3D Structure Prediction and Structural Stability Profiling of Cub-Mutated iRhom2

By

Aqsa Khalid

Fall-2018-MSBI-3-00000273964

Supervised by:

Dr Mehak Rafiq

A THESIS SUBMITTED IN THE PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE in Bioinformatics

School of Interdisciplinary Engineering and Sciences (SINES)

National University of Science & Technology (NUST)

07-10-2021

Dedication

This thesis is dedicated to my beloved parents for their constant support and love.

Certificate of Originality

I hereby declare that the results presented in this research work titled “3D structure prediction and structural stability profiling of cub-mutated iRhom2” are generated by myself. Moreover, none of its contents is plagiarised nor set forth for any evaluation or higher education purposes. I have acknowledged/referenced all the literary content used for support in this research work.

_____

Aqsa Khalid

00000273964

Acknowledgement

First and foremost, praise is to Almighty Allah, who is the most merciful and the greatest creator of all known and unknown worlds, Who gave me the opportunity, strength and determination to complete my research work. Without His grace and mercy, I am nothing. I would like to record my deep and sincere gratitude and heartfelt thanks to my supervisor, Dr. Mehak Rafiq, for her continuous support, advice, the patience of listening and understanding to give better suggestions and constructive criticism. I appreciate her contributions, including time, ideas and support at every stage of my research. I am also highly indebted to my co-supervisor Dr. Rehan Zafar Paracha, for his continuous efforts and help during my research phase. I would also like to thank my GEC member and HOD, Dr. Ishrat Jabeen, for always guiding me whenever I required her guidance and mentorship. I would also like to thank my GEC Dr. Maria Rafiq for her support and guidance throught my research phase. I would also like to present my gratitude towards former principal of Research Centre for Modelling and Simulation (SINES) NUST, Dr. Rizwan Riaz and current principal of SINES for providing us with the collaborative and constructive environment to conduct our research project. I would highly appreciate all the faculty members at SINES, NUST, who have helped me directly or indirectly complete my thesis and study. I appreciate the support and continuous motivation of my dear friends Maleeha Ahmed, Farhana Riaz and Tayyaba Alvi. My special and affectionate gratitude is to my parents, Tahira amma, my husband and all family, the ones who can never be thanked enough for their overwhelming love, care, financial and moral support and without whose proper guidance, I would never be able to complete my higher education.

Contents

List of Abbreviations

EGFR	Epidermal growth factor receptor
Derlin	Degradation in endoplasmic reticulum protein
TMEM	Transmembrane Protein
iRhom1	Inactive Rhomboid 1
iRhom2	Inactive Rhomboid 2
UBAC	Ubiquitin-associated domain-containing protein
RHBDD	Rhomboid domain-containing protein
TMDs	Transmembrane domains
ER	Endoplasmic Reticulum
IRHD	iRhom Homology Domain
H1F1a	transcription factor hypoxia-inducible factor – 1
RACK1	Receptor of activated protein C kinase-1
ADAM17	ADAM metallopeptidase domain 1
STING	Stimulator Of Interferon Response CGAMP Interactor
TRAP	Translocon-associated protein
VISA	Virus-induced signalling adaptor
A.A	Amino acid
IRHD	iRhom Homology Domain
TOC	Tylosis with oesophageal cancer
GPCRs	G protein-coupled receptors
PMA	Phorbol 12-myristate 13-acetate
3D	3 Dimension
BLAST	Basic local alignment search tool
AMBER	Assisted Model Building with Energy Refinement
CHARMM	Chemistry at Harvard Macromolecular Mechanics
GROMACS	GROningen MAchine for Chemical Simulations
BER	Brendnsen Thermostat
NMR	Nuclear magnetic resonance

List of Figures

List of Tables

Abstract

The iRhoms are part of the rhomboid family, highly conserved among all sequenced metazoans. However, they are known as pseudoproteases as they lack vital amino acid residues required for the catalysis of serine proteases. The iRhom2 consists of a large cytoplasmic N-terminus and an inactive catalytic domain of an unknown function. iRhom2 is an unstable protein, which performs a crucial part in regulating the EGFR pathway by the maturation of TNF-α converting enzyme, TACE or ADAM-17. The up and downregulation of the EGFR pathway by changing iRhom2 can lead to several human diseases, including breast cancer, Alzheimer's disease, TOC and many others. The stability of iRhom2 can be increased via some specific mutations induced at the N-terminus. The deletion of 268 amino acids from the N-terminus results in a gain of function mutation in iRhom2; this mutation is known as cub-mutation, observed in mice. The cub-mutated iRhom2 can hyperactivate or downregulate the EGFR pathway. The hyperactivation leads to increased wound healing, inflammation and enhanced tumorigenesis, thus making this pseudoprotease more stable than wild-type iRhom2. However, the downregulation of the EGFR pathway cannot make cub-mutated iRhom2 stable than wild-type iRhom2. The contradiction about the stability of cub-mutated iRhom2 is investigated in this study. The 3D protein structure prediction, protein structure evaluation and MD simulation techniques were used to solve this contradiction. Ab-initio protein modelling technique was used to predict cub-domain (1-268) and complete cytoplasmic N-terminus domain (1-351) of iRhom2 for comparative analysis of both domains. The evaluations of these predicted domains' structures by various evaluation methods result in possible accurate models. In addition, the MD simulation of both iRhom2 domains was completed to check the structural accuracy and stability of these domains. The MD simulation results showed that the complete cytoplasmic N-terminus domain is modelled as a less compact and unstable structure. However, the cub-domain of iRhom2 resulted in a more stable and compact structure than the complete cytoplasmic N-terminus domain. The contradiction about the stability of cub-mutated iRhom2 is resolved; however, extensive insilico and in-vivo research are required to explore the functional stability of iRhom2 further.

Introduction

1.1 Rhomboid Superfamily

Rhomboids were known as a family of contentious intramembrane proteases; after 19 years of research, these proteases are still anonymous. This unique family of proteases is conserved in all eukaryotes and prokaryotes. Being intracellular proteases, rhomboids were challenging to identify in the first place. They were later identified by the detailed analysis of the Epidermal growth factor receptor (EGFR) by studying *Drosophila melanogaster*. EGFR was known to regulate crucial intercellular signalling functions. They were reported to be responsible for the EGFR regulation in drosophila and tumour necrosis factor-alpha (TNF-α) in mammals [1]. The rhomboid superfamily consists of six transmembrane domains with a catalytic serine on one transmembrane domain and histidine on the other to make it an active serine protease Figure 1.1.

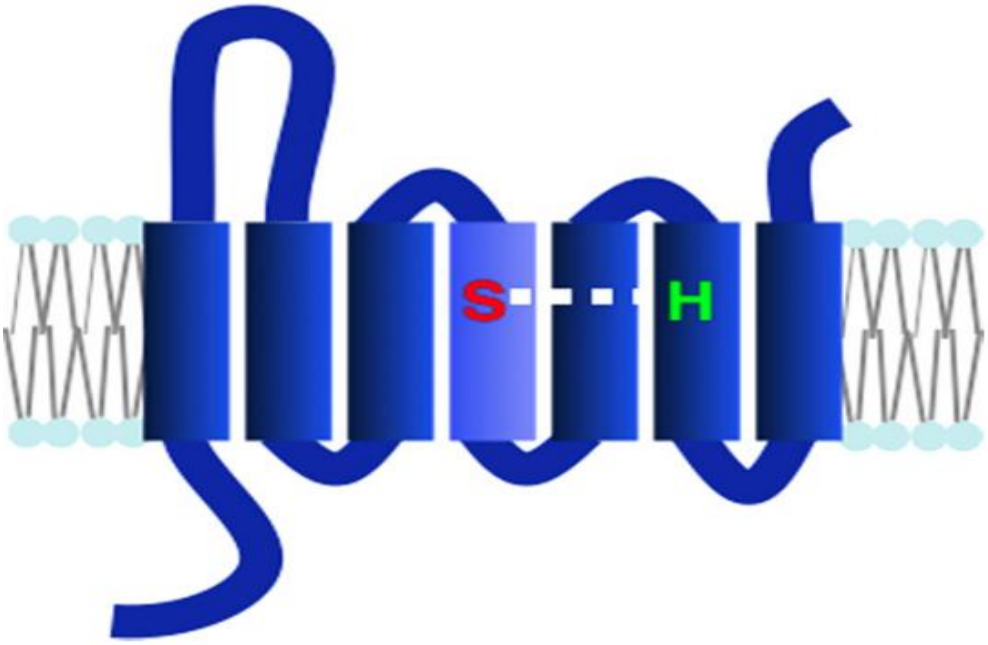

Figure 1.1 Schematic picture of a rhomboid protease - intramembrane serine protease with serine shown in red and histidine in green. The domain with catalytic serine is shorter than the rest. [2]

Later on, different mechanisms of rhomboids were discovered, including the Proteolytic cleavage of polypeptide chains of other membrane-anchored proteins. [2] Rhomboid proteases are of great interest due to their diverse biological role and their involvement in human diseases such as cancer, parasite infections, skin diseases, and many others. Because of their importance, rhomboids were identified as a potential target against many diseases [3]. At present, there are 14 individuals from the rhomboid family, five are rhomboid proteases, and nine are homologous to a rhomboid family. Classification of the rhomboid superfamily is shown below in ***Figure 1.2***.

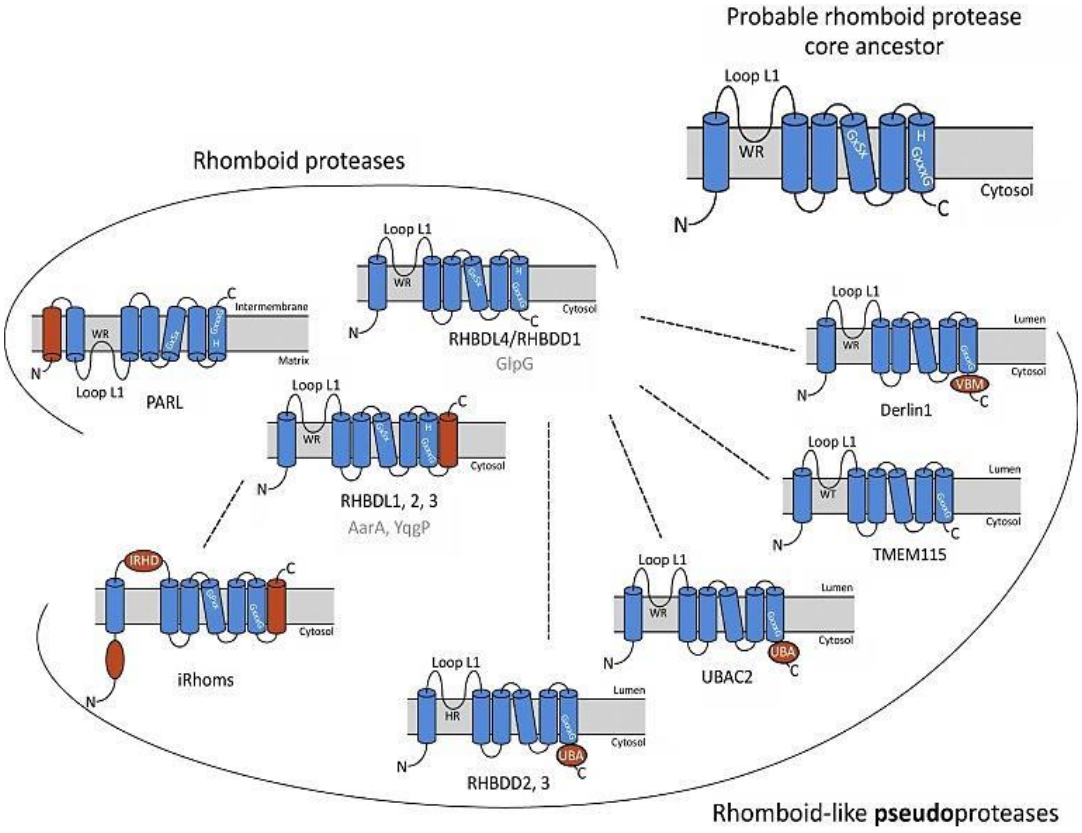

Figure 1.2 Rhomboids classified as a diverse family of active and pseudoproteases, including degradation in endoplasmic reticulum protein 1 (derlin1), Transmembrane Protein115 (TMEM115). Ubiquitin-associated domain-containing protein 2 (UBAC2), Rhomboid domain-containing protein 2 (RHBDD2), and Inactive Rhomboids (iRhoms)[4]

This family of proteases varies from RHBDFs to Inactive rhomboids (iRhoms) and Darlins, where six ancestral transmembrane domains (TMDs) are conserved among all

family members (shown in blue). For RHBDLs, the protease active site motifs ‘GxSG’ and ‘H’ form a catalytic dyad between TMDs helices four and six, whereas iRhoms have a ‘GPxG’ sequence. Unique domains like the iRhom homology domain (IRHD) are highlighted in red [5].

1.2 Rhomboid like Pseudoproteases

The members of the Rhomboid family are known as inactive because of the lack or absence of critical active residues present in core protein. iRhoms, Derlins, UBAC2, TMEM115, and RHBDDs belong to the pseudoprotease sub-class of the Rhomboid Superfamily. The genes responsible for the regulation and transcription of these inactive proteases are well preserved among many species. It is assumed that pressure exists to conserve these inactive proteins with significance for life in a proteolytic-independent way. It is observed that if normal rhomboid proteins become catalytically disabled, they do not act like inactive rhomboids (iRhom), suggesting that other domains present in rhomboids like pseudoproteases are crucial for their survival and proper functioning *(Luo et al*., 2016a; Luo and Shu, 2017).

1.2.1 Inactive Rhomboids

iRhoms lack essential amino acids required for the catalytic activity of rhomboid superfamily members. iRhoms contain seven transmembrane domains (TMDs), predominantly located in the endoplasmic reticulum (ER). Usually, with rhomboids, serine residue at the N-terminus makes it able to perform the catalytic tasks. However, in iRhoms, serine is replaced by proline which makes it a pseudoprotease. (i.e. GxS was replaced as GxP) Figure 1.3. iRhoms also has a long cytoplasmic N-terminus and highly conserved, cysteine-rich, luminal domain named ‘iRhom Homology Domain’ (IRHD). This unique domain is responsible for linking Transmembrane domain 1 (TMD1) and Transmembrane domain 2 (TMD2) [3].

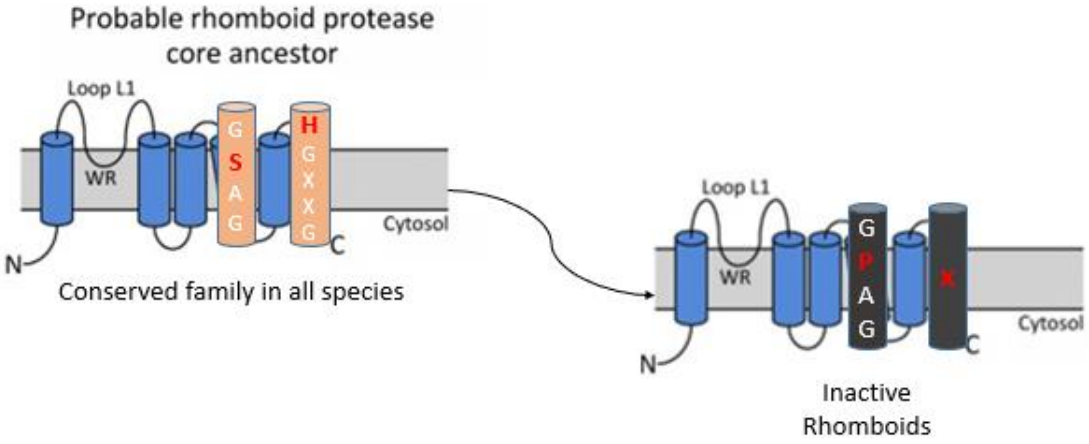

Figure 1.3 Replacement of active residues from Rhomboids to iRhoms. A schematic of how active residues are swapped from rhomboids andj transform into iRhoms. The 'ser' residue in the 4th domain and (Ha, Akiyama and Xue, 2013)

1.3 Role of iRhoms

iRhoms might have lost their catalytic activity in the process of evolution, but they kept the crucial non-protease functions. The first understanding of the physiological functions of iRhoms was clear after studying *Drosophila Melanogaster* genetically. iRhoms control Epidermal Growth Factor Receptor (EGFR) by degrading EGF-like ligands using a crucial protein quality control mechanism, i.e., endoplasmic reticulum-associated degradation (ERAD). TNFa signalling pathway is also controlled by iRhoms [9]. The function of iRhoms in mammals is to degrade similar proteins, thus proving that this function of EGF-like ligands is well-preserved [10]. Although the exact link of iRhoms specifically regulating EGRF only is yet to be discovered. Moreover, iRhoms are reported to be involved in protein turnout and stability. iRhoms are also responsible for stabilising protein turnover by regulating client proteins. iRhom1 expression is observed in many tissues, but the expression of iRhom2 is minimal. The protein turnover of iRhom1 is discussed in Table 1.1.

