols liberation. In K. Kieslich, C. P. Van der Beek, J. A. M. de Bont, & W. J. J. Van den Tweel

- (Eds.), *Studies in Organic Chemistry* (pp. 149–157). Elsevier. [https://doi.org/10.1016/S0165-3253\(98\)80018-7](https://doi.org/10.1016/S0165-3253(98)80018-7)
- Gupta, C. L., Akhtar, S., & Bajpai, P. (2014). In silico protein modeling: Possibilities and limitations. *Excli Journal*, 13, 513–515. <https://pubmed.ncbi.nlm.nih.gov/26417278>
- Haghighi, O., Davaeifar, S., Zahiri, H. S., Maleki, H., & Noghabi, K. A. (2020). Homology Modeling and Molecular Docking Studies of Glutamate Dehydrogenase (GDH) from Cyanobacterium *Synechocystis* sp. PCC 6803. *International Journal of Peptide Research and Therapeutics*, 26(2), 783–793. <https://doi.org/10.1007/s10989-019-09886-4>
- Halgren, T. A., Murphy, R. B., Friesner, R. A., Beard, H. S., Frye, L. L., Pollard, W. T., & Banks, J. L. (2004). Glide: A new approach for rapid, accurate docking and scoring: Enrichment factors in database screening. *Journal of Medicinal Chemistry*, 47(7), 1750–1759. <https://doi.org/10.1021/jm030644s>
- Hart, T. N., & Read, R. J. (1992). A multiple-start Monte Carlo docking method. *Proteins: Structure, Function, and Bioinformatics*, 13(3), 206–222. <https://doi.org/10.1002/prot.340130304>
- Hillisch, A., Pineda, L. F., & Hilgenfeld, R. (2004). Utility of homology models in the drug discovery process. *Drug Discovery Today*, 9(15), 659–669. [https://doi.org/10.1016/s1359-6446\(04\)03196-4](https://doi.org/10.1016/s1359-6446(04)03196-4)
- Homeyer, N., & Gohlke, H. (2012). Free Energy Calculations by the Molecular Mechanics Poisson–Boltzmann Surface Area Method. *Molecular Informatics*, 31(2), 114–122. <https://doi.org/10.1002/minf.201100135>
- Hooft, R. W. W., Vriend, G., Sander, C., & Abola, E. E. (1996). Errors in protein structures. *Nature*, 381(6580), 272–272. <https://doi.org/10.1038/381272a0>
- Hotz, C., & Gibson, R. S. (2007). Traditional Food–Processing and Preparation Practices to Enhance the Bioavailability of Micronutrients in Plant–Based Diets. *The Journal of Nutrition*, 137(4), 1097–1100. <https://doi.org/10.1093/jn/137.4.1097>
- Hou, X., Du, J., Zhang, J., Du, L., Fang, H., & Li, M. (2013). How to Improve Docking Accuracy of AutoDock4.2: A Case Study Using Different Electrostatic Potentials. *Journal of Chemical Information and Modeling*, 53(1), 188–200. <https://doi.org/10.1021/ci300417y>

- Howlin, B. J. (1993). Chapter 3. Molecular modelling. *Annual Reports Section "C" (Physical Chemistry)*, 90(0), 45–66. <https://doi.org/10.1039/PC9939000045>
- Huang, N., Kalyanaraman, C., Irwin, J. J., & Jacobson, M. P. (2006). Physics-based scoring of protein–ligand complexes: Enrichment of known inhibitors in large-scale virtual screening. *Journal of Chemical Information and Modeling*, 46(1), 243–253. <https://doi.org/10.1021/ci0502855>
- Huang, S. Y., & Zou, X. (2010). Inclusion of solvation and entropy in the knowledge-based scoring function for protein–ligand interactions. *Journal of Chemical Information and Modeling*, 50(2), 262–273. <https://doi.org/10.1021/ci9002987>
- Idibie, C. A., Davids, H., & Iyuke, S. E. (2007). Cytotoxicity of purified cassava linamarin to a selected cancer cell lines. *Bioprocess and Biosystems Engineering*, 30(4), 261–269. <https://doi.org/10.1007/s00449-007-0122-3>
- Ikediodi, C., & Onyike, E. (1982). The use of linamarase in gari production. *Process Biochemistry*, 17(4), 2–5.
- Ishchenko, A. V., & Shakhnovich, E. I. (2002). SMoG2001 (SMoG2001): An Improved Knowledge-Based Scoring Function for Protein–Ligand Interactions. *Journal of Medicinal Chemistry*, 45(13), 2770–2780. <https://doi.org/10.1021/jm0105833>
- Isorna, P., Polaina, J., Latorre–García, L., Cañada, F. J., González, B., & Sanz–Aparicio, J. (2007). Crystal structures of *Paenibacillus polymyxa* beta–glucosidase B complexes reveal the molecular basis of substrate specificity and give new insights into the catalytic machinery of family I glycosidases. *Journal of Molecular Biology*, 371(5), 1204–1218. <https://doi.org/10.1016/j.jmb.2007.05.082>
- Jackson, R. M. (2002). Q–fit: A probabilistic method for docking molecular fragments by sampling low energy conformational space. *Journal of Computer–aided Molecular Design*, 16(1), 43–57. <https://doi.org/10.1023/A:1016307520660>
- Jeng, W. Y., Wang, N. C., Lin, M. H., Lin, C. T., Liaw, Y. C., Chang, W. J., Liu, C. I., Liang, P. H., & Wang, A. H. (2011). Structural and functional analysis of three β –glucosidases from bacterium *Clostridium cellulovorans*, fungus *Trichoderma reesei* and termite *Neotermes koshunensis*. *Journal of Structural Biology*, 173(1), 46–56. <https://doi.org/10.1016/j.jsb.2010.07.008>

- Jianmin, G., Gall, R., & Zuomin, W. (2001). Dynamic Damping and Stiffness Characteristics of the Rolling Tire. *Tire Science and Technology*, 29(4), 258–268. <https://doi.org/10.2346/1.2135243>
- Jones–Hertzog, D. K., & Jorgensen, W. L. (1997). Binding affinities for sulfonamide inhibitors with human thrombin using Monte Carlo simulations with a linear response method. *Journal of Medicinal Chemistry*, 40(10), 1539–1549. <https://doi.org/10.1021/jm960684e>
- Jørgensen, K., Morant, A. V., Morant, M., Jensen, N. B., Olsen, C. E., Kannangara, R., Motawia, M. S., Møller, B. L., & Bak, S. (2011). Biosynthesis of the cyanogenic glucosides linamarin and lotaustralin in cassava: Isolation, biochemical characterization, and expression pattern of CYP71E7, the oxime–metabolizing cytochrome P450 enzyme. *Plant Physiology*, 155(1), 282–292. <https://doi.org/10.1104/pp.110.164053>
- Jorgensen, W. L., Maxwell, D. S., & Tirado–Rives, J. (1996). Development and Testing of the OPLS All–Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *Journal of the American Chemical Society*, 118(45), 11225–11236. <https://doi.org/10.1021/ja9621760>
- Jorgensen, W. L., & Tirado–Rives, J. (1988). The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. *Journal of the American Chemical Society*, 110(6), 1657–1666. <https://doi.org/10.1021/ja00214a001>
- Jorgensen, W. L., & Tirado–Rives, J. (2005). Potential energy functions for atomic–level simulations of water and organic and biomolecular systems. *Proceedings of the National Academy of Sciences*, 102(19), 6665–6670.
- Joseph–McCarthy, D., Thomas, B. E. T., Belmarsh, M., Moustakas, D., & Alvarez, J. C. (2003). Pharmacophore–based molecular docking to account for ligand flexibility. *Proteins*, 51(2), 172–188. <https://doi.org/10.1002/prot.10266>
- Katebi, A. R., Kloczkowski, A., & Jernigan, R. L. (2010). Structural interpretation of protein–protein interaction network. *Structural Biology*, 10 (1), 1-4. <https://doi.org/10.1186/1472-6807-10-s1-s4>