Table 1.1 Functions of iRhom1 in different species

Organism	Localisation	Function	Reference
KO Mouse	Different tissues and cells	Severe phenotype shows damaged tissues and organs. The weaker phenotype has shown less damaged tissues due to the difference in genetic background.	[11] (Li, Maretzky, Weskamp, Monette, Qing, P. D. A. Iss*uree, et al*>, 2015)
Human and flies	Endoplasmic Reticulum (ER)	Regulate the activity of the proteasome under endoplasmic stress. IRhom1 regulates the turnover of cytoplasmic proteins. IRhoms carries out the oxygen-dependent degradation of transcription factor hypoxia-inducible factor – 1 a (HIF1a) by using the receptor of activated protein C kinase-1 (RACK1).	[13] [14]
Multiple species	ER and Golgi	ADAM 17 dependent regulation of the (EGFR) pathway by working mutually with iRhom2 Regulation of the catabolic process of proteasomal protein	(Siggs *et al*>, 2014; Li, Maretzky, Weskamp, Monette, Qing, P. D. A. Issuree,>*et al*>, 2015)[15]
Human	Cancer cells	Cell migration and cell population proliferation Negative regulation of protein secretion	[16]

iRhoms complete many biological tasks by working co-dependently on each other. For example, the regulation (maturation) of ADAM metallopeptidase domain 17 **(**ADAM17) is solely controlled by iRhom1 and iRhom2, which demonstrates that iRhoms trigger all functions of ADAM17, not only the release of TNFa but also they are involved in the release of many EGFR ligands and many other substrates. However, mostly iRhoms are responsible for the regulation of inflammation and growth factor signalling. An example of iRhoms not being involved in the regulation of TACE or, indeed, inflammation at all is its modification of the cellular immune response of cells to DNA viruses by controlling the

protein turnover of STING (Stimulator Of Interferon Response CGAMP Interactor) (L*uo et al*>, 2016).Figure 1.4.

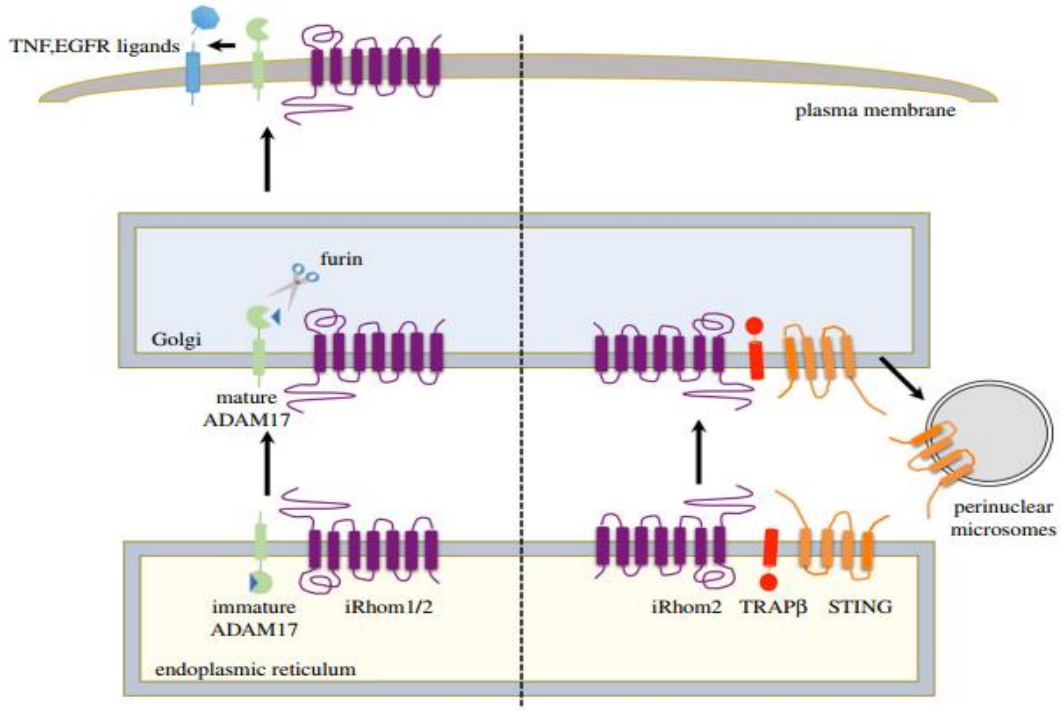

Figure 1.4 The role of iRhoms in protein trafficking. A schematic diagram to show the participation of iRhoms in regulating the membrane proteins. ADAM17 (green) and STING (orange). On the left, iRhom 1 and 2 are illustrated to show the trafficking of ADAM17 from ER to Golgi apparatus, where it later goes for furin mediated maturation by cleaving its predomain. iRhoms regulate the trafficking of STING from ER to microsomes via Golgi with the help of translocon-associated protein beta (TRAPb) (red). [6]**.**

The involvement of iRhoms in the disease is highly dependent on both of these factors. iRhom1 is reported in the cell proliferation of cancers, including breast cancer [17] and colorectal cancer [18]. In addition, both iRhoms are unexpectedly involved in heart diseases [19], [20], but iRhom2 is considered functionally more significant compared to iRhom1 in many terms.

1.3.1 Why is iRhom2 Crucial?

iRhom2 is a vital factor for the regulation of TACE. Loss of iRhom2 results in the blockage of TACE maturation, which leads to the defective shedding of TNF-a. IRhom2 binds to ADAM17/TACE to help its transportation from ER. So without the support of iRhom2,

ADAM17 is unable to leave the ER and then cannot be trafficked to Golgi apparatus to remove its inhibitory pre domain by furin [11]. Therefore in mice, the inactivation of iRhom2 leads to the inactivation of TACE simultaneously [21]. The targeting of iRhom2 effectively inactivated ADAM17 in mice immune cells without affecting its processes in other cells. The inactivation of TACE will stop there the usual immune response of the cells and normal cell proliferation. The protective function of iRhom2 with TNF-a has shown that both mice and humans have conserved ADAM17/TNF-a pathway. Therefore, iRhom2 is considered a crucial novel target for the treatment of EGFR/TNF-a dependent pathologies [9]. Different KO studies have been performed to have a better understanding of the catalytic activity of iRhom2. In mice, severe phenotype changes (excessive wound healing, increased inflammation and accelerated tumorigenesis) have been observed compared to other rhomboid superfamily members. [9], [22], [23]

1.4 Mutations in iRhom2

Most mutations occur at the highly conserved area of the iRhom2 cytoplasmic domain; one of these mutations is responsible for the tylosis with esophageal cancer (TOC), which is a rare inherited syndrome (Saarinen *et al.*,; Ellis *et al.*, 2015). These mutations are specifically lead to an increase in the activity of TACE and constitutive shedding of EGFR ligands [26] (Table 1.2). Nevertheless, the systematic importance of the four a.a region (figure 1.4) where all TOC mutations happen is still unclear. Nevertheless, there is some molecular insight into the regulatory functions of the N-terminal domain due to the mutations. The simulation by GPCRs (G protein-coupled receptor) or PMA (Phorbol 12-myristate 13-acetate) at the plasma membrane are responsible for the phosphorylation of the iRhom2 N-terminal cytoplasmic domain. The phosphorylation in the well-defined sites leads to the addition of 14-3-3 Proteins [27]. The recruitment of these proteins is sufficient and essential to initiate the ADAM17- dependent shedding [27]. The defective mutants of phosphorylation still support constitutive shedding by TACE/ADAM17, which shows that phosphorylation controls precisely the simulated shedding of ADAM17 substrates.

Many mutations occurring at the N-terminal cytoplasmic domain-encoding region of iRhom2 are conserved in mammals. Several knockout studies have been performed to

determine the exact or proper response of these mutations on the overall functioning of iRhom2. In mice, another mutation was reported as an "uncovered" mutation [28]. iRhom2*Uncv* leads to the sudden hair shaft and inner root sheath differentiation in mice, which results in a hairless phenotype (Table 1.2). As compared to wild type Protein, iRhom2*uncv* in mice showed hyperkeratosis and hyperproliferation in the epidermis along with hypertrophy of sebaceous glands.

Recently a new substrate of iRhom2 was identified, further defining the molecular details of its regulation. A protein named FRM8 or iTAP (FERM Domain containing 8) binds with iRhom2 between amino acids 200 to 300 [29], [30]. FRMD8 is essential for the maturation of TACE/ADAM17, and it has also been known to stabilise both members of the sheddase complex at the cell surface [29], [31]. Therefore, it is crucial to observe the bonding of iTAP with iRhom2, which seems independent of 14-3-3 proteins, but there is no confirmation about any relationship between phosphorylation-dependent 14-3-3 binding and stabilisation of the complex by iTAP [29], [31].

Table 1.2 Mutations in iRhom2 and their effect on ADAM17 [32]

Mutations in iRhom2	**Effects on ADAM17 activation**		
	Maturation	**Constitutive Shedding**	**Induced shedding**
Δ(N-terminus deletion)	Reduced	Increased α No difference	Reduced
Δ IRHD	Reduced	N.A.	N.A.
Cub	Reduced	Increased β Reduced	Reduced
Tylosis+	Increased	Increased Increased α	N.A.
Uncovered	Reduced	Reduced	N.A.

αIncreased TNFR shedding.

βGreater levels of AREG secretion independent of TACE activity

1.4.1 iRhom2 and Disease Relevance

iRhom2 is considered a vital target for cancer and others disorders because iRhoms are critical results of ADAM17, primarily mediated by TNF-A signalling. ADAM17 have a crucial role in regulating many inflammatory diseases. However, other pathological disorders are prepared to be linked with iRhom2, including neurodegeneration, infections, skin and heart diseases. In some cases, iRhoms are reported to be directly linked with many diseases, but, primarily, iRhoms are considered to be indirectly linked with an extensive range of physiological and pathological processes [32] (Table 1.3)

Table 1.3 The pathologies and physiological effects of loss of function of iRhom2 in mouse and human tissues together with relevant clients [32]

Disease	**Phenotypic readout**	**Client**
Tylosis with oesophageal cancer (TOC)	• Palmoplantar hyperkeratosis, increased risk of oesophageal cancer • Adenoma formation and decreased survival • Complete hair loss of mice at birth • Increased wound healing	EGFR ligands
Gastric Cancer-associated fibroblasts	Diffuse type gastric ulcers	TGF-b1, ADAM17
Neurological disease	Alzheimer's (speculated)	n.a.
Acute lung injury after intestinal ischemia-reperfusion	Reduced Apoptosis	ADAM17, TNF-α
Renal dysfunction	Significant protection against tissue inflammation, kidney damage	Reduced ADAM17, EGFR

1.5 Problem statement

Hyper-activation of EGFR due to excessive release of AREG by cub mutated iRhom2 may cause different types of diseases. However, unavailability of its 3D structure, the stability of this mutated pseudoprotease is poorly understood. Therefore, 3D comparative modelling, MD simulations and Protein-Protein interactions of iRhom2 can determine its mechanism.

1.6 objectives

- To check similarity of Cub iRhom2 with Human iRhom2 by using multiple sequence alignment techniques.
- To predict the 3D structure of iRhom2 and its cub mutated isoform by using Ab-initio and threading based modelling approaches.
- To examine the stability of iRhom2 in comparison with cub mutated isoform by Molecular Dynamics Simulation and protein-protein interaction to examine the binding affinity of amphiregulin with cub-iRhom2 and wild-type.

Literature Review

2.1 iRhom2

iRhom2 is one of the inactive rhomboid proteases in humans which do not have the principal catalytic motif present in active rhomboids [33]. It is thought that iRhom2 has lost its catalytic function during evolutionary changes. Although it is implicated to have non-protease functions in regulating EGF and TNF-α signalling pathways [34]. iRhom2 has some key features, having pathological and physiological roles in variety of organism i.e. involved in breast cancer [35], ovarian cancer, tylosis with esophageal cancer, colorectal cancer [32], alzheimer disease [9] and many skin diseases [36].

2.2 iRhom2 Domain Structures and Functions

The easier way to understand the structure of iRhom2 is by breaking down the protein into various domains. Each domain has a functional significance. iRhom2 is a large protein that consists of 856 amino acids (a.a). The structure of iRhom2 is divided into these modules; Transmembrane, luminal and N-terminal domains.

2.2.1 Transmembrane Domain

The main characteristic of the rhomboid-like superfamily is the conserved 6-TMD core which is also known as a "rhomboid-like" like structure. The difference amount this group of proteins derives from some of them having a 7th TMD like iRhom2 [3]. The primary function of these six conserved TMDs is the recognition of other substrate TMDs [4]. In the case of mouse iRhom2, 1387F sinecure mutation (point mutation)(Figure 2.1) in the first TMD predicts that the Transmembrane helix will be titled in the membrane lipid bilayer [37]. It is assumed that this might provide the interaction site for the substrate to bond with client TMDs. VISA (virus-induced signalling adaptor) is a Transmembrane protein that requires the first TMD [38]. In the case of STING interaction, the first TMD has also been observed to be crucial [6]. Overall, it is clear that the first TMD of iRhom2 is essential, but other domains are also vital.

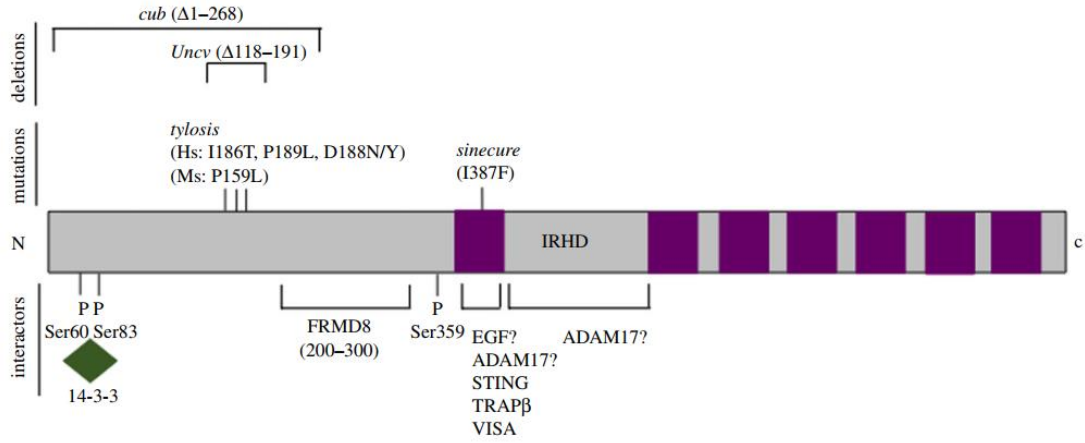

Figure 2.1 The structural design of iRhom2 Domains with its phosphorylation sites and interactors. TMDs are shown (purple) along with the highly conserved IRHD and the long cytoplasmic N-terminal. This cytoplasmic tail is responsible for some functionally crucial 114-3-3 binding sites (green). This architecture also shows various known mapped mutations on the N-terminal. These mutations are reported to be involved in human disease and proposed binding sites for known interactors of iRhom2 [32]

2.2.2 IRHD (iRhom Homology Domain)

The extracellular loop domain between two TMDs 1 and 2 is called IRHD. This domain is somewhat mysterious. IRHD is the most highly conserved domain among iRhoms, i.e. 65% between iRhom1and iRhom2. However, the function of this domain is still unclear (‘Adrain C, Freeman M. 2012; Lemberg and Adrain, 2016). The total length of IRHD is 230 a.a, and it contains 16 conserved cysteine residues. These residues were predicted to form disulphide bonds, which depicts a complex 3D structure. However, the 3D structure of IRHD (or any other iRhom) has not been solved yet because it does not resemble any other known domain. Furthermore, the Deletion of IRHD from iRhom2 prevents the maturation of Adam 17, which leads to the decreased binding to ADAM17 (immature) in the ER [27] (Figure 1.4). iRhom2 without IRHD was also detected at the same level on the cell surface as wild-type iRhom2 [27], which concludes that for the trafficking of iRhom2, IRHD is likely to be unnecessary.

2.2.3 N-terminal Cytoplasmic Domain

One of the most prominent features of iRhom2 is its long N-terminal cytoplasmic domain. This domain has many crucial but undefined functions. This specific domain

is proven to be very complicated. It contains both positive and negative regulatory elements for the proper functioning of iRhom2. This domain is made up of almost 350 a.a, and it is the least conserved domain among iRhom1 and iRhom2 (45%), which indicates the difference in function between both iRhoms [32]. Furthermore, most of the cytoplasmic N-terminal is highly prone to mutations. These mutations are significant to determine the behaviour of iRhoms in normal and diseased conditions.

2.2.4 Role of iRhom2 in TNF-α

iRhom2 is upstream regulator of TACE (TNF-α converting enzyme) [34] also recognised as ADAM17. Several experiments have been performed on the flies and mice with mutated genes of iRhom2, exposed its indirect role in the discharge of TNF and EGF family ligands inactive form by directly affecting the breakdown associated with the endoplasmic reticulum (ER) and trafficking of TACE from ER to the plasma membrane [34].

ADAM17 or TACE is a metallo-protease that is responsible for the release of some membrane-bound proteins. ADAM17 is also known for the primary release of TNF-α (an inflammatory cytokine) and numerous EGFR ligands. iRhom2 is responsible for trafficking and maturation of ADAM17. A recent discovery about ADAM17/iRhom2 sheddase is that FERM8 (part of ADAM17/iRhom2 sheddase complex) protein is necessary for the stability of this complex at the outer membrane of the cell. FRMD8 containing FERM domain is a poorly understood protein and not well characterized yet. iRhom2 has large cytoplasmic N-terminus, FRMD8 performs its function by attaching it to N-terminus. If FRMD8 is not attached to ADAM17/iRhom2 complex, it undergoes to lysosomal degradation. This phenomenon results in the reduced release of ADAM17 ultimately reduced the release of TNF-α and EGF receptor ligands. This study is confirmed by several experiments on human macrophages and mouse tissues hence proving FRMD8 role in the controlled release of several cytokine i.e. TNF-α and growth factor signals (EGFR ligands) [29].

TACE controls the TNF-α and EGFR pathway, both have significant roles in growth and infection [40]. TACE is essential for the discharge of TNF-α and EGFR that are key activators for inflammation [40]. Tumour Necrosis Factor (alpha) is important in

the body as ‘emergency call’ because it assists the body in combat against infectious diseases, but if mutations occur or due to some other reasons its release is uncontrollable TNF-α can cause inflammatory arthritis which is an inflammatory disease[41]. TACE is regulated by iRhom2, and they both are activated by a small stimulus in the body, i.e. pro-inflammatory mediators in blood and growth factors. Therefore iRhoms are imperative for TACE activity and show co-expression with TACE[42].

2.2.5 Role of iRhoms in EGFR pathway

Epidermal growth factor receptors (EGFR) belong to the receptor tyrosine kinase family involved in many cellular regulation processes which include in the survival of cell, differentiation and proliferation. Family of EGFR ligand incorporates TGF-α, EGF, amphiregulin (AR), epigen, heparin-binding EGF-like development factor (HB-EGF), epiregulin and betacellulin (BTC) [38]. An interesting fact of EGFR is that t plays an effective role in cancer as an activated oncogene. Overexpression of EGFR has been reported in a wide variety of cancers i.e. head and neck, breast, cervical, ovarian and lung cancers and has been associated per their deprived prognosis [43]. Whereas in case of ovarian cancer gene of EGFR is overexpressed in about 4-22% of cases [32], [44] especially in case of epithelial ovarian cancer about 13% of amplified expression has been reported [45]. Several studies have been reported the implication of iRhoms in EGFR signalling regulation pathway. EGFR signalling pathway is regulated by the active rhomboid proteases that are stimulated during wings growth and development. A strong association of rhomboid proteases in the signalling of EGFR was established using the Drosophila primordium that was sensitive developing wing to reveal the ectopic activity of EGFR.

Experiments suggest that iRhom1 co-expression with an active rhomboid (HB-EGF) are responsible for extreme wing phenotypes in Drosophila. TACE can control various physiologically important EGFR signalling pathways by releasing several ligands of the EGFR [46]. iRhoms deficient Drosophila induced sleep like phenotype, close to the phenotype detected in elevated EGFR pathway activation. These outcomes show that iRhoms of Drosophila are associated with inhibition of EGFR signalling regulation pathway. ADAM17/TACE is also responsible for the release of different EGFR ligands. In this way, TACE can control a varied range of therapeutically and physiologically

roles that were significant in EGFR signalling [33], [47]. Studies showed the involvement of iRhom1 and iRhom2 in the EGFR regulation signalling in mice, featured iRhoms significant role in the signalling pathway of EGFR [3], [34]. Human iRhom1 or mouse iRhom2 co-expression with ligands of EGF family in cells of COS7 involved in the downregulation of all the ligands of EGFR [48]. In another way, some TACE substrates release by the rapid stimulation of mutant embryonic fibroblasts of iRhom2 resulted in the downregulation of amphiregulin, HB-EGF and shedding of epiregulin, however, TGF-α shedding was not altered. In the case, the level of mature TACE was not changed, suggesting the involvement of iRhom2 itself in determining TACE-dependent substrate selectivity detaching [3], [33].

Silencing for the gene of RHBDF1 through siRNA, in cell lines of cancer decrease the cell movement level and multiplication and auto phagocytosis or induced apoptosis in cells of cancer [43]. Furthermore, iRhom1 is essential for the survival of epithelial cancer cells in humans and might be associated with GPCR-mediated transactivation of EGFR [10], [49]. So, these outcomes show that iRhoms were not just promoting EGFR ligands forward trafficking from ER to Golgi, yet in addition inhibit ER transfer of the EGFR ligands through proteasome with the help of ERAD [39] and iRhoms that might be effective targets for treating TACE/EFGR-dependent pathological issues [12]. iRhoms are suggested to be upstream regulators of TACE (TNF-α converting enzyme) [34]. However, it may also be known as ADAM17 (ADAM metallopeptidase domain 17) which also involves in different diseases. Several experiments have been performed on the flies and mice with mutated genes of iRhoms, elucidated their indirect role in the discharge of TNF and EGF family ligands inactive form by directly affecting the breakdown associated with the endoplasmic reticulum (ER) and trafficking of TACE from ER to the plasma membrane [11]

In contrast to an early postulate that the iRhoms are directly involved in active rhomboid proteases inhibition [46], it decreases the level of substrates in growth factor through activating their degradation. Moreover, iRhom1 in human and iRhom2 in mouse have been revealed that signalling pathway down-regulation of EGF through binding of ligands of EGF in the ER and directing them for ERAD, which triggered because of quality control framework of ER [50].

In the light of all these facts, it is concluded that iRhom-mediated ERAD in EGF signalling pathway regulation of EGF family ligands is shared by Drosophila and mammals as well but there is no strong proof for iRhoms physiological role in ERAD regulation in mammals.

2.2.6 Cub-Mutation and Stability of iRhom2

The curly bare mutation (cub) is reported as a genomic deletion of most of the N-terminus cytoplasmic domain in mouse iRhom2 (Hosur et al., 2014; Owen M. Siggs et al., 2014). This mutation is a spontaneously occurring mutation in the mouse, which results in a loss of hair phenotype, wound healing and tumorigenesis. The cub deletion removes the first 268 amino acids from the iRhom2 cytoplasmic N-terminal domain, but it is not a loss of function mutation. During normal state (wild-type) iRhom2 induce migration of primary mice keratinocytes [38], it promotes the survival of human squamous epithelial cancer cells, it helps in the clearance of the misfolded protein from endoplasmic reticulum membranes in mammalian cell lines [39] and also the ADAM17 dependent regulation of substrate selectivity for the metalloprotease shedding in mouse embryonic fibroblasts (MEFs) [38], [40].

iRhom2 is a short-lived protein, which means that it is unstable because, after the secretion of an EGFR ligand, i.e. AREG, iRhom2 degrade via lysosome (stabilising mature ADAM17). However, cub-mutated iRhom2 tends to hyperactivates the EGFR pathway by stimulating the secretion of AREG, independent of ADAM17. This hyperactivation of the EGFR pathway thus making cub-mutated iRhom2 more stable, which leads to excessive cell proliferation and enhanced tumorigenesis. This hypothesis makes cub-mutated iRhom2 a vital therapeutic target [22]. Although there exists some conflict about this hypothesis because this hypothesis was tested by [11] disagreement, their results showed that the EGFR pathway is downregulated by cub-mutated iRhom2. Furthermore, they observe that cub mutated embryonic fibroblasts release less AREG, which makes them less able to bind with ADAM17 for it is maturation [37]. Further studies are required to clarify this disagreement.

3 Methodology

3.1 Global Alignment

Global alignment is a "global optimisation" that pushes the alignment across the entire query sequences' length. Needleman and Wunsch's dynamic programming algorithm is used for global alignment, which is a slow but accurate method.

The hypothesis for this project was based on the similarity between the sequence and structure of mammalian iRhom2. The stability of iRhom2 in mice was studied in mice. To study a similar concept for Human iRhom2, global sequence alignment was required to identify similarities among mammalian proteins for further analysis [51], [52].

3.2 3D Structure prediction

The interpretation of the 3D structure of a protein from its primary structure for the prediction of its folding into the secondary and tertiary structure is known as computational protein modelling. Protein structure prediction is one of the primary goals developed by the Fields of bioinformatics and biochemistry; 3D Protein structure prediction is crucial for drug designing and biotechnology. For quality assurance, every two years, the existing servers for 3D structure predictions are tested by CASP (Critical Assessment of Techniques for Protein Structure Prediction) [53], [54].

In this project, the 3D structure prediction was required because no crystal structure of iRhom2 is solved yet (both mouse and human) or any inactive member (pseudoproteases) of the rhomboid superfamily [32]. iRhom2 being a protein with transmembrane domains, is challenging to model because transmembrane proteins are proven to be problematic to study because of their flexibility, partially hydrophobic surfaces and lack of stability [55]. It was not feasible and crucial to predict the complete iRhom2 structure because the research question of this project is based on some vital point mutations [36], IRHD and complete cytoplasmic N-Terminus. The remaining rhomboid family conserved part of iRhom2 was not involved in its structure stability profiling [32]. 3D protein modelling has

three types, i.e., Homology (comparative) modelling, threading (fold based) and ab-initio (De-novo). However, for this project, the most suitable method of 3D structure prediction is ab-initio or threading-based protein modelling, as the BLAST search results of iRhom2 showed a similarity level between iRhom2 and its homologs in PDB is less than 18%; no homologue was found. [32].

3.2.1 Robetta

De-novo based or ab-initio structure prediction servers usually produce decoys and choose the best possible confirmation, considering them based on the thermodynamic stability and energy state. The best possible confirmation should have the least entropy and free energy. Ab-initio protein structure prediction is a computationally expensive and extensive method [54].

Robetta is one of the servers, which offers both comparative and ab-initio modelling. Robetta parses the query sequence into putative domains and decides between the Comparative and De-novo methods of structure prediction based on the availability of homologues in the PDB [56]. Ab-initio protein structure prediction was used for this project, as no homologue of iRhom2 human was found.

3.2.2 I-TASSER

I-TASSER (Iterative Threading Assembly Refinement) is a server for Protein modelling and structure-based function annotation. I-TASSER searches and identifies similar structural folds or multiple threads by using LOMETS. I-TASSER constructs full-length atomic models. The I-TASSER prediction accuracy is based on the relative clustering structural density scoring function (C-score) and consensus significance score of different threading templates. The best structure among all predicted structures will be selected based on RMSD, c-score and TM score (a structural similarity measurement value between 0,1). I-TASSER was rated the best 3D structure prediction server by the CASP experiment. [57], [58].

For this project, both Robetta and I-TASSER were used to generate structures for wild-type iRhom2 (1-856 a.a) and cub-mutated iRhom2 (287-856 a.a). The optimal model will be selected after the structure evaluation.

3.3 Structure Evaluation

It is crucial to select the best possible confirmation of a structure from the candidate models' pool generated by different structure prediction methods. [59].

3.3.1 SAVES

SAVES server is a meta server, which runs six programs simultaneously to check and validate 3D protein structures before and after model refinement. The programs used for this project are mentioned below

3.3.2 ERRAT

ERRAT is based on a novel procedure to differentiate correct and incorrect conformations. Different protein model building errors add different atoms at random positions; these errors can be distinguished from atoms' optimal distributions by using a quadratic error function. This function differentiates the collection of pairwise interactions from 9 residue sliding windows in a database of more than 90 reliable 3D protein structures. Misregistered parts of the target protein can be spotted by studying the pattern of non-bonded interactions from each window [60].

3.3.3 Verify3D

Calculates the compatibility of a 3D Protein model with its amino acid sequence by assigning a structural class based on the position and environment (polar, nonpolar, alpha, beta, loops) and compare the target (result) to the optimal models [61].

3.3.4 PROCHECK

PROCHECK analyses the stereochemical properties of the target protein structure in detail. The output is shown in plots (Postscript format) and comprehensive residue by residue listings [62], [63].

3.3.4.1 Ramachandran Plot

Procheck qualify proteins by generating different plots including Ramachandran plot shows stereochemistry and geometry of protein complex by establishing that no part of the molecule is present in the electrostatically unfavorable region. [64]

3.3.5 WHATCHECK

This server is originated from the subset of many protein evaluation tools from WHAT IF program. This server evaluates the target protein model by checking its stereochemical properties extensively [65].

3.4 Molecular Dynamics Simulations

Molecular dynamics (MD) is a molecular mechanics program intended to imitate atoms' movement within a molecule. MD can be done on a molecule to produce altered conformation upon energy minimisation, giving a range of regular conformations. The MD method is used to calculate the molecular motion of several particles that are interacting classically. In MD, motions of molecules are within some specific time frame, i.e. one-time step (fs); atoms have some velocities and are subjected to some forces [5]. The Atomic representation is the one that leads to the as closest reproduction of the existing systems [66]. The MD simulation is used to observe the conformational changes in molecules by assimilating Newton’s second law of motion (F=ma). The motion of a single molecule or a large number of molecules can be studied by this method. Therefore, by using MD, the stability and flexibility of the system can be evaluated. [66].