- Kato, Y., & Terada, H. (2014). Determination method of linamarin in cassava products and beans by ultra high performance liquid chromatography with tandem mass spectrometry. *Shokuhin Eiseigaku Zasshi*, 55(3), 162–166. <https://doi.org/10.3358/shokueishi.55.162>
- Kepp, O., Menger, L., Vacchelli, E., Adjemian, S., Martins, I., Ma, Y., Sukkurwala, A. Q., Michaud, M., Galluzzi, L., Zitvogel, L., & Kroemer, G. (2012). Anticancer activity of cardiac glycosides. *OncImmunology*, 1(9), 1640–1642. <https://doi.org/10.4161/onci.21684>
- Ketudat, J. R., & Esen, A. (2010). β -Glucosidases. *Cellular and Molecular Life Sciences*, 67(20), 3389–3405. <https://doi.org/10.1007/s00018-010-0399-2>
- Kim, H. H., Hyun, J. S., Choi, J., Choi, K. E., Jee, J. G., & Park, S. J. (2018). Structural ensemble-based docking simulation and biophysical studies discovered new inhibitors of Hsp90 N-terminal domain. *Scientific Reports*, 8(1), 368. <https://doi.org/10.1038/s41598-017-18332-8>
- Kim, S., Chen, J., Cheng, T., Gindulyte, A., He, J., He, S., Li, Q., Shoemaker, B. A., Thiessen, P. A., Yu, B., Zaslavsky, L., Zhang, J., & Bolton, E. E. (2021). PubChem in 2021: New data content and improved web interfaces. *Nucleic Acids Research*, 49(D1), D1388–D1395. <https://doi.org/10.1093/nar/gkaa971>
- Kim, T. R., Oh, S., Yang, J. S., Lee, S., Shin, S., & Lee, J. (2012). A simplified homology-model builder toward highly protein-like structures: An inspection of restraining potentials. *Journal of Computational Chemistry*, 33(24), 1927–1935. <https://doi.org/10.1002/jcc.23024>
- Kmiecik, S., Gront, D., Kolinski, M., Wieteska, L., Dawid, A. E., & Kolinski, A. (2016). Coarse-Grained Protein Models and Their Applications. *Chemical Reviews*, 116(14), 7898–7936. <https://doi.org/10.1021/acs.chemrev.6b00163>
- Kohring, K., Wiesner, J., Altenkämper, M., Sakowski, J., Silber, K., Hillebrecht, A., Haebel, P., Dahse, H. M., Ortmann, R., Jomaa, H., Klebe, G., & Schlitzer, M. (2008). Development of benzophenone-based farnesyltransferase inhibitors as novel antimalarials. *Medicinal Chemistry*, 3(8), 1217–1231. <https://doi.org/10.1002/cmdc.200800043>

- Kokh, D. B., Wade, R. C., & Wenzel, W. (2011). Receptor flexibility in small-molecule docking calculations. *Computational Molecular Science*, 1(2), 298–314. [https:// doi. org/ https:// doi. org/ 10.1002/wcms.29](https://doi.org/10.1002/wcms.29)
- Kollman, P. (1993). Free energy calculations: Applications to chemical and biochemical phenomena. *Chemical Reviews*, 93(7), 2395–2417. <https://doi.org/10.1021/cr00023a004>
- Kongsaeree, P. T., Ratananikom, K., Choengpanya, K., Tongtubtim, N., Sujiwattanasat, P., Porncharoenp, C., Onpium, A., & Svasti, J. (2010). Substrate specificity in hydrolysis and transglucosylation by family 1 β -glucosidases from cassava and Thai rosewood. *Journal of Molecular Catalysis B: Enzymatic*, 67(3–4), 257–265. [https:// doi. org/ 10. 1016/ j. molcatb.2010.09.003](https://doi.org/10.1016/j.molcatb.2010.09.003)
- Krammer, A., Kirchhoff, P. D., Jiang, X., Venkatachalam, C. M., & Waldman, M. (2005). LigScore: A novel scoring function for predicting binding affinities. *Journal of Molecular Graphics and Modelling*, 23(5), 395–407. [https:// doi. org/ https:// doi. org/ 10. 1016/ j. jmglm. 2004.11.007](https://doi.org/10.1016/j.jmglm.2004.11.007)
- Krieger, E., Darden, T., Nabuurs, S. B., Finkelstein, A., & Vriend, G. (2004). Making optimal use of empirical energy functions: Force-field parameterization in crystal space. *Proteins: Structure, Function, and Bioinformatics*, 57(4), 678–683. [https:// doi. org/ https:// doi. org/ 10. 1002/ prot.20251](https://doi.org/10.1002/prot.20251)
- Krieger, E., Koraimann, G., & Vriend, G. (2002). Increasing the precision of comparative models with YASARA NOVA: A self-parameterizing force field. *Proteins*, 47(3), 393–402. <https://doi.org/10.1002/prot.10104>
- Krisch, J., Takó, M., Papp, T., & Vágvölgyi, C. (2010). Characteristics and potential use of glucosidases from Zygomycetes. *Current Research, Technology and Education Topics in Applied Microbiology and Microbial Biotechnology*, 2010, 891-895.
- Krivov, G. G., Shapovalov, M. V., & Dunbrack, R. L. (2009). Improved prediction of protein side-chain conformations with SCWRL4. *Proteins*, 77(4), 778–795. [https:// doi. org/ 10. 1002/ prot.22488](https://doi.org/10.1002/prot.22488)
- Kuhn, B., & Kollman, P. A. (2000). Binding of a Diverse Set of Ligands to Avidin and Streptavidin: An Accurate Quantitative Prediction of their Relative Affinities by a

- Combination of Molecular Mechanics and Continuum Solvent Models. *Journal of Medicinal Chemistry*, 43(20), 3786–3791. <https://doi.org/10.1021/jm000241h>
- Kumari, R., Kumar, R., & Lynn, A. (2014). G_mmpbsa—a GROMACS tool for high-throughput MM–PBSA calculations. *Journal of Chemical Information and Modeling*, 54(7), 1951–1962. <https://doi.org/10.1021/ci500020m>
- Kuntz, I. D., Blaney, J. M., Oatley, S. J., Langridge, R., & Ferrin, T. E. (1982). A geometric approach to macromolecule–ligand interactions. *Journal of Molecular Biology*, 161(2), 269–288. [https://doi.org/https://doi.org/10.1016/0022-2836\(82\)90153-X](https://doi.org/https://doi.org/10.1016/0022-2836(82)90153-X)
- Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J., & Higgins, D. G. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics*, 23(21), 2947–2948. <https://doi.org/10.1093/bioinformatics/btm404>
- Laskowski, R. A., MacArthur, M. W., Moss, D. S., & Thornton, J. M. (1993). PROCHECK: A program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography*, 26(2), 283–291. <https://doi.org/https://doi.org/10.1107/S0021889892009944>
- Lee, H. M., & Moon, A. (2016). Amygdalin Regulates Apoptosis and Adhesion in Hs578T Triple–Negative Breast Cancer Cells. *Biomolecules & Therapeutics*, 24(1), 62–66. <https://doi.org/10.4062/biomolther.2015.172>
- Legoll, F., & Monneau, R. (2002). Designing reversible measure invariant algorithms with applications to molecular dynamics. *The Journal of Chemical Physics*, 117(23), 10452–10464. <https://doi.org/10.1063/1.1519842>
- Levitt, M. (1992). Accurate modeling of protein conformation by automatic segment matching. *Journal of Molecular Biology*, 226(2), 507–533. [https://doi.org/10.1016/0022-2836\(92\)90964-1](https://doi.org/10.1016/0022-2836(92)90964-1)
- Levitt, M. (2007). Growth of novel protein structural data. *Proceedings of the National Academy of Sciences*, 104(9), 3183. <https://doi.org/10.1073/pnas.0611678104>
- Levy, R. M., Zhang, L. Y., Gallicchio, E., & Felts, A. K. (2003). On the nonpolar hydration free energy of proteins: Surface area and continuum solvent models for the solute–solvent

- interaction energy. *Journal of the American Chemical Society*, 125(31), 9523–9530. <https://doi.org/10.1021/ja029833a>
- Li, D., Li, X., Dang, W., Tran, P. L., Park, S. H., Oh, B. C., Hong, W. S., Lee, J. S., & Park, K. H. (2013). Characterization and application of an acidophilic and thermostable β -glucosidase from *Thermofilum pendens*. *Journal of Bioscience and Bioengineering*, 115(5), 490–496. <https://doi.org/10.1016/j.jbiosc.2012.11.009>
- Li, D., Li, X., Dang, W., Tran, P. L., Park, S. H., Oh, B. C., Hong, W. S., Lee, J. S., & Park, K. H. (2013). Characterization and application of an acidophilic and thermostable β -glucosidase from *Thermofilum pendens*. *Journal of Bioscience and Bioengineering*, 115(5), 490–496. <https://doi.org/10.1016/j.jbiosc.2012.11.009>
- Lin, J. H., Perryman, A. L., Schames, J. R., & McCammon, J. A. (2002). Computational drug design accommodating receptor flexibility: The relaxed complex scheme. *Journal of the American Chemical Society*, 124(20), 5632–5633. <https://doi.org/10.1021/ja0260162>
- Lindorff-Larsen, K., Piana, S., Palmo, K., Maragakis, P., Klepeis, J. L., Dror, R. O., & Shaw, D. E. (2010). Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins*, 78(8), 1950–1958. <https://doi.org/10.1002/prot.22711>
- Lovell, S. C., Davis, I. W., Arendall Iii, W. B., De Bakker, P. I. W., Word, J. M., Prisant, M. G., Richardson, J. S., & Richardson, D. C. (2003). Structure validation by $\text{C}\alpha$ geometry: ϕ, ψ and $\text{C}\beta$ deviation. *Proteins: Structure, Function, and Bioinformatics*, 50(3), 437–450. <https://doi.org/https://doi.org/10.1002/prot.10286>
- Lykkesfeldt, J., & Møller, B. L. (1994). Cyanogenic glucosides in cassava, *Manihot esculenta*. *Acta Chemica Scandinavica*, 48, 178–180.
- Ma, B., Shatsky, M., Wolfson, H. J., & Nussinov, R. (2002). Multiple diverse ligands binding at a single protein site: A matter of pre-existing populations. *Protein science: A publication of the Protein Society*, 11(2), 184–197. <https://doi.org/10.1110/ps.21302>
- MacKerell, A. D., Banavali, N., & Foloppe, N. (2000). Development and current status of the CHARMM force field for nucleic acids. *Biopolymers*, 56(4), 257–265. [https://doi.org/10.1002/1097-0282\(2000\)56:4<257:Aid-bip10029>3.0.Co;2-w](https://doi.org/10.1002/1097-0282(2000)56:4<257:Aid-bip10029>3.0.Co;2-w)

- Magrane, M., & UniProt, C. (2011). UniProt Knowledgebase: A hub of integrated protein data. *Database: The Journal of Biological Databases and Curation*, 2011, bar009–bar009. <https://doi.org/10.1093/database/bar009>
- Mark, P., & Nilsson, L. (2001). Structure and Dynamics of the TIP3P, SPC, and SPC/E Water Models at 298 K. *The Journal of Physical Chemistry A*, 105(43), 9954–9960. <https://doi.org/10.1021/jp003020w>
- Martyna, G. J., Hughes, A., & Tuckerman, M. E. (1999). Molecular dynamics algorithms for path integrals at constant pressure. *The Journal of Chemical Physics*, 110(7), 3275–3290. <https://doi.org/10.1063/1.478193>
- May, A., & Zacharias, M. (2005). Accounting for global protein deformability during protein–protein and protein–ligand docking. *Biochimica et Biophysica Acta*, 1754(1–2), 225–231. <https://doi.org/10.1016/j.bbapap.2005.07.045>
- Mazlan, N. S., & Khairudin, N. B. (2016). A molecular dynamics study of Beta–Glucosidase B upon small substrate binding. *Journal of Biomolecular Structure and Dynamics*, 34(7), 1486–1494. <https://doi.org/10.1080/07391102.2015.1081570>
- McMahon, J. M., White, W. L. B., & Sayre, R. T. (1995). Review Article: Cyanogenesis in cassava (*Manihot esculenta* Crantz). *Journal of Experimental Botany*, 46(7), 731–741. <https://doi.org/10.1093/jxb/46.7.731>
- McMartin, C., & Bohacek, R. S. (1997). QXP: Powerful, rapid computer algorithms for structure–based drug design. *Journal of Computer-Aided Molecular Design*, 11(4), 333–344. <https://doi.org/10.1023/a:1007907728892>
- McWilliam, H., Li, W., Uludag, M., Squizzato, S., Park, Y. M., Buso, N., Cowley, A. P., & Lopez, R. (2013). Analysis Tool Web Services from the EMBL–EBI. *Nucleic Acids Research*, 41(Web Server issue), W597–600. <https://doi.org/10.1093/nar/gkt376>
- Meiler, J., & Baker, D. (2006). Rosetta Ligand: Protein–small molecule docking with full side–chain flexibility. *Proteins*, 65(3), 538–548. <https://doi.org/10.1002/prot.21086>
- Meirovitch, H. (2007). Recent developments in methodologies for calculating the entropy and free energy of biological systems by computer simulation. *Current Opinion in Structural Biology*, 17(2), 181–186. <https://doi.org/10.1016/j.sbi.2007.03.016>