After evaluating both structures, i.e., wild-type iRhom2 and the cub-mutated iRhom2, the next step was MD for the stability profiling of these structures.

3.4.1 Molecular Operating Environment (MOE)

For MD, Molecular Operating Environment (MOE) software (2015 version issued by the Chemical Computing Group) was used. This software is a complete package that offers different services for molecular modelling, pharmacophore discovery, docking, and simulations. For the MDs of both wild-type and cub-mutated iRhom2, MOE was used. MD was performed by selecting some specific parameters (force field) [67].

3.4.2 Force Fields

To create a realistic environment for every atom of the system to be simulated, the forces acting on each atom will be obtained by deriving equations. These equations are known as the "force fields" The equations are complex but simple to calculate. [66], [68] Force Fields represents molecular features of a system as:

- Springs (bond length and angles)
- Periodic function (bond rotation)
- Lennard John potential
- Coulomb's law (van der Waals)
- Electrostatic interactions

Currently, used force fields in atomistic MDs differ in the way they are parameterized. However, that is how a typical force field expression might look like (Figure 3.1)

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

Figure 3.1 The expression shows the contribution of bonded atoms, angles, torsions, non-bonded atoms, lennard john potential, and electrostatic forces towards the overall energy of the system (Equals to "U, i.e. overall energy of the system") [69]

After calculating the force field, Newton's law of motion is used to calculate the velocities, accelerations, and current position of the atoms [67]. The most popular force fields are:

- Assisted Model Building with Energy Refinement (AMBER) [70]
- Chemistry at Harvard Macromolecular Mechanics **(**CHARMM) [71]
- GROningen Molecular Simulation **(**GROMOS) [72]
- Optimized Potentials for Liquid Simulations-All atom (OPLS-AA) [73]
- Berendnsen Thermostat (BER) [74]

3.4.3 GROMACS

MD simulations were performed using GROningen MAchine for Chemical Simulations (GROMACS) v5.1.0, following GROMACS tutorial-1 (Protein in water)[75] while utilising the parameter files provided in the tutorial, with the OPLS-AA force field for all simulations. Periodic boundary conditions were used. Newton's equations of motion were incorporated through the leap-frog algorithm. [76]. A 2fs (femtosecond) integration time step was utilized, the temperature was sustained at 300K (kelvin) using a Nosé-Hoover thermostat with a coupling time constant of 1.0ps (picoseconds). The rhombic dodecahedron system box was used, which is ~71% volume of a simple cubic box. For Van der Waals (vdW) and electrostatic interactions, cutoffs of 1.0nm (nanometer) were applied. All systems were minimized and then equilibrated for a total of 100 ps, including NVT (constant number, volume and temperature) and NPT (constant number, volume and pressure) with the Berendsen weak coupling method. These parameters were used following the GROMACS tutorial-1 (Protein in water), making them grossly generalized and applicable on average to any MD simulation [75].

To perform molecular dynamics simulations, a Linux environment is preferred; hence Ubuntu v2.02 was installed. The GUI used to manipulate input files were PyMol and MOE [77], similarly to visualize the VMD v1.9.3 [78] was used. The simulations were carried out on the supercomputing facility of the research and education centre (ScREC) at Research Center for modeling & Simulation (RCMS) NUST.

3.5 Analysis of MD Simulation

After the completion of the MD simulation, a complex form of information is generated. Cartesian coordinates of every atom of the system are recorded at each time step, and the whole process of MD simulation includes thousands to millions of steps. [78]. Thus, further analysis is required to interpret and understand the results of MD simulation. This research was based on the MD simulation of “protein in water” using GROMACS. Different parameters were selected and interpreted for this research [79], [80].

3.5.1 System Parameters

System parameters include different parameters to interpret the whole system's properties (including the water molecules, protein molecules) and different electrostatic forces acting on that system. [81]

3.5.1.1 Pressure

MD simulations of small flexible systems often require accurate local pressure calculations. It is necessary since the peripheral forces acting on the soft inhomogeneous protein system—for example, interfacial free energies calculations, osmotic pressure gradients measurements and assessments of coarse-grained hydrodynamic theories [82].

MD simulation for this study requires equilibration of pressure. Using GROMACS, this step was done under NPT (isothermal-isobaric) ensemble, whereas temperature, number of particles and pressure were kept constant [75].

3.5.1.2 Density

The density of a protein (mass divided by volume) is a physical property. Density calculation is crucial for MD simulation because it is easier to calculate. After all, protein density and packing of amino acids are the critical components of many biochemical and biophysical properties [83].

Density measurement for MD simulation falls under NPT equilibration (like pressure). For this project, density estimation is done to check the stability of the density across the process of MD. Desirable results expect density values to stabilize over time [75].

3.5.1.3 Potential Energy

In order to determine the correct molecular arrangement in space, energy must be minimized because the drawn chemical structures are energetically unfavourable. The potential energy of molecules contains various high-energy components, such as stretching, bending, and torsion. Therefore, when the energy minimization program is executed, it immediately reaches the local minimum energy value. The energy minimization may stop after the first stable conformer that is structurally closest to the original molecular arrangement is found. At this time, it is defined as the local minimum of energy, and changes in structure result in low-level changes in energy. Therefore, minimization can be stopped [84].

Potential energy minimization (PEM) (Yingling, no date) was required to observe structure stability for this project. This technique required eliminating high energies in both predicted models and achieving global minima (close to native structure).

3.5.1.4 Radius of Gyration

The radius of gyration or gyrate describes the overall spread of the molecule. In the case of proteins, the radius of gyration measures the protein's overall compactness [85]. The radius of gyration is calculated in GROMACS by summing up the protein centre into total mass w.r.t. [75]. For this project, calculating the radius of gyration was crucial to check the folded protein structures' stability.

3.6 Structure Comparison

After the MD simulation and evaluation of both structures, it was crucial to check the structural variability of cub-mutated iRhom2 from wild-type iRhom2. 3D Structure superimposition technique was used to fulfil this task.

3.6.1 Chimera

UCSF Chimera is a package of different integrated sequence-structure analysis tools. MatchMaker is a package of chimera which was used to superimpose both of the structures. MatchMaker first creates pairwise sequence alignment of protein structures, then fit the paired residues. Residue type and secondary structure information may be used to perform initial sequence alignment. Later structural-sequence alignment is performed to find similar folds [86].

Results

This study was focused on the 3D structure prediction of cub-mutated and wild-type iRhom2 stability profiling. Different types of computational techniques were applied to accomplish these goals; the following are the outcomes of different steps of methodology:

4.1 Global Alignment

According to the literature, the theory of stability of cub-mutated iRhom2 is based on mammalian iRhom2, specifically mice. Therefore, the research question was based on the hypothesis of whether the cub-mutated iRhom2 in humans shows stability as observed in mice [22], [34]. This hypothesis is based on the similarities (sequence and structure) observed among mammalian iRhoms [9]. Therefore, it was essential to check the sequence similarity of mouse and human, using global alignment. RHDF2_MOUSE and RHDF2_HUMAN sequences were extracted from UniProt [87], both sequences are then submitted to EMBOSS Needle (Needleman-Wunsch Global align) module. Figure 4.1 illustrates the global alignment result.

```
#=======================================
#
# Aligned_sequences: 2
# 1: RHDF2_MOUSE
# 2: RHDF2_HUMAN
# Matrix: EBLOSUM62
# Gap_penalty: 1
# Extend_penalty: 1
#
# Length: 865
# Identity:     763/865 (88.2%)
# Similarity:   794/865 (91.8%)
# Gaps:          47/865 ( 5.4%)
# Score: 4112
#
#
#=======================================
```

Figure 4.1 Global alignment results: Sequence identity and sequence similarity with gaps were shown in percentages and overall count.

Sequence identity (homologue) is 88.2%, which is considered an appropriate score value in this scenario. High sequence identity implies the common evolutionary ancestor, leading

to the same physicochemical properties among both species [88]. Rhomboids are common ancestors for derlins and iRhoms [9], [32]. Therefore, sequence identity and sequence similarity (same amino acid unit) between human and mouse iRhom2 are greater than 85%. This result provided enough evidence to study the proposed hypothesis about the stability of cub-mutated iRhom2.

4.2 3D Protein Modelling

One of the essential objectives of this research project was 3D protein modelling. After the global alignment, it was evident that mice iRhom2 and human iRhom2 are highly similar and homologous proteins [34]. It is known by literature that the structures of iRhoms and Derlins are not solved by NMR (Nuclear magnetic resonance) or X-ray crystallography to date [9], [32]. 3D structures of transmembrane proteins are notoriously difficult to solve [89]; therefore, the research can be continued by solving the 3D structure of wild-type and cub mutated iRhom2 in humans.

Homology modelling is one of the most accurate computational protein modelling techniques [88], [90]. However, the cub-domain (269-351 a.a) and cytoplasmic N-terminus or wild-type domain (1-351 a.a) of iRhom2 could not find a single homologue after pBlast (PDB search). These domains are unique to iRhoms [32], [34]; therefore, no homologues were found in PDB.

Ab-initio and fold-recognition based computational protein modelling techniques were suitable for this project because both methods are considered efficient for 3D modelling of proteins with similarity in PDB less than 20% [54].

4.2.1 Fold Recognition Based Modelling (I-TASSER)

As mentioned in the second chapter I-TASSER was used to predict 3D models of cub-mutated and wild-type iRhom2. First, the complete cytoplasmic N-terminus (1-351) of iRhom2 was submitted for modelling in I-TASSER.

4.2.1.1 Cytoplasmic N-Terminus Domain (1-351) The I-TASSER server generated five different models, their respective quality estimations, i.e. c-score, RMSD and TM-score, are mentioned in Figure 4.2.

Name	C-score	Exp.TM-Score	Exp.RMSD	No.of decoys	Cluster density
Model1:	-1.37	0.55+-0.15	9.7+-4.6	600	0.2500
Model2:	-1.58			600	0.2020
Model3:	-2.80			600	0.0597
Model4:	-3.89			126	0.0202
Model5:	-4.00			114	0.0181

Figure 4.2 C-score and other evaluations of models generated by I-TASSER

Model1 was selected out of the four models generated by I-TASSER. C-score shows a -1.37 confidence level of model1, which is the highest of other models. The favourable confidence score is between -5,2. TM-score is co-related with c-score for high-rank models [57]. C-score and TM-score indicate a mid-rank structural quality. The cluster density of the model1 is highest than other models too. Figure 4.3 shows the 3D structure of model1. The uniqueness of iRhoms majorly depends on the long cytoplasmic N-terminus; the hypothesized structure of iRhoms suggests the same long loop at the edge like model1 predicted by I-TASSER [32].

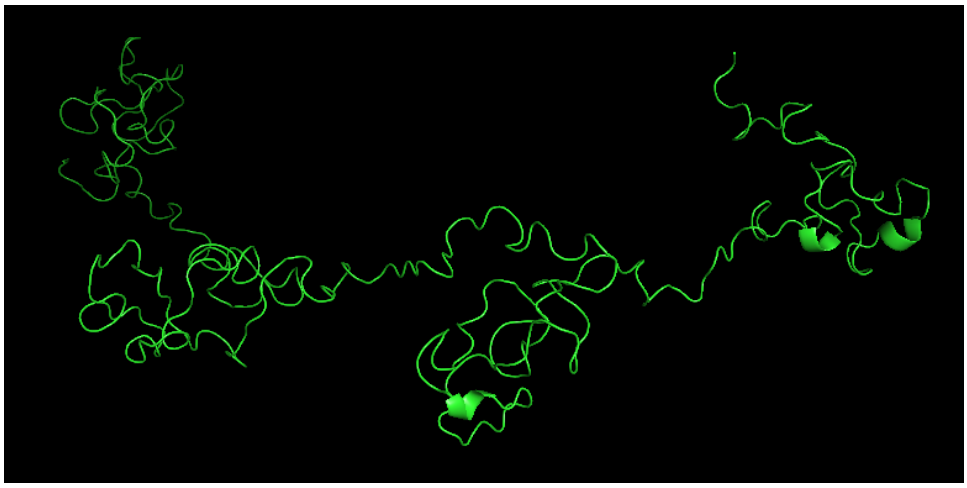

Figure 4.3 Predicted model of complete cytoplasmic N-Terminus. Most part is consists of loops, some helices and sheets

Every model generated by I-TASSER was evaluated individually before finally selecting a model for MD simulations.

4.2.1.2 Cub-Mutated Domain (269-351)

I-TASSER predicted the cub-domain of iRhom2, which lacks most of the n-terminus (only 85 a.a). Five models were generated, their respective quality estimations, i.e., c-score, RMSD and TM-score, are mentioned in Figure 4.4.

Name	C-score	Exp.TM-Score	Exp.RMSD	No.of decoys	Cluster density
Model1:	-3.20	0.36+-0.12	10.6+-4.6	2333	0.0598
Model2:	-4.31			742	0.0197
Model3:	-2.68			630	0.1012
Model4:	-3.13			620	0.0646
Model5:	-4.55			575	0.0156

Figure 4.4 C-score and other evaluations of models generated by I-TASSER

The c-score for the cub-domain is between -3.20 to -4.55. Model1 was selected for further studies because of overall structural accuracy. Model4 shows a c-score of -3.13, which is less than Model1, TM-Score, RMSD, No. of decoys represent better structural quality than Model1. Figure 4.5 shows the predicted Model1 by I-TASSER.

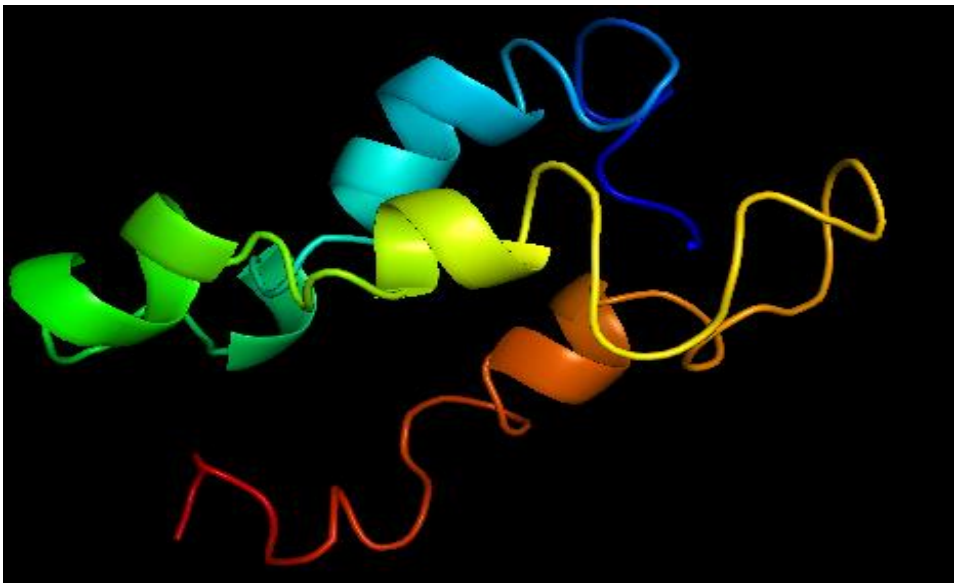

Figure 4.5 Predicted model of cub-domain of Human iRhom2. Helices can be spotted connected by loops.

The structure of the complete N-terminus shows fewer helices than cub-domain; therefore, it can be concluded that deletion of 1-268 a.a shows a drastic change in the structure of the remaining N-terminus of iRhom2.

4.2.2 Ab-initio Modelling (Robetta)

The second method opted for iRhom2 domain structure prediction is ab-initio protein modelling. As mentioned earlier in chapter 2, ab-initio modelling was required because no homologues were found for iRhom2 cub-domain or complete N-terminus in the PDB.