- Meliciani, I., Klenin, K., Strunk, T., Schmitz, K., & Wenzel, W. (2009). Probing hot spots on protein–protein interfaces with all–atom free–energy simulation. *The Journal of Chemical Physics*, 131(3), 034114. <https://doi.org/10.1063/1.3177008>
- Meng, E. C., Shoichet, B. K., & Kuntz, I. D. (1992). Automated docking with grid–based energy evaluation. *Journal of Computational Chemistry*, 13(4), 505–524. <https://doi.org/10.1002/jcc.540130412>
- Miller, M. D., Kearsley, S. K., Underwood, D. J., & Sheridan, R. P. (1994). Flog: A system to select ‘quasi–flexible’ ligands complementary to a receptor of known three–dimensional structure. *Journal of Computer–aided Molecular Design*, 8(2), 153–174. <https://doi.org/10.1007/BF00119865>
- Mitchell, J. B. O., Laskowski, R. A., Alex, A., Forster, M. J., & Thornton, J. M. (1999). BLEEP—potential of mean force describing protein–ligand interactions: Calculation of binding energies and comparison with experimental data. *Journal of Computational Chemistry*, 20(11), 1177–1185. [https://doi.org/10.1002/\(SICI\)1096-987X\(199908\)20:11<1177::AID-JCC8>3.0.CO;2-0](https://doi.org/10.1002/(SICI)1096-987X(199908)20:11<1177::AID-JCC8>3.0.CO;2-0)
- Mlingi, N., Poulter, N. H., & Rosling, H. (1992). An outbreak of acute intoxications from consumption of insufficiently processed cassava in Tanzania. *Nutrition Research*, 12(6), 677–687. [https://doi.org/10.1016/S0271-5317\(05\)80565-2](https://doi.org/10.1016/S0271-5317(05)80565-2)
- Moitessier, N., Therrien, E., & Hanessian, S. (2006). A method for induced–fit docking, scoring, and ranking of flexible ligands. Application to peptidic and pseudopeptidic beta–secretase (BACE 1) inhibitors. *Journal of Medicinal Chemistry*, 49(20), 5885–5894. <https://doi.org/10.1021/jm050138y>
- Montagnac, J. A., Davis, C. R., & Tanumihardjo, S. A. (2009a). Nutritional Value of Cassava for Use as a Staple Food and Recent Advances for Improvement. *Comprehensive Reviews in Food Science and Food Safety*, 8(3), 181–194. <https://doi.org/10.1111/j.1541-4337.2009.00077.x>
- Montagnac, J. A., Davis, C. R., & Tanumihardjo, S. A. (2009b). Processing Techniques to Reduce Toxicity and Antinutrients of Cassava for Use as a Staple Food. *Comprehensive Reviews in Food Science and Food Safety*, 8(1), 17–27. <https://doi.org/10.1111/j.1541-4337.2008.00064.x>

- Morant, A. V., Jørgensen, K., Jørgensen, C., Paquette, S. M., Sánchez-Pérez, R., Møller, B. L., & Bak, S. (2008). Beta-Glucosidases as detonators of plant chemical defense. *Phytochemistry*, 69(9), 1795–1813. <https://doi.org/10.1016/j.phytochem.2008.03.006>
- Morris, G. M., Goodsell, D. S., Halliday, R. S., Huey, R., Hart, W. E., Belew, R. K., & Olson, A. J. (1998). Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *Journal of Computational Chemistry*, 19(14), 1639–1662. [https://doi.org/https://doi.org/10.1002/\(SICI\)1096-987X\(19981115\)19:14<1639:AID-JCC10>3.0.CO;2-B](https://doi.org/https://doi.org/10.1002/(SICI)1096-987X(19981115)19:14<1639:AID-JCC10>3.0.CO;2-B)
- Muegge, I. (2006). PMF scoring revisited. *Journal of Medicinal Chemistry*, 49(20), 5895–5902. <https://doi.org/10.1021/jm050038s>
- Muegge, I., & Martin, Y. C. (1999). A general and fast scoring function for protein–ligand interactions: A simplified potential approach. *Journal of Medicinal Chemistry*, 42(5), 791–804. <https://doi.org/10.1021/jm980536j>
- Muhammed, M. T., & Aki-Yalcin, E. (2019). Homology modeling in drug discovery: Overview, current applications, and future perspectives. *Chemical Biology & Drug Design*, 93(1), 12–20. <https://doi.org/10.1111/cbdd.13388>
- Murugan, K., Sekar, K., & Al-Sohaibani, S. (2012). Detoxification of cyanides in cassava flour by linamarase of *Bacillus subtilis* KM05 isolated from cassava peel. *African Journal of Biotechnology*, 11(28), 7232–7237.
- Nambisan, B. (1994). *Evaluation of the effect of various processing techniques on cyanogen content reduction in cassava*. International Workshop on Cassava Safety 375. <https://www.google.com>
- Nambisan, B. (1999). Cassava latex as a source of linamarase for determination of linamarin. *Journal of Agricultural and Food Chemistry*, 47(2), 372–373. <https://doi.org/10.1021/jf980768r>
- Nambisan, B. (2011). Strategies for elimination of cyanogens from cassava for reducing toxicity and improving food safety. *Food and Chemical Toxicology*, 49(3), 690–693. <https://doi.org/10.1016/j.fct.2010.10.035>