4.2.2.1 Cytoplasmic N-Terminus Domain (1-351)

Robetta also generates five models for each submitted query sequence. Models are ranked in ascending order w.r.t to quality of the structure. C-score (confidence level) is mentioned collectively for all five models. C-score for Robetta ranges between 0 to 1; closer to one show high-quality models, and closer to zero represents poor quality models. The confidence level for ab-initio models is the average pairwise TM-score of the top ten Rosetta scoring models. For the cytoplasmic N-terminus domain c-score is 0.11.

Figure 4.6 shows model1 predicted by Robetta. This structure physically differs a lot from the structure predicted by I-TASSER. The significant difference in structure can be observed due to different modelling algorithms.

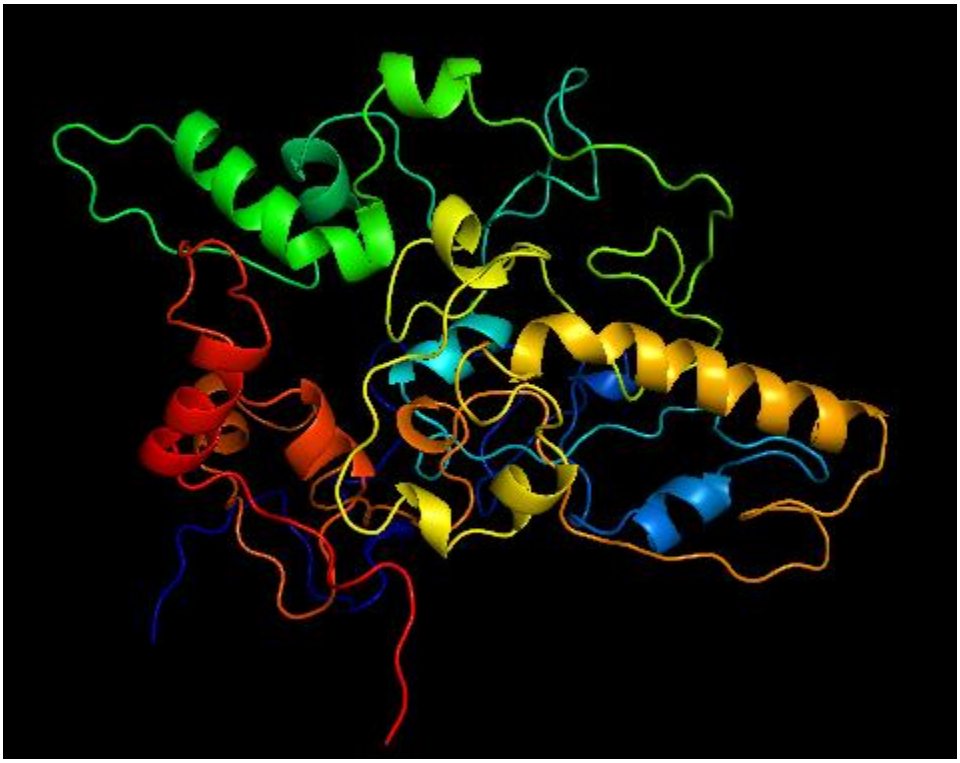

Figure 4.6 iRhom2 complete cytoplasmic N-terminus domain. Loops can be observed. Many coils and sheets can be seen prominently in this structure.

4.2.2.2 Cub-Mutated Domain (268-351)

As mentioned earlier, the confidence level (c-score) is the single qualitative measure for ab-initio models in the Robetta server. C-score for the predicted cub-domain is 0.22. As mentioned previously, the c-score for the complete N-terminus domain was 0.11, which shows poor structure quality. Cub-domain predicted model shows better structure quality; however, the overall quality of structures predicted by Robetta is relatively low compared to I-TASSER. Structure evaluation will reveal more about model quality in detail. Figure 4.7 shows model1 predicted by Robetta.

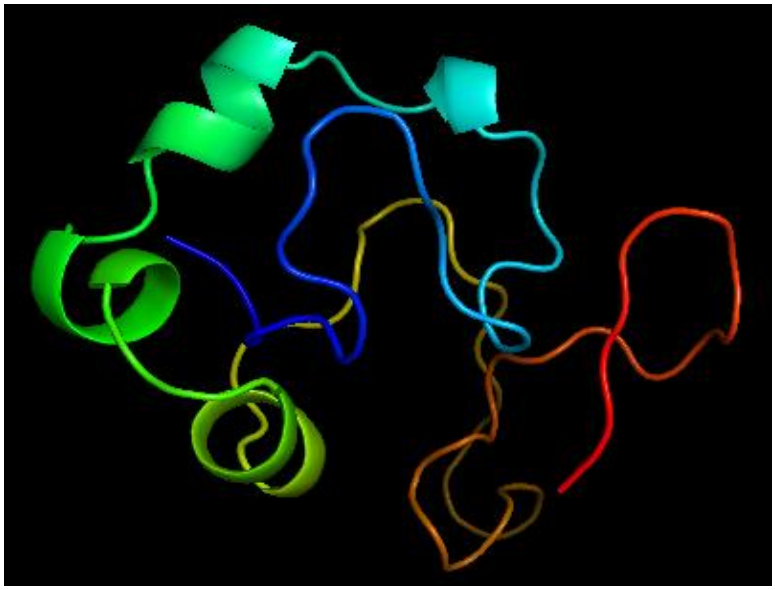

Figure 4.7 iRhom2 cub-domain structure predicted by Robetta. Different loops and helices can be observed.

4.3 Structure Evaluation by SAVES

Different models were predicted by I-TASSER and Robetta; as mentioned earlier, out of several predicted models, only single models were selected for each domain. The SAVES server evaluated every predicted model; therefore, the models that outperformed others during evaluations were analysed by MD simulation.

4.3.1 I-TASSER Domains Structure Evaluation

Every model generated by I-TASSER was submitted to the SAVES server; usually, model1 showed better statistics than other models after evaluations; therefore, we discussed model1 evaluations from each domain in this section.

First of all, model1 of the cytoplasmic N-terminus domain was evaluated. Figure 4.8 shows the results of evaluation. Verify3D shows 23% of residues have averaged 3D-1D score >= 2.0, which shows “fail” 3D structure modelling. For an extended protein sequence of 351 a.a, this score shows a misfolded structure. As mentioned in the literature, verify 3D compares the predicted 3D structure with its 1D or sequence positions and environment (physicochemical properties). [91]. PROVE results shows 10.3% outliers from model1,

which predicts a precarious structure. ERRAT shows an overall structure quality of approximately 76.5%. ERRAT score of complete N-terminus domain is better than other evaluations by different servers. ERRAT works by analysing the statistics of non-bonded interactions between different atom types. [60]. Therefore, it can be considered that the iRhom2 complete N-terminus domain predicted by I-TASSER can be used for further studies.

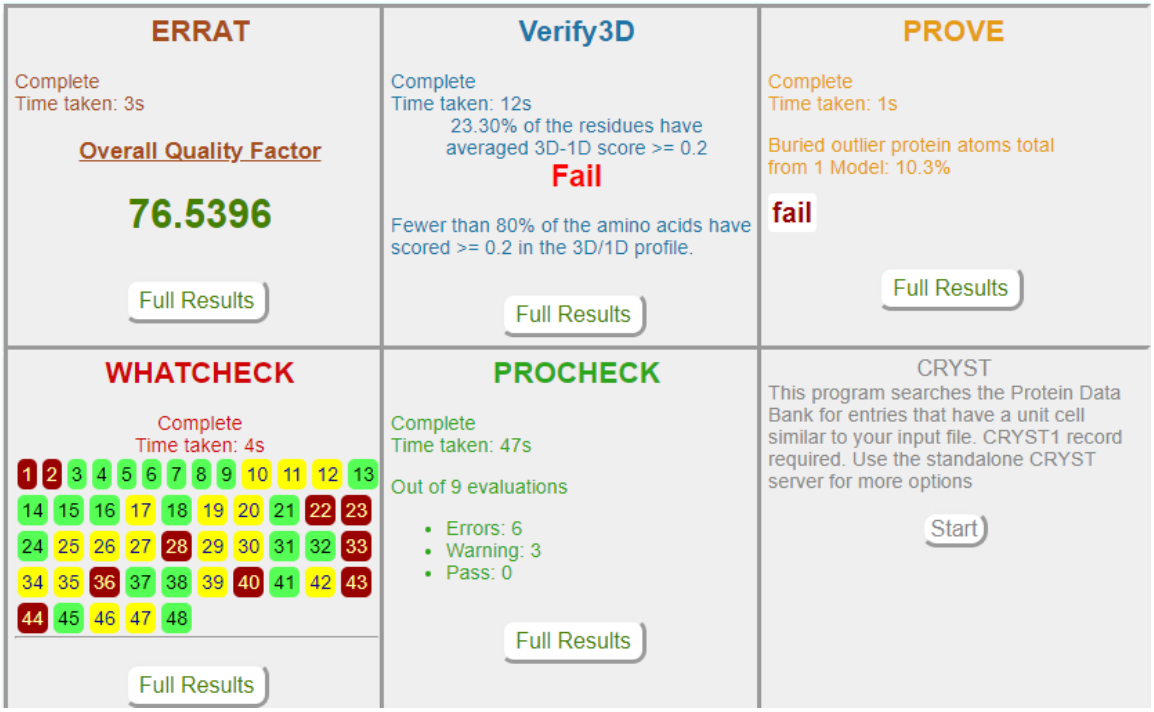

Figure 4.8 Shows SAVES structure evaluation results for iRhom2 complete cytoplasmic N-terminus domain

4.3.1.1 Ramachandran plot for complete cytoplasmic N-terminus domain

PROCHECK server shows six errors, three warnings and 0 pass evaluations out of 9 evaluations. Ramachandran plot also shows 38.6% core residues, 48.3% residues in the allowed region, 7.9% in the generously allowed region and 5.2% disallowed region (Figure 4.9). These results by PROCHECK [63] explains that this domain structure has many unacceptable steric clashes and subpar geometrical assembly.

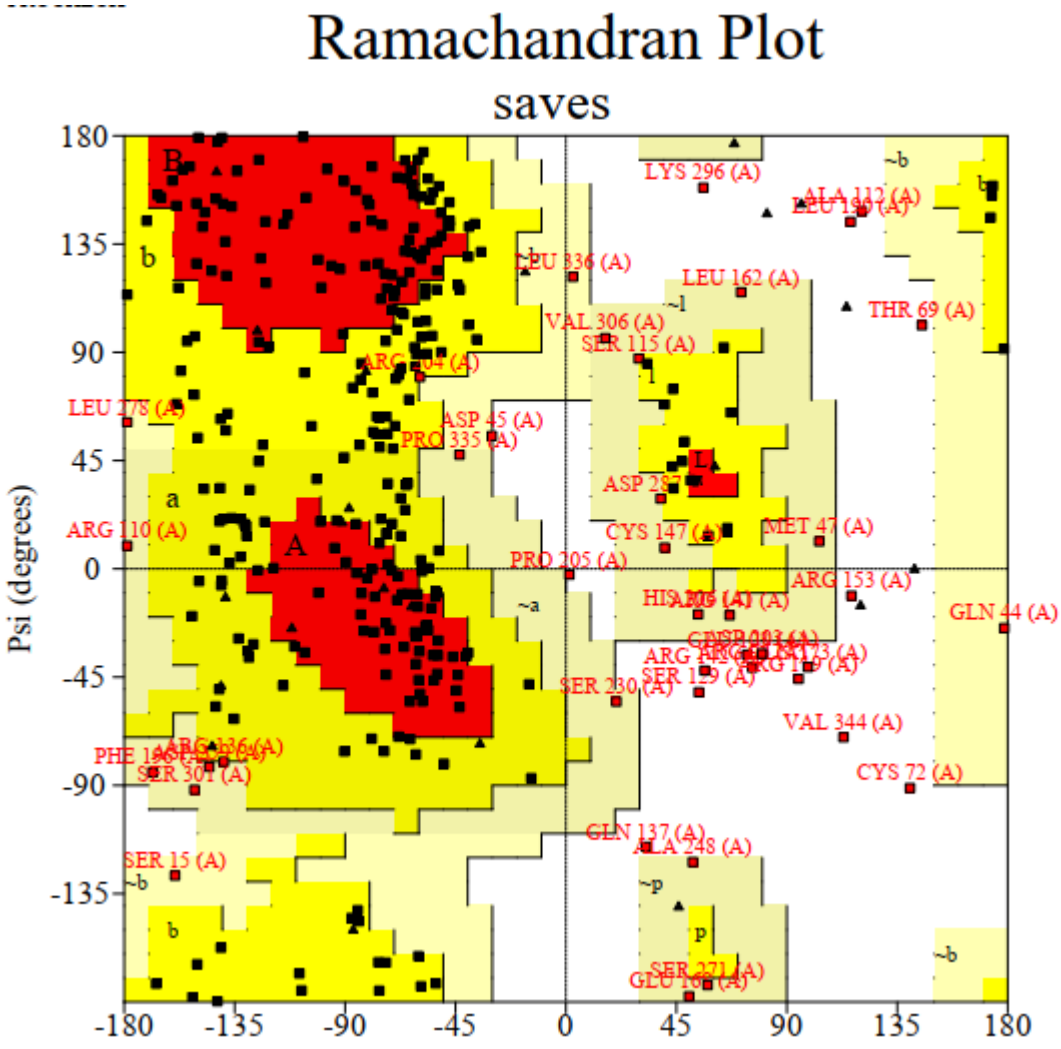

Figure 4.9 Ramachandran plot of iRhom2 complete cytoplasmic N-terminus domain

The second domain predicted by I-TASSER was cub-mutated iRhom2 (cub part removed from N-terminus only). An evaluation was done on this domain by SAVES for further analysis; results are mentioned in Figure 4.10. Verify3D shows 56 % of the residues have averaged 3D-1D score >= 0.2. This estimation indicates a “fail” model because more than 80% of a.a residues should score the average residue score >=0.2. PROVE shows 5.05 buried outliers in the model. ERRAT shows an overall structure quality of approximately 77%. ERRAT score of iRhom2 cub-domain is better than other evaluations by different servers.

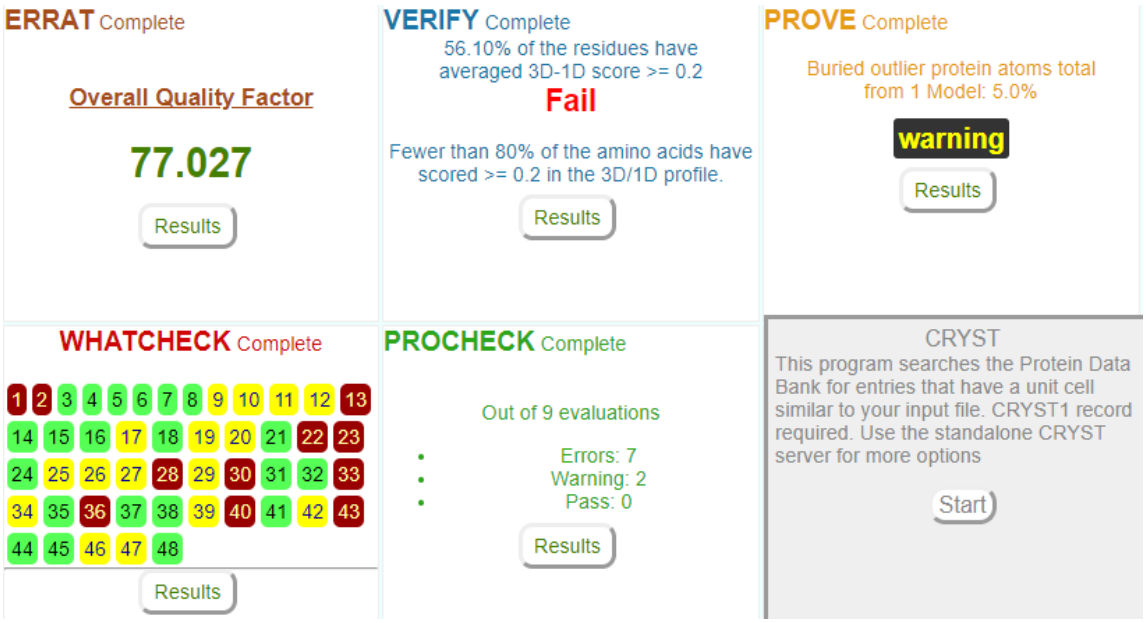

Figure 4.10 SAVES structure evaluation results for iRhom2 Cub-domain predicted by I-TASSER

4.3.1.2 Cub-domain Ramachandran Plot

PROCHECK server shows seven errors, two warnings and 0 pass evaluations out of 9 evaluations. Ramachandran plot also shows 46% core residues, 31.7% residues in the allowed region, 12% in the generously allowed region and 9.5% disallowed region (Figure 4.11). These results by PROCHECK [63] explains that this domain structure has many unacceptable steric clashes and subpar geometrical assembly.

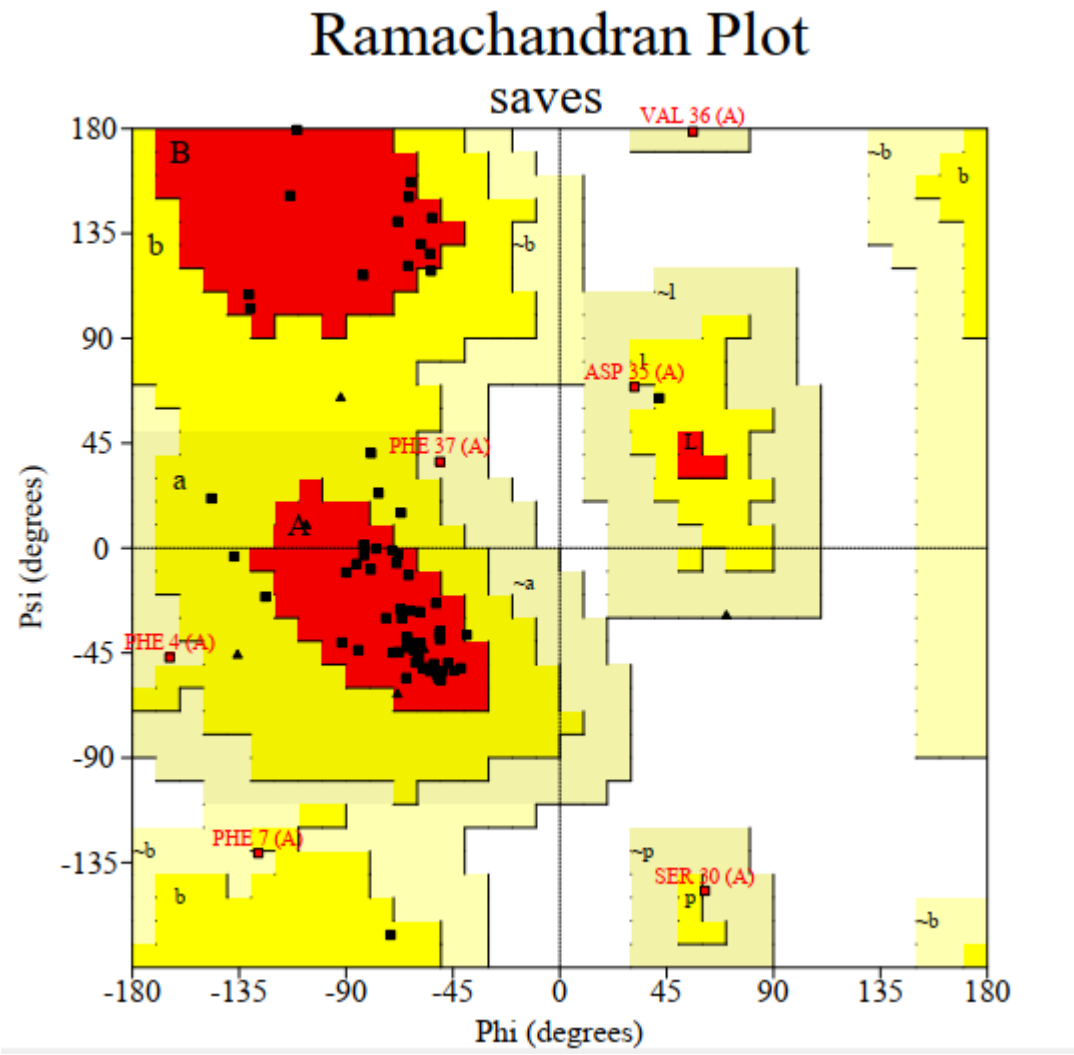

Figure 4.11 Ramachandran plot of iRhom2 cub-domain

4.3.2 Robetta Domains Structure Evaluation

Ab-initio based modelling by Robetta results in a unique set of models for both IRhom2 domains. First of all, model1 of the cytoplasmic N-terminus domain was evaluated. Results are mentioned in Figure 4.12. Verify3D shows approximately 78% of residues have averaged 3D-1D score >= 2.0, which shows “fail” 3D structure modelling. For an extended protein sequence of 351 a.a, this score shows a misfolded structure; however, this score by verify3D for the N-terminus domain is relatively better than the domain predicted by I-TASSER. ERRAT evaluation results show an overall quality factor of 88.9%. ERRAT score is somewhat improved for the domains predicted by Robetta.

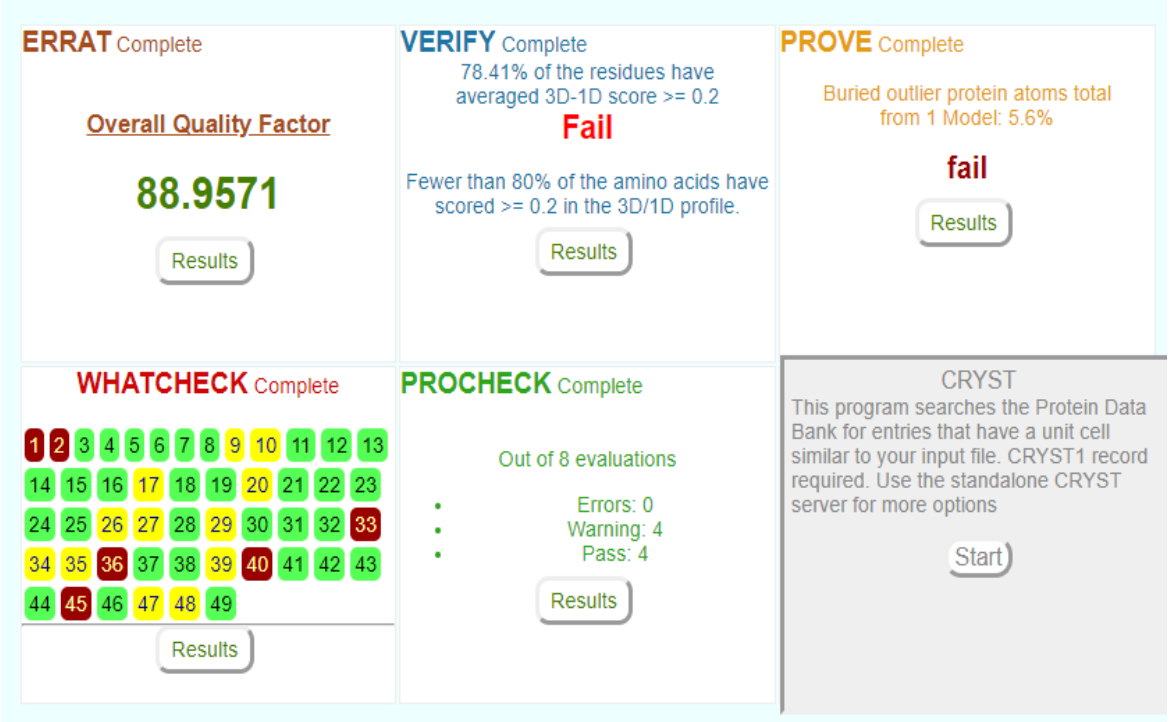

Figure 4.12 SAVES structure evaluation results for iRhom2 Complete cytoplasmic N-terminus predicted by Robetta

4.3.2.1 Ramachandran plot of iRhom2 complete cytoplasmic N-terminus

PROCHECK server shows zero errors, four warnings and four pass evaluations out of 8 evaluations. Ramachandran plot also shows 86.9% core residues, 13.1% residues in the allowed region, 0% in the generously allowed region and 0% in the disallowed region (Figure 4.13). These results by PROCHECK [63] explains that this domain structure has lesser steric clashes and a desirable geometrical assembly.

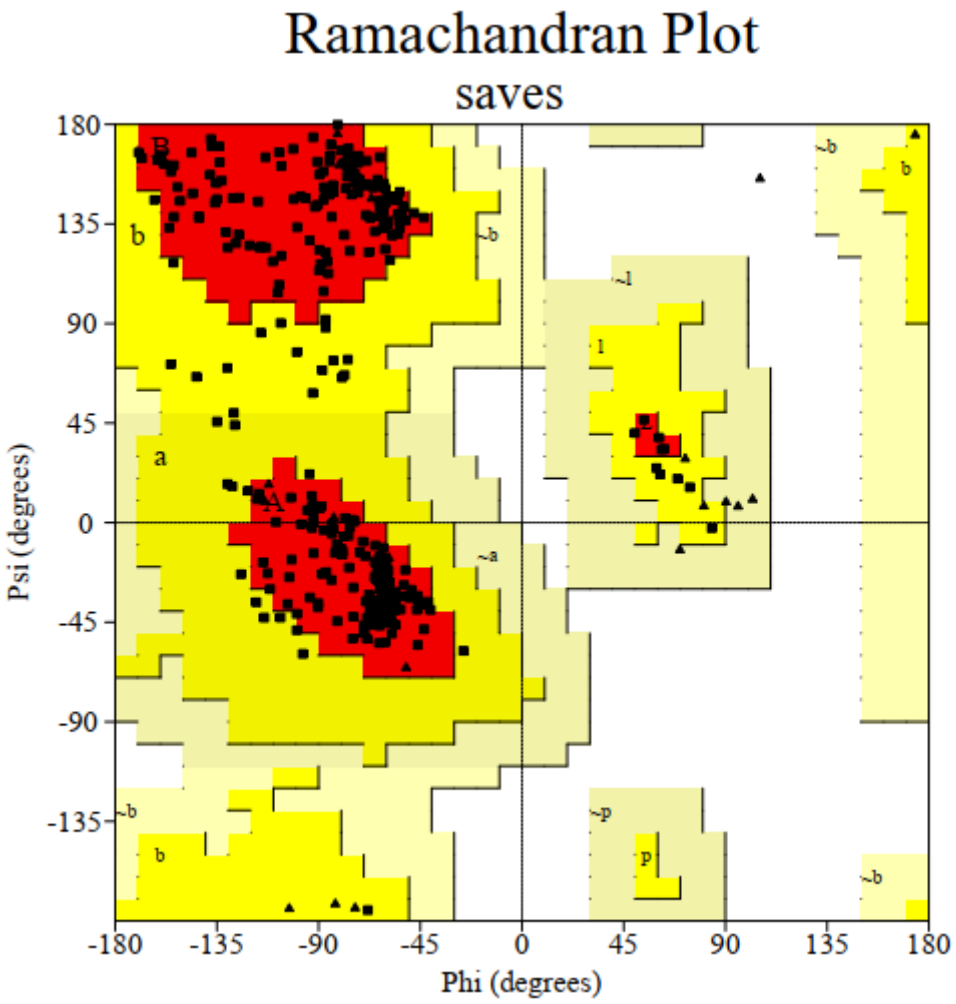

Figure 4.13 Ramachandran plot of iRhom2 complete N-terminus domain

The second iRhom2 domain predicted by Robetta is the cub-domain (269-351 a.a). The evaluation results are mentioned in Figure 4.14. Verify3D shows approximately 84.15% of residues have averaged 3D-1D score >= 2.0, at least more than 80% of the amino acids have scored >= 2.0, which shows "Pass" 3D structure modelling. For a short length structure like cub-domain with only 82 residues, evaluation of this model shows desirable results. ERRAT shows an overall structure quality of approximately 98.6%. ERRAT score of iRhom2 cub-domain predicted by Robetta is better than domains predicted by the I-TASSER server. Robetta is known to produce quality 3D models for smaller length sequences. ERRAT works by analysing the statistics of non-bonded interactions between different atom types; a higher ERRAT score leads to the non-random distribution of

different atoms [60]. WHATCHECK shows more minor errors than other domains evaluations. However, PROVE shows 5.6% buried outlier protein atoms.

After accessing all evaluations of domain models by Robetta and I-TASSER, it can be assumed that for iRhom2, ab-initio based modelling showed better structure quality than threading based modelling techniques.

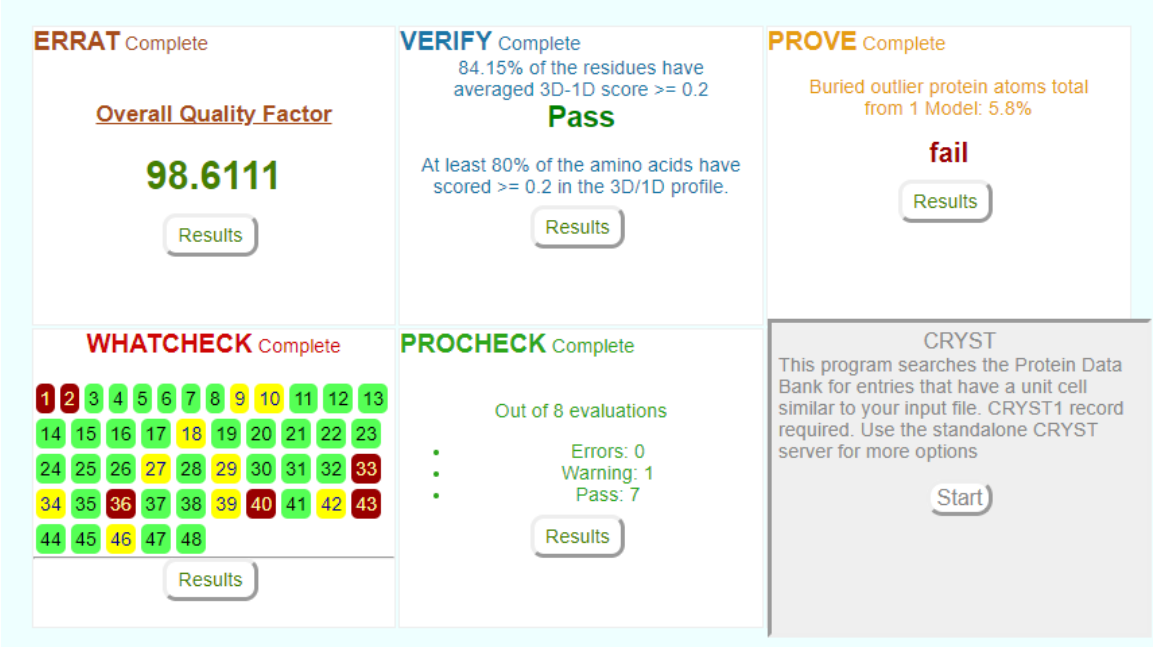

Figure 4.14 SAVES structure evaluation results for iRhom2 cub-domain predicted by Robetta

4.3.2.2 Ramachandran plot of cub-domain

PROCHECK server shows zero errors, only one warning and seven pass evaluations out of eight evaluations. Ramachandran plot also shows 93.7% core residues, 6.3% residues in the allowed region, 0% in the generously allowed region and 0% in the disallowed region (Figure 4.15). These results by PROCHECK [63] explains that this domain structure has almost no steric clashes and a desirable geometrical assembly.