- Nicholls, A., & Honig, B. (1991). A rapid finite difference algorithm, utilizing successive over-relaxation to solve the Poisson–Boltzmann equation. *Journal of Computational Chemistry*, 12(4), 435–445. <https://doi.org/10.1002/jcc.540120405>
- Noronha, G., Barrett, K., Boccia, A., Brodhag, T., Cao, J., Chow, C. P., Dneprovskaia, E., Doukas, J., Fine, R., Gong, X., Gritzen, C., Gu, H., Hanna, E., Hood, J. D., Hu, S., Kang, X., Key, J., Klebansky, B., Kousba, A., ... & Zhu, H. (2007). Discovery of [7–(2,6–dichlorophenyl)–5–methylbenzo[1,2,4] triazin–3–yl] –[4–(2–pyrrolidin–1 ylethoxy) phenyl] amine—a potent, orally active Src kinase inhibitor with anti–tumor activity in preclinical assays. *Bioorganic & Medicinal Chemistry Letters*, 17(3), 602–608. <https://doi.org/10.1016/j.bmcl.2006.11.006>
- O'Sullivan, O., Suhre, K., Abergel, C., Higgins, D. G., & Notredame, C. (2004). 3DCoffee: Combining protein sequences and structures within multiple sequence alignments. *Journal of Molecular Biology*, 340(2), 385–395. <https://doi.org/10.1016/j.jmb.2004.04.058>
- Oluwole, O. S., Onabolu, A. O., Link, H., & Rosling, H. (2000). Persistence of tropical ataxic neuropathy in a Nigerian community. *Journal of Neurology Neurosurgery and Psychiatry*, 69(1), 96–101. <https://doi.org/10.1136/jnnp.69.1.96>
- Oostenbrink, C., Villa, A., Mark, A. E., & van Gunsteren, W. F. (2004). A biomolecular force field based on the free enthalpy of hydration and solvation: The GROMOS force–field parameter sets 53A5 and 53A6. *Journal of Computational Chemistry*, 25(13), 1656–1676. <https://doi.org/10.1002/jcc.20090>
- Osbourn, A. E. (1996). Preformed Antimicrobial Compounds and Plant Defense against Fungal Attack. *The Plant Cell*, 8(10), 1821–1831. <https://doi.org/10.1105/tpc.8.10.1821>
- Parenti, M. D., & Rastelli, G. (2012). Advances and applications of binding affinity prediction methods in drug discovery. *Biotechnology Advances*, 30(1), 244–250. <https://doi.org/10.1016/j.biotechadv.2011.08.003>
- Parrinello, M., & Rahman, A. (1981). Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied Physics*, 52, 7182. <https://doi.org/10.1063/1.328693>
- Patel, B., Singh, V., & Patel, D. (2019). *Structural Bioinformatics*. https://doi.org/10.1007/978-3-030-02634-9_9

- Paul, L., Shadrack, D. M., Mudogo, C. N., Mtei, K. M., Machunda, R. L., & Ntie-Kang, F. (2021). Structural characterization of cassava linamarase–linamarin enzyme complex: An integrated computational approach. *Journal of Biomolecular Structure and Dynamics*, 2021, 1–9. <https://doi.org/10.1080/07391102.2021.1925156>
- Pei, J., Wang, Q., Liu, Z., Li, Q., Yang, K., & Lai, L. (2006). PSI-DOCK: Towards highly efficient and accurate flexible ligand docking. *Proteins*, 62(4), 934–946. <https://doi.org/10.1002/prot.20790>
- Peng, J. (2013). *Statistical inference for template-based protein structure prediction*. [https://www.google.com.arXiv/preprint arXiv:1306.4420](https://www.google.com.arXiv/preprint/arXiv:1306.4420).
- Petrey, D., Xiang, Z., Tang, C. L., Xie, L., Gimpelev, M., Mitros, T., Soto, C. S., Goldsmith–Fischman, S., Kernytsky, A., Schlessinger, A., Koh, I. Y., Alexov, E., & Honig, B. (2003). Using multiple structure alignments, fast model building, and energetic analysis in fold recognition and homology modeling. *Proteins*, 53(6), 430–435. <https://doi.org/10.1002/prot.10550>
- Petrucchioli, M., Brimer, L., Cicalini, A. R., & Federici, F. (1999). The linamarase of *Mucor circinelloides* LU M40 and its detoxifying activity on cassava. *Journal of Applied Microbiology*, 86(2), 302–310. <https://doi.org/10.1046/j.1365-2672.1999.00655.x>
- Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004). UCSF Chimera: A visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, 25(13), 1605–1612. <https://doi.org/10.1002/jcc.20084>
- Poppenberger, B., Berthiller, F., Lucyshyn, D., Sieberer, T., Schuhmacher, R., Krska, R., Kuchler, K., Glössl, J., Luschnig, C., & Adam, G. (2003). Detoxification of the Fusarium mycotoxin deoxynivalenol by a UDP–glucosyltransferase from *Arabidopsis thaliana*. *Journal of Biological Chemistry*, 278(48), 47905–47914. <https://doi.org/10.1074/jbc.M307552200>
- Rarey, M., Kramer, B., Lengauer, T., & Klebe, G. (1996). A fast flexible docking method using an incremental construction algorithm. *Journal of Molecular Biology*, 261(3), 470–489. <https://doi.org/10.1006/jmbi.1996.0477>

- Rashin, A. A. (1990). Hydration phenomena, classical electrostatics, and the boundary element method. *The Journal of Physical Chemistry*, 94(5), 1725–1733. <https://doi.org/10.1021/j100368a005>
- Rastelli, G., Del Rio, A., Degliesposti, G., & Sgobba, M. (2010). Fast and accurate predictions of binding free energies using MM–PBSA and MM–GBSA. *Journal of Computational Chemistry*, 31(4), 797–810. <https://doi.org/10.1002/jcc.21372>
- Rivadeneira–Domínguez, E., Vázquez–Luna, A., Rodríguez–Landa, J. F., & Díaz–Sobac, R. (2013). Neurotoxic effect of linamarin in rats associated with cassava (*Manihot esculenta* Crantz) consumption. *Food and Chemical Toxicology*, 59, 230–235. <https://doi.org/10.1016/j.fct.2013.06.004>
- Robert, X., & Gouet, P. (2014). Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Research*, 42(W1), W320–W324. <https://doi.org/10.1093/nar/gku316>
- Robertson, M. J., Tirado–Rives, J., & Jorgensen, W. L. (2015). Improved Peptide and Protein Torsional Energetics with the OPLS–AA Force Field. *Journal of Chemical Theory and Computation*, 11(7), 3499–3509. <https://doi.org/10.1021/acs.jctc.5b00356>
- Roy, A., Kucukural, A., & Zhang, Y. (2010). I–TASSER: A unified platform for automated protein structure and function prediction. *Nature Protocols*, 5(4), 725–738. <https://doi.org/10.1038/nprot.2010.5>
- Sali, A., & Blundell, T. L. (1993). Comparative protein modelling by satisfaction of spatial restraints. *Journal of Molecular Biology*, 234(3), 779–815. <https://doi.org/10.1006/jmbi.1993.1626>
- Samanthi, K. A. U., Welideniya, D. T., Acharige, A. D., Samarakoon, S. R., Rathnayaka, R. K., De Silva, M., Perera, S. S., Pieris, C., Wanninayake, U. K., Jayathilaka, A., Karunaratne, V., Amaratunga, G. A. J., & Gunasekera, D. S. (2020). An efficient and high–yielding method for extraction and purification of linamarin from Cassava: In vitro biological evaluation. *Natural Product Research*, 1–4. <https://doi.org/10.1080/14786419.2020.1744136>
- Sanni, A., Franz, C., Schillinger, U., Huch, M., Guigas, C., & Holzapfel, W. (2013). Characterization and Technological Properties of Lactic Acid Bacteria in the Production

- of “Sorghurt,” a Cereal–Based Product. *Food Biotechnology*, 27(2), 178–198. [https:// doi. org/ 10.1080/08905436.2013.781949](https://doi.org/10.1080/08905436.2013.781949)
- Sanz–Aparicio, J., Hermoso, J. A., Martínez–Ripoll, M., González, B., López–Camacho, C., & Polaina, J. (1998). Structural basis of increased resistance to thermal denaturation induced by single amino acid substitution in the sequence of beta–glucosidase A from *Bacillus polymyxa*. *Proteins*, 33(4), 567–576. [https:// doi. org/ 10. 1002/ \(sici\)1097–0134\(19981201\)33:4<567:aid–prot9>3.0.co;2–u](https://doi.org/10.1002/(sici)1097-0134(19981201)33:4<567:aid-prot9>3.0.co;2-u)
- Sanz–Aparicio, J., Hermoso, J. A., Martínez–Ripoll, M., Lequerica, J. L., & Polaina, J. (1998). Crystal structure of beta–glucosidase A from *Bacillus polymyxa*: Insights into the catalytic activity in family 1 glycosyl hydrolases. *Journal of Molecular Biology*, 275(3), 491–502. <https://doi.org/10.1006/jmbi.1997.1467>
- Schames, J. R., Henchman, R. H., Siegel, J. S., Sotriffer, C. A., Ni, H., & McCammon, J. A. (2004). Discovery of a Novel Binding Trench in HIV Integrase. *Journal of Medicinal Chemistry*, 47(8), 1879–1881. <https://doi.org/10.1021/jm0341913>
- Shen, M. Y., & Sali, A. (2006). Statistical potential for assessment and prediction of protein structures. *Protein Science: A publication of the Protein Society*, 15(11), 2507–2524. <https://doi.org/10.1110/ps.062416606>
- Shim, Y. Y., Olivia, C. M., Liu, J., Boonen, R., Shen, J., & Reaney, M. J. T. (2016). Secoisolariciresinol Diglucoside and Cyanogenic Glycosides in Gluten–free Bread Fortified with Flaxseed Meal. *Journal of Agricultural and Food Chemistry*, 64(50), 9551–9558. <https://doi.org/10.1021/acs.jafc.6b03962>
- Shoichet, B. K., Leach, A. R., & Kuntz, I. D. (1999). Ligand solvation in molecular docking. *Proteins*, 34(1), 4–16. [https://doi.org/10.1002/\(sici\)1097–0134\(19990101\)34:1<4:aid–prot2>3.0.co;2–6](https://doi.org/10.1002/(sici)1097-0134(19990101)34:1<4:aid-prot2>3.0.co;2-6)
- Silano, V., Bansul, H., & Bozzini, A. (1981). *Improvement of Nutritional Quality of Food Crops: A State of the Art Report* (9251011664). <https://www.google.com>
- Singh, G., Verma, A. K., & Kumar, V. (2016). Catalytic properties, functional attributes and industrial applications of beta–glucosidases. *3 Biotechnology*, 6(1), 1–3. [https:// doi. org/ 10. 1007/ s13205–015–0328–z](https://doi.org/10.1007/s13205-015-0328-z)

- Sinko, W., Lindert, S., & McCammon, J. A. (2013). Accounting for receptor flexibility and enhanced sampling methods in computer-aided drug design. *Chemical Biology & Drug Design*, 81(1), 41–49. <https://doi.org/10.1111/cbdd.12051>
- Siritunga, D., & Sayre, R. T. (2003). Generation of cyanogen-free transgenic cassava. *Planta*, 217(3), 367–373.
- Söding, J. (2005). Protein homology detection by HMM–HMM comparison. *Bioinformatics*, 21(7), 951–960. <https://doi.org/10.1093/bioinformatics/bti125>
- Souda, R., Kawanowa, H., Kondo, M., & Gotoh, Y. (2003). Hydrogen bonding between water and methanol studied by temperature-programmed time-of-flight secondary ion mass spectrometry. *The Journal of Chemical Physics*, 119(12), 6194–6200. <https://doi.org/10.1063/1.1602055>
- Srisomsap, C., Subhasitanont, P., Techasakul, S., Surarit, R., & Svasti, J. (1999). Synthesis of homo- and hetero-oligosaccharides by Thai rosewood β -glucosidase. *Biotechnology Letters*, 21(11), 947–951. <https://doi.org/10.1023/A:1005626209655>
- Stepper, J., Dabin, J., Eklof, J. M., Thongpoo, P., Kongsaree, P., Taylor, E. J., Turkenburg, J. P., Brumer, H., & Davies, G. J. (2013). Structure and activity of the *Streptococcus pyogenes* family GH1 6-phospho- β -glucosidase SPy1599. *Acta Crystallographica. Section D, Biological Crystallography*, 69(Pt 1), 16–23. <https://doi.org/10.1107/s0907444912041005>
- Sun, Q., Xing, Y., Qiu, C., & Xiong, L. (2014). The Pasting and Gel Textural Properties of Corn Starch in Glucose, Fructose and Maltose Syrup. *PLoS One*, 9(4), e95862. <https://doi.org/10.1371/journal.pone.0095862>
- Svasti, J., Phongsak, T., & Sarnthima, R. (2003). Transglucosylation of tertiary alcohols using cassava β -glucosidase. *Biochemical and Biophysical Research Communications*, 305(3), 470–475. [https://doi.org/10.1016/s0006-291x\(03\)00793-9](https://doi.org/10.1016/s0006-291x(03)00793-9)
- Swope, W. C., Andersen, H. C., Berens, P. H., & Wilson, K. R. (1982). A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters. *The Journal of Chemical Physics*, 76(1), 637–649. <https://doi.org/10.1063/1.442716>