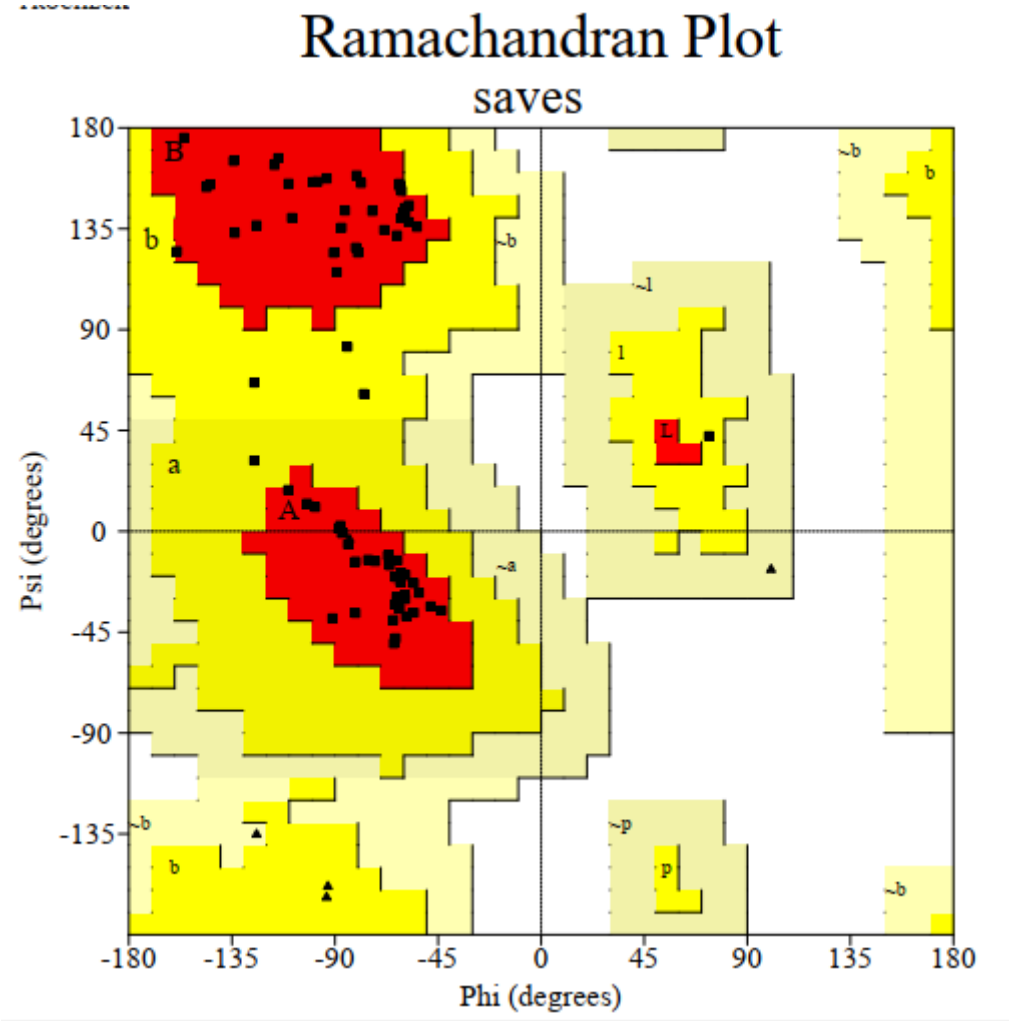

Figure 4.15 Ramachandran plot of iRhom2 cub-domain

4.4 Structure Comparison

In order to observe the side-by-side difference between domains predicted by Robetta and I-TASSER, structure comparison (superimposition) was crucial for this project. A superimposition method was used to select domains predicted by servers with better structure quality for MD simulations. Chimera was used to complete this objective.

4.4.1 Complete N-terminus domain (1-351) Structure Comparison

I-TASSER and Robetta both predicted structures for iRhom2 complete cytoplasmic N-terminus. Figure 4.16 shows the superimposition between models selected after SAVES

evaluations. The matchmaker tool was used to align the models. Structure predicted by I-tasser was given as a reference structure, while structure predicted by Robetta was selected as "structure to match." It can be hypothesized that different 3D protein modelling techniques would fold structures differently. However, it was observed after superimposition that there is a drastic change between both models. I-TASSER model is extended as a complete loop with few sheets and helices. Structures with mostly loops are considered unstable, misfolded or at a high energy state. Literature also mentions a long cytoplasmic N-terminus with many loops [32]; however, the I-TASSER model shows excessive loops compared to the model generated by Robetta.

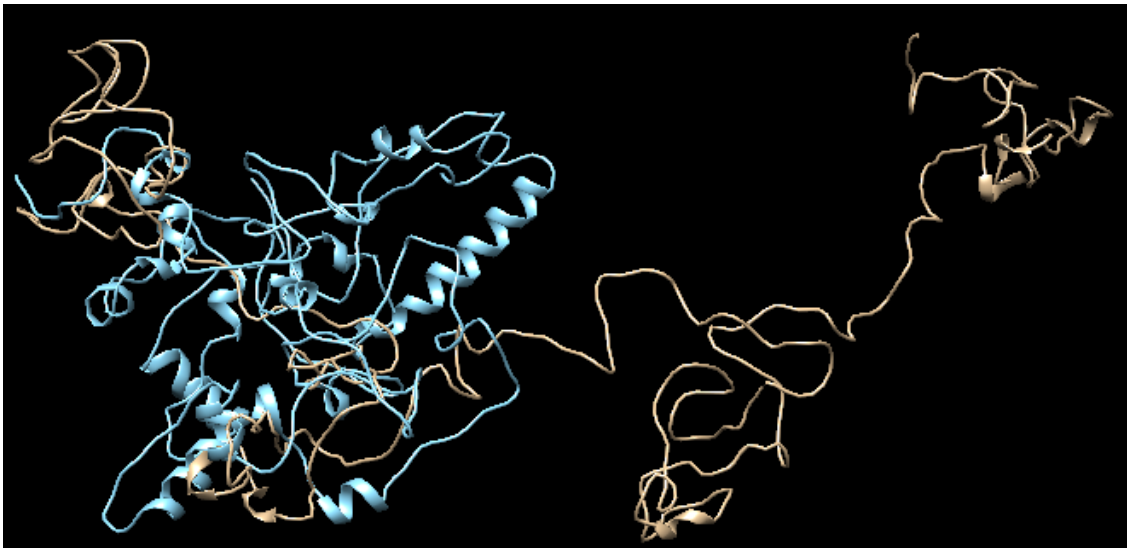

Figure 4.16 Superimposed complete N-terminus domain models by chimera. The golden model was predicted by I-TASSER, and Robetta predicted the blue model.

4.4.2 Cub-Domain (269-351) Structure Comparison

The cub-domain structure consists of only 82 amino acids. I-TASSER and Robetta both predicted structures for the iRhom2 cub domain. Figure 4.17 shows the superimposition between models selected after SAVES evaluations. As mentioned earlier, the chimera tool matchmaker was used to align both models. Structure predicted by I-TASSER was given as a reference structure, while the structure predicted by Robetta was selected as "structure to match". The superimposition shows a poor alignment between both models. I-TASSER predicted a more compact model as compared to the structure predicted by Robetta.

However, SAVES evaluation results show that Robetta models are of high quality as compared to I-TASSER.

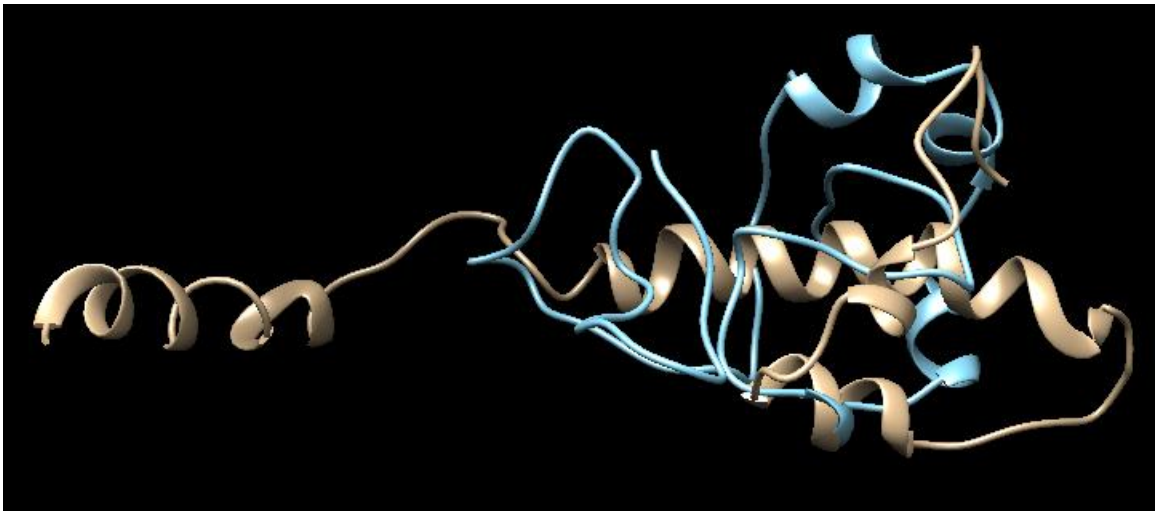

Figure 4.17 Superimposed complete cub-domain models by chimera. The golden model was predicted by I-TASSER, and Robetta predicted the blue model.

4.5 MD Simulation Analysis

MD simulations were carried out using protein structures generated by Robetta and thoroughly evaluated by the SAVES server. In addition, the dynamic behaviours of iRhom2 domains were examined to clarify the dynamic stability of these structures. The stability profiling depends on the collective analysis of these structures by different parameters using GROMACS. Therefore, both domains were compared side by side based on different parameters.

4.5.1 Energy Minimisation

After topology generation, the definition of ions, and the solvation of iRhom2 domains, there were the least possible steric clashes; we had a solvated electroneutral system ready for the energy minimization step. Energy minimization was done by GROMACS MD engine, “mdrun.” After energy minimization of both iRhom2 domains, potential energy graphs were generated for both, Figure 4.18 and Figure 4.19 explains potential energy against time in picoseconds (ps). The complete cytoplasmic N-terminus domain shows a final potential energy value of -1.9384938e+06. The steepest energy value was

achieved at the 959 steps. The complete cytoplasmic N-terminus domain was a more extensive system, and it is known by literature that an extensive system takes more time to reach the lowest energy conformation [80], [92]. For cub-domain (268-351) final potential energy value was -3.8536091e+05. The steepest energy value was achieved at the 370 steps. The graph shows few energy distortions at the start and the steepest potential energy values in kJ/mol plotted against time (ps). The cub-domain was a less extensive system; hence, it took fewer steps to reach the lowest energy conformation.

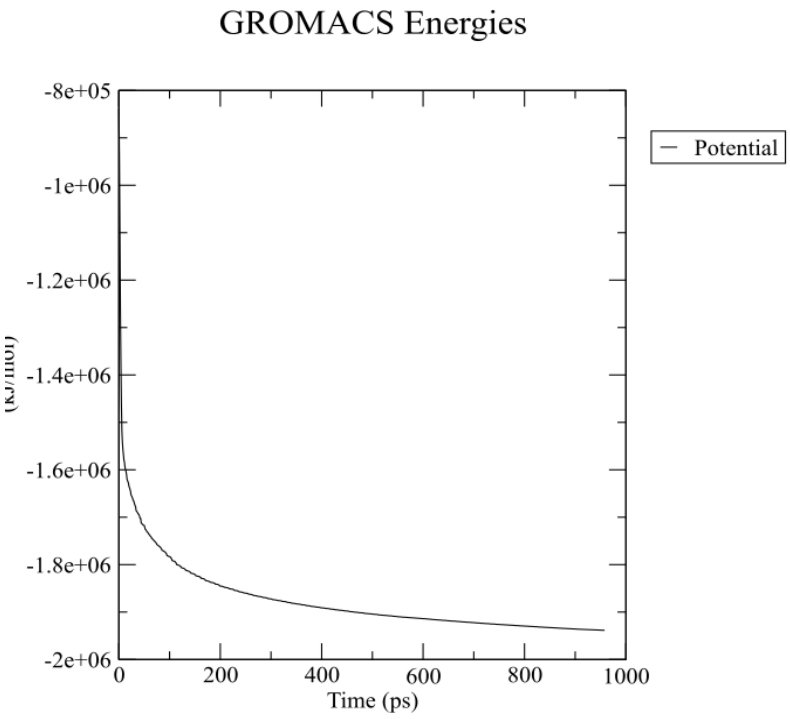

Figure 4.18 Potential energy plot for complete cytoplasmic N-terminus domain (1-351).

A massive change in potential energy values w.r.t can be observed. The low potential energy value of the complete cytoplasmic N-terminus domain shows a more unstable structure than the cub-domain, which shows a higher potential energy value. Therefore, after energy minimization, it can be observed that the cub-domain of iRhom2 is more stable than the complete cytoplasmic N-terminus.

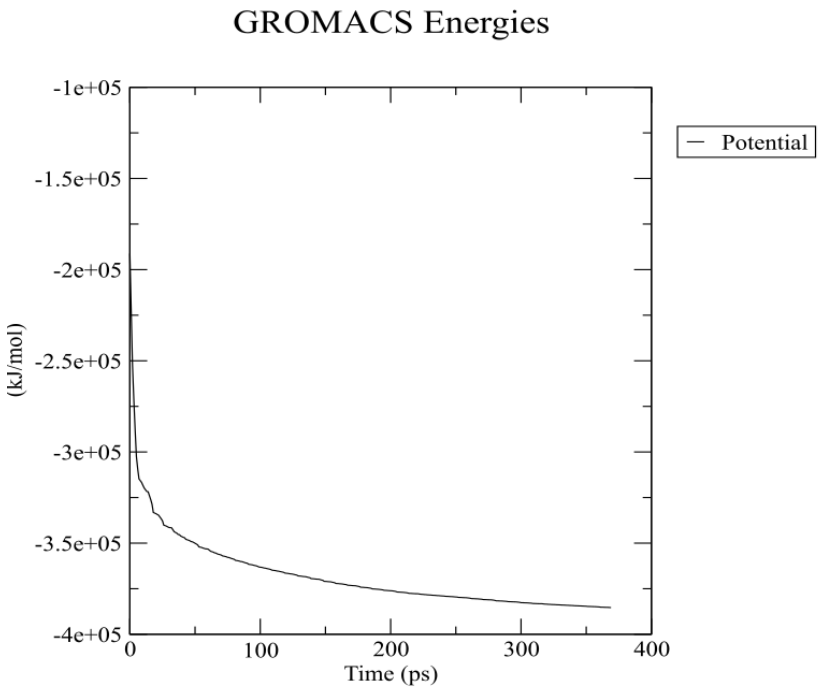

Figure 4.19 Potential energy plot for iRhom2 cub-domain (268-351).

However, the stability of iRhom2 domains cannot be determined solely by final potential energy values because no entropic term was involved while calculating potential energy. Moreover, potential energy depends on the respective force fields; with the change of force fields, potential energy would be different, unlike the kinetic energy of the system.

4.5.2 NVT Equilibration

In order to solvate ions around proteins, equilibration was done. Generally, the solvent around protein is optimized within itself and not around the protein. Therefore, to establish proper orientation around protein, it was crucial to bring our system to the optimal temperature, i.e. 300 K [75]. Equilibration is done under NVT ensemble, with a constant number of particles, temperature and volume. NVT equilibration was conducted at 100-ps; this time constraint was considered sufficient for both iRhom2 domains to reach the desired temperature.