- Tan, C., Tan, Y. H., & Luo, R. (2007). Implicit Nonpolar Solvent Models. *The Journal of Physical Chemistry B*, 111(42), 12263–12274. <https://doi.org/10.1021/jp073399n>
- Taylor, J. S., & Burnett, R. M. (2000). Darwin: A program for docking flexible molecules. *Proteins: Structure, Function, and Bioinformatics*, 41(2), 173–191. [https://doi.org/10.1002/1097-0134\(20001101\)41:2<173:AID-PROT30>3.0.CO;2-3](https://doi.org/10.1002/1097-0134(20001101)41:2<173:AID-PROT30>3.0.CO;2-3)
- Taylor, R. D., Jewsbury, P. J., & Essex, J. W. (2003). FDS: Flexible ligand and receptor docking with a continuum solvent model and soft-core energy function. *Journal of Computational Chemistry*, 24(13), 1637–1656. <https://doi.org/10.1002/jcc.10295>
- Teeri, T. T. (1997). Crystalline cellulose degradation: New insight into the function of cellobiohydrolases. *Trends in Biotechnology*, 15(5), 160–167. [https://doi.org/10.1016/S0167-7799\(97\)01032-9](https://doi.org/10.1016/S0167-7799(97)01032-9)
- Teles, F. F. F. (2002). Chronic Poisoning by Hydrogen Cyanide in Cassava and its Prevention in Africa and Latin America. *Food and Nutrition Bulletin*, 23(4), 407–412. <https://doi.org/10.1177/156482650202300418>
- Thompson, J. D., Higgins, D. G., & Gibson, T. J. (1994). Clustal W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, 22(22), 4673–4680. <https://doi.org/10.1093/nar/22.22.4673>
- Tian, W., Chen, C., Lei, X., Zhao, J., & Liang, J. (2018). CASTp 3.0: computed atlas of surface topography of proteins. *Nucleic Acids Research*, 46(W1), W363–W367. <https://doi.org/10.1093/nar/gky473>
- Tietze, S., & Apostolakis, J. (2007). GlamDock: Development and validation of a new docking tool on several thousand protein–ligand complexes. *Journal of Chemical Information and Modeling*, 47(4), 1657–1672. <https://doi.org/10.1021/ci7001236>
- Totrov, M., & Abagyan, R. (1997). Flexible protein–ligand docking by global energy optimization in internal coordinates. *Proteins: Structure, Function, and Bioinformatics*, 29(S1), 215–220. [https://doi.org/10.1002/\(SICI\)1097-0134\(1997\)1+<215:AID-PROT29>3.0.CO;2-Q](https://doi.org/10.1002/(SICI)1097-0134(1997)1+<215:AID-PROT29>3.0.CO;2-Q)

- Trosset, J. Y., & Scheraga, H. A. (1999). Prodock: Software package for protein modeling and docking. *Journal of Computational Chemistry*, 20(4), 412–427. [https:// doi. org/ https:// doi. org/10.1002/\(SICI\)1096-987X\(199903\)20:4<412: AID-JCC3>3.0.CO;2-N](https://doi.org/10.1002/(SICI)1096-987X(199903)20:4<412::AID-JCC3>3.0.CO;2-N)
- Trott, O., & Olson, A. J. (2010). AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry*, 31(2), 455–461. <https://doi.org/10.1002/jcc.21334>
- Tuccinardi, T. (2009). Docking-based virtual screening: Recent developments. *Combinatorial Chemistry and High Throughput Screening*, 12(3), 303–314. [https:// doi. org/ 10. 2174/ 138620709787581666](https://doi.org/10.2174/138620709787581666)
- Tuccinardi, T., Manetti, F., Schenone, S., Martinelli, A., & Botta, M. (2007). Construction and validation of a RET TK catalytic domain by homology modeling. *Journal of Chemical Information and Modeling*, 47(2), 644–655. <https://doi.org/10.1021/ci6004383>
- Tuccinardi, T., Poli, G., Dell'Agnello, M., Granchi, C., Minutolo, F., & Martinelli, A. (2015). Receptor-based virtual screening evaluation for the identification of estrogen receptor β ligands. *Journal of Enzyme Inhibition and Medicinal Chemistry*, 30(4), 662–670. <https://doi.org/10.3109/14756366.2014.959946>
- Valdés-Tresanco, M. S., Valdés-Tresanco, M. E., Valiente, P. A., & Moreno, E. (2020). AMDock: a versatile graphical tool for assisting molecular docking with Autodock Vina and Autodock4. *Biology Direct*, 15(1), 12. <https://doi.org/10.1186/s13062-020-00267-2>
- Van Dijk, M., Van Dijk, A. D., Hsu, V., Boelens, R., & Bonvin, A. M. (2006). Information-driven protein–DNA docking using HADDOCK: It is a matter of flexibility. *Nucleic Acids Research*, 34(11), 3317–3325. <https://doi.org/10.1093/nar/gkl412>
- Van Gunsteren, W. F., & Berendsen, H. J. C. (1988). A Leap-frog Algorithm for Stochastic Dynamics. *Molecular Simulation*, 1(3), 173–185. [https:// doi. org/ 10. 1080/ 08927028808080941](https://doi.org/10.1080/08927028808080941)
- Venken, T., Krnavek, D., Münch, J., Kirchhoff, F., Henklein, P., De Maeyer, M., & Voet, A. (2011). An optimized MM/PBSA virtual screening approach applied to an HIV-1 gp41 fusion peptide inhibitor. *Proteins*, 79(11), 3221–3235. <https://doi.org/10.1002/prot.23158>

- Verdonk, M. L., Chessari, G., Cole, J. C., Hartshorn, M. J., Murray, C. W., Nissink, J. W., Taylor, R. D., & Taylor, R. (2005). Modeling water molecules in protein–ligand docking using GOLD. *Journal of Medicinal Chemistry*, 48(20), 6504–6515. <https://doi.org/10.1021/jm050543p>
- Verlet, L. (1967). Computer "Experiments" on Classical: Thermodynamical Properties of Lennard–Jones Molecules. *Physical Review*, 159, 98. <https://doi.org/10.1103/PhysRev.159.98>
- Vetter, J. (2000). Plant cyanogenic glycosides. *Toxicon*, 38(1), 11–36. [https://doi.org/10.1016/S0041-0101\(99\)00128-2](https://doi.org/10.1016/S0041-0101(99)00128-2)
- Vyas, V. K., Ukawala, R. D., Ghate, M., & Chintha, C. (2012). Homology modeling a fast tool for drug discovery: Current perspectives. *Indian Journal of Pharmaceutical Sciences*, 74(1), 1–17. <https://doi.org/10.4103/0250-474x.102537>
- Wagoner, J. A., & Baker, N. A. (2006). Assessing implicit models for nonpolar mean solvation forces: The importance of dispersion and volume terms. *Proceedings of the National Academy of Sciences*, 103(22), 8331. <https://doi.org/10.1073/pnas.0600118103>
- Wajant, H., Mundry, K. W., & Pfizenmaier, K. (1994). Molecular cloning of hydroxynitrile lyase from *Sorghum bicolor* (L.). Homologies to serine carboxypeptidases. *Plant Molecular Biology*, 26(2), 735–746. <https://doi.org/10.1007/bf00013758>
- Wang, G., & Dunbrack, R. L. (2004). Scoring profile–to–profile sequence alignments. *Protein Science: A Publication of the Protein Society*, 13(6), 1612–1626. <https://doi.org/10.1110/ps.03601504>
- Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A., & Case, D. A. (2004). Development and testing of a general amber force field. *Journal of Computational Chemistry*, 25(9), 1157–1174. <https://doi.org/10.1002/jcc.20035>
- Wang, W., & Kollman, P. A. (2000). Free energy calculations on dimer stability of the HIV protease using molecular dynamics and a continuum solvent model. *Journal of Molecular Biology*, 303(4), 567–582. <https://doi.org/10.1006/jmbi.2000.4057>
- Wang, W., & Kollman, P. A. (2000). Free energy calculations on dimer stability of the HIV protease using molecular dynamics and a continuum solvent model¹¹Edited by B. Honig.

- Journal of Molecular Biology*, 303(4), 567–582. <https://doi.org/10.1006/jmbi.2000.4057>
- Wang, W., & Kollman, P. A. (2001). Computational study of protein specificity: The molecular basis of HIV–1 protease drug resistance. *Proceedings of the National Academy of Sciences*, 98(26), 14937. <https://doi.org/10.1073/pnas.251265598>
- Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F. T., De Beer, T. A. P., Rempfer, C., Bordoli, L., Lepore, R., & Schwede, T. (2018). SWISS–MODEL: Homology modelling of protein structures and complexes. *Nucleic Acids Research*, 46(W1), W296–w303. <https://doi.org/10.1093/nar/gky427>
- Werner, T., Morris, M. B., Dastmalchi, S., & Church, W. B. (2012). Structural modelling and dynamics of proteins for insights into drug interactions. *Advanced Drug Delivery Reviews*, 64(4), 323–343. <https://doi.org/10.1016/j.addr.2011.11.011>
- White, W. L. B., Arias–Garzon, D. I., McMahon, J. M., & Sayre, R. T. (1998). Cyanogenesis in cassava. The role of hydroxynitrile lyase in root cyanide production. *Plant Physiology*, 116(4), 1219–1225. <https://doi.org/10.1104/pp.116.4.1219>
- Wiederstein, M., & Sippl, M. J. (2007). ProSA–web: Interactive web service for the recognition of errors in three–dimensional structures of proteins. *Nucleic Acids Research*, 35(Web Server issue), W407–410. <https://doi.org/10.1093/nar/gkm290>
- Wobeto, C., Corrêa, A. D., Abreu, C. M. P. D., Santos, C. D. D., & Pereira, H. V. (2007). Antinutrients in the cassava (*Manihot esculenta* Crantz) leaf powder at three ages of the plant. *Food Science and Technology*, 27(1), 108–112.
- Wu, S., Skolnick, J., & Zhang, Y. (2007). Ab initio modeling of small proteins by iterative TASSER simulations. *BMC Biology*, 5(1), 1–17. <https://doi.org/10.1186/1741-7007-5-17>
- Wu, S. Y., McNae, I., Kontopidis, G., McClue, S. J., McInnes, C., Stewart, K. J., Wang, S., Zheleva, D. I., Marriage, H., Lane, D. P., Taylor, P., Fischer, P. M., & Walkinshaw, M. D. (2003). Discovery of a novel family of CDK inhibitors with the program LIDAEUS: Structural basis for ligand–induced disordering of the activation loop. *Structure*, 11(4), 399–410. [https://doi.org/10.1016/s0969-2126\(03\)00060-1](https://doi.org/10.1016/s0969-2126(03)00060-1)

- Yemm, R. S., & Poulton, J. E. (1986). Isolation and characterization of multiple forms of mandelonitrile lyase from mature black cherry (*Prunus serotina* Ehrh.) seeds. *Archives of Biochemistry and Biophysics*, 247(2), 440–445. [https://doi.org/10.1016/0003-9861\(86\)90604-1](https://doi.org/10.1016/0003-9861(86)90604-1)
- Yeoh, H. H., & Sun, F. (2001). Assessing cyanogen content in cassava-based food using the enzyme-dipstick method. *Food and Chemical Toxicology*, 39(7), 649–653. [https://doi.org/10.1016/S0887-2333\(01\)00005-4](https://doi.org/10.1016/S0887-2333(01)00005-4)
- Yin, B., Gu, H., Mo, X., Xu, Y., Yan, B., Li, Q., Ou, Q., Wu, B., Guo, C., & Jiang, C. (2019). Identification and molecular characterization of a psychrophilic GH1 β -glucosidase from the subtropical soil microorganism *Exiguobacterium* sp. GXG2. *AMB Express*, 9(1), 159. <https://doi.org/10.1186/s13568-019-0873-7>
- Ytreberg, F. M., Swendsen, R. H., & Zuckerman, D. M. (2006). Comparison of free energy methods for molecular systems. *The Journal of Chemical Physics*, 125(18), 184114. <https://doi.org/10.1063/1.2378907>
- Yusuf, U. F., Fakhru'l, R. A. R. R., Iyuke, S. E., Billa, N., Abdullah, N., & Umar-Tsafe, N. (2006). An in vitro inhibition of human malignant cell growth of crude water extract of cassava (*Manihot esculenta* Crantz) and commercial linamarin. *In vitro*, 28(1), 1- 2.
- Zarogoulidis, P., Darwiche, K., Sakkas, A., Yarmus, L., Huang, H., Li, Q., Freitag, L., Zarogoulidis, K., & Malecki, M. (2013). Suicide Gene Therapy for Cancer: Current Strategies. *Journal of Genetic Syndromes & Gene Therapy*, 4, 16849. <https://doi.org/10.4172/2157-7412.1000139>
- Zhang, Y., & Skolnick, J. (2005). TM-align: A protein structure alignment algorithm based on the TM-score. *Nucleic Acids Research*, 33(7), 2302–2309. <https://doi.org/10.1093/nar/gki524>
- Zhu, J., Fan, H., Periole, X., Honig, B., & Mark, A. E. (2008). Refining homology models by combining replica-exchange molecular dynamics and statistical potentials. *Proteins: Structure, Function, and Bioinformatics*, 72(4), 1171–1188. <https://doi.org/10.1002/prot.22005>
- Zou, Z., & Yang, J. (2019). Genomics analysis of the light-harvesting chlorophyll a/b-binding (Lhc) superfamily in cassava (*Manihot esculenta* Crantz). *Gene*, 702, 171–181. <https://doi.org/10.1016/j.gene.2019.03.071>

RESEARCH OUTPUTS

(i) Publications

Paul, L., Celestin, N. M., Kelvin, M. M., Revocatus, L. M., & Fidele, N. K. (2019). A computer-based approach for developing linamarase inhibitory agents. *De gruyter*, 2020, 1-13. **Error! Hyperlink reference not valid.**

Paul, L., Shadrack, D. M., Mudogo, C. N., Mtei, K. M., Machunda, R. L., & Ntie-Kang, F. (2021). Structural characterization of cassava linamarase–linamarin enzyme complex: An integrated computational approach. *Journal of Biomolecular Structure and Dynamics*, 2021, 1–9. <https://doi.org/10.1080/07391102.2021.1925156>

Paul, L., Shadrack, D. M., Mudogo, C. N., Mtei, K. M., Machunda, R. L., & Ntie-Kang, F. (2022). Determination of linamarin Stability in different solvents by using computational approach. *Processes*, 10, 1-18.

(ii) Poster Presentation