The NVT equilibration graphs of the iRhom2 complete cytoplasmic N-terminus domain and cub-domain are mentioned below (Figure 4.20 & Figure 4.21). The graph for the iRhom2 complete cytoplasmic N-terminus domain (Figure 4.20) shows fluctuations in temperature across the time frame of 100 ps. For this scenario, the temperature of the system quickly reaches the target. On the other hand, for this domain of iRhom2, an abrupt shift of temperature can be observed around 30-80 ps. Therefore, the system can be assumed as slightly unstable.

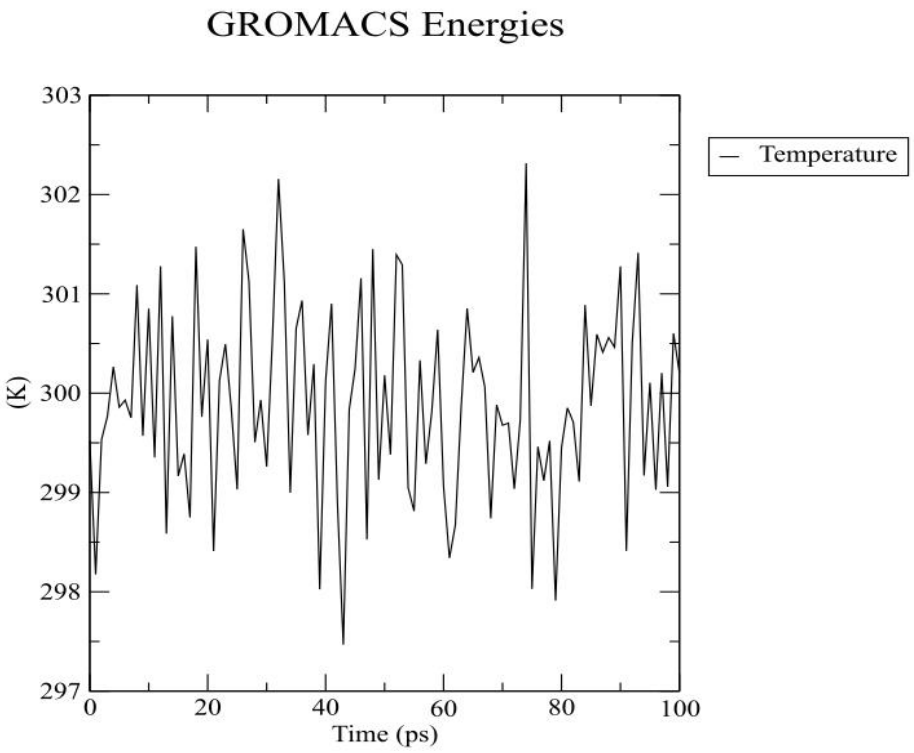

Figure 4.20 NVT equilibration plot of iRhom2 complete cytoplasmic N-terminus domain

The iRhom2 cub-domain, the graph (Figure 4.21) shows that the system reaches the desired temperature quickly. The cub-domain is relatively a less extensive system than the complete cytoplasmic N-terminus domain. Therefore, a shorter period of equilibration around 50 ps might be sufficient. The consistent behaviour of temperature with respective time frames shows a stable behaviour of the iRhom2 cub-domain.

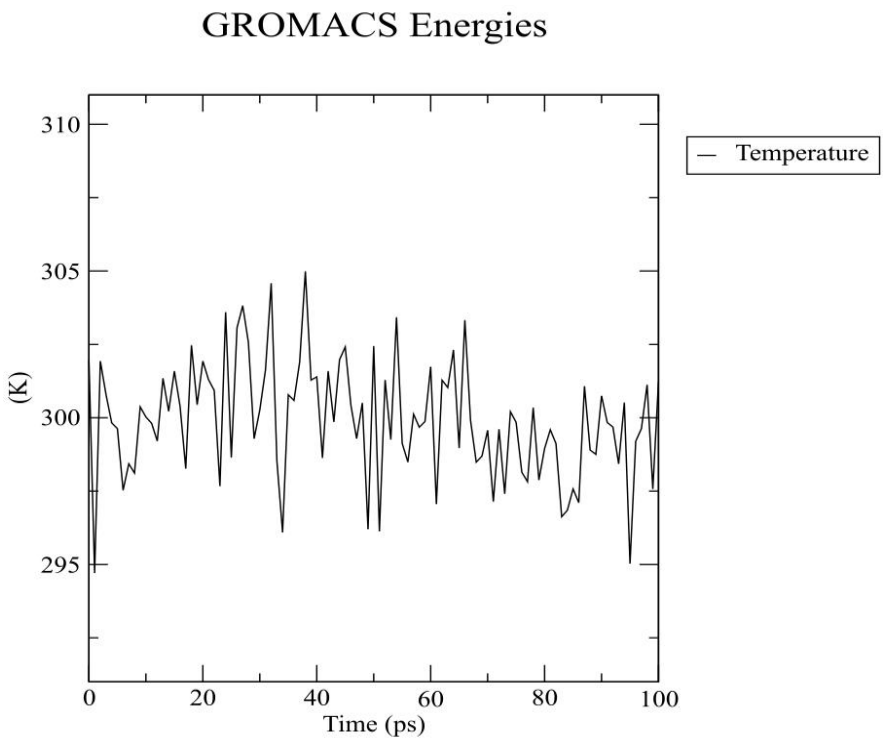

Figure 4.21 NVT equilibration plot of iRhom2 cub-domain

4.5.3 NPT Equilibration (Pressure)

The second part of equilibration is NPT equilibration. The temperature was stabilised; the next step was pressure equilibration. The number of particles, temperature and pressure, are all constant. The NPT equilibration was run for 100 ps, grompp was called with mdrun. The graphs generated after running NPT for iRhom2 domains are mentioned below (Figure 4.22 & Figure 4.23); the cytoplasmic N-terminus domain equilibration shows fluctuations among pressure values for a total of 100-ps. However, these fluctuations were expected; the pressure bar was set to one initially, and then the bar reached up to ± 300. The graph generated after the cub-domain shows more fluctuating pressure than the complete cytoplasmic N-terminus domain. The pressure bar was set to one initially for cub-domain, and the scale reached up to ± 600. The fluctuations in pressure are observed widely throughout MD simulations [75]. The considerable change in pressure fluctuations among both domains shows that the cub-domain system required more pressure than the complete cytoplasmic N-terminus domain to be equilibrated.

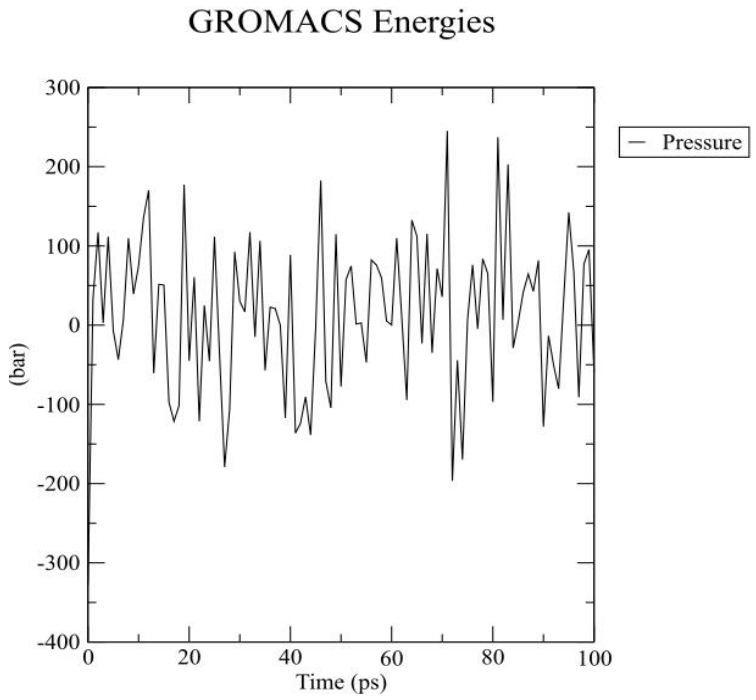

Figure 4.22 NPT (pressure) equilibration plot of iRhom2 complete cytoplasmic N-terminus domain

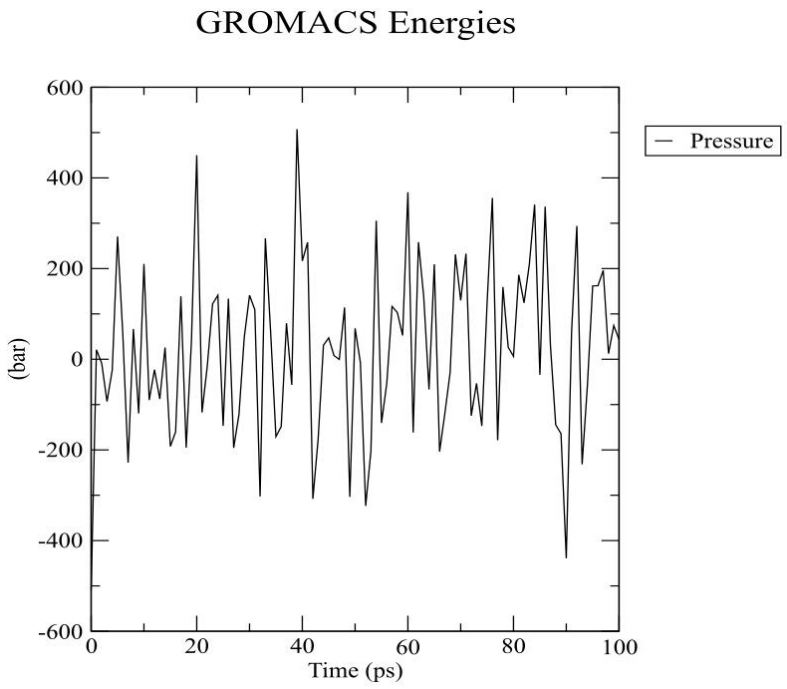

Figure 4.23 NPT (pressure) equilibration plot of iRhom2 cub-domain

4.5.4 NPT Equilibration (Density)

After NPT pressure equilibration, density equilibration was required to complete the equilibration process to proceed towards the production MD step. The number of particles, temperature and pressure, are all constant. The NPT equilibration was run for 100 ps once for both. The graphs generated after running NPT for iRhom2 domains are mentioned below (Figure 4.24 & Figure 4.25);

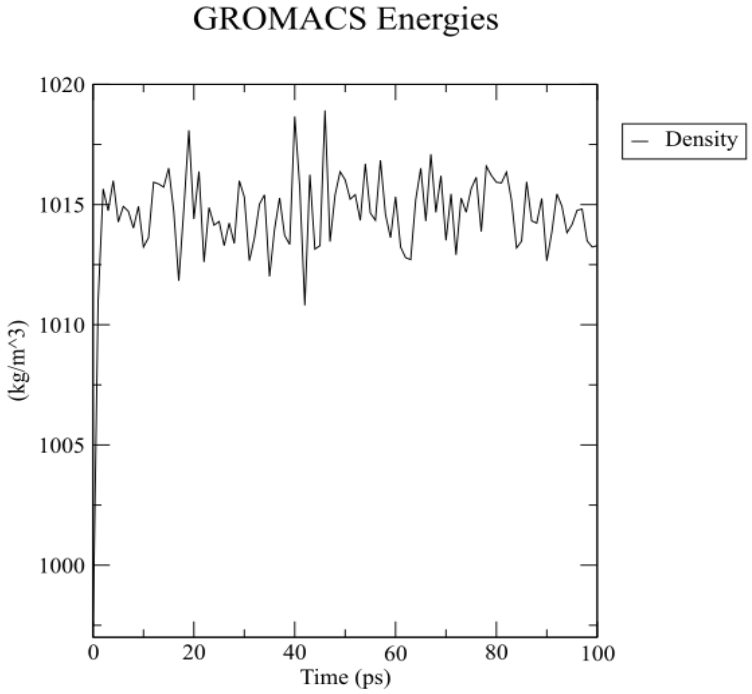

Figure 4.24 NPT (density) equilibration plot of iRhom2 complete cytoplasmic N-terminus domain

The complete cytoplasmic N-terminus domain shows some fluctuations across the entire equilibration period. The average kg/m^3 for 100 ps is approximately 1000-1056. For cub-domain, the graph shows density values stability across the 100 ps time frame. The average kg/m^3 for 100 ps is approximately 1000-1015. The equilibration results were expected due to structural differences; the iRhom2 cytoplasmic N-terminus domain consists of more loops than the cub-domain.

Moreover, the sequence length of the cub-domain only consists of 83 amino acids. Therefore, the cub-domain was equilibrated at least amount of density. The domains were

equilibrated with respect to pressure and density at that step; the next step was production MD. The MD was for one ns, and this was a relatively short MD simulation; however, it was considered appropriate for this particular task. The goal was to observe and conclude the stability profile of iRhom2 domains.

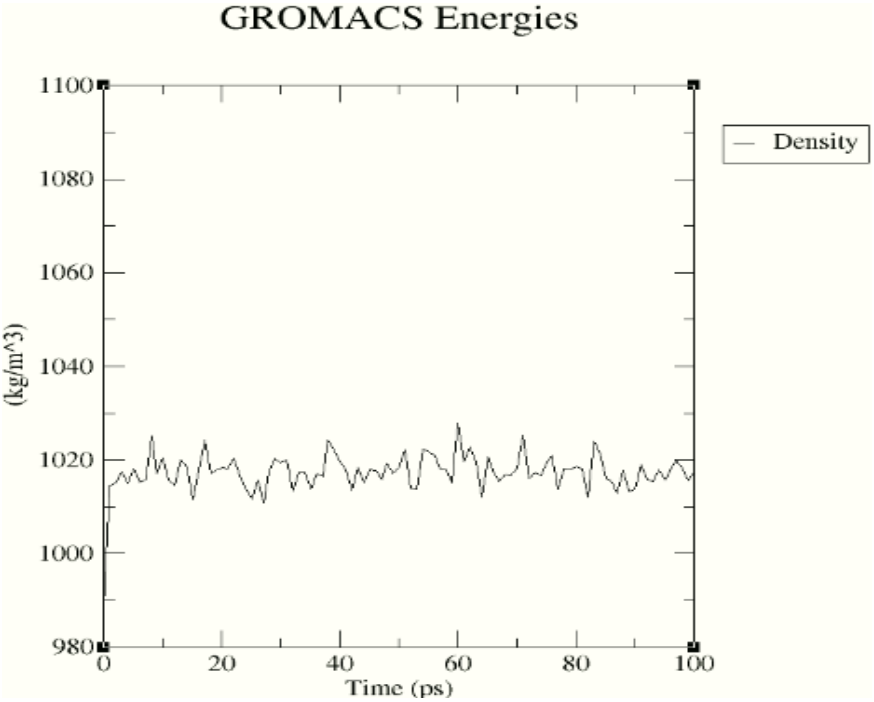

Figure 4.25 NPT (density) equilibration plot of iRhom2 cub-domain

4.5.5 Radius of Gyration

The calculation of the radius of gyration is another way to analyse MD simulation results for protein stability. The radius of gyration calculates the distance between the centre of the mass protein and both termini of the protein. The compactness and folding of the protein tell a lot about the structural stability of a protein (Lobanov, Bogatyreva, and Galzitskaya, 2008). The radius of gyration for iRhom2 domains is explained in 'nm' against time (Figure 4.26 and Figure 4.27). The radius of the gyration plot for complete cytoplasmic N-terminus (Figure 4.26) shows that at 0 ps, the value of 'Rg' kickstarted at 2.26, which is a high value, and over time, the slop of the plot kept on fluctuating. Therefore, we can say that variant behaviour of iRhom2 complete cytoplasmic N-terminus remained unstable and did not stay as a compact structure throughout 1000 ps at 300k temperature.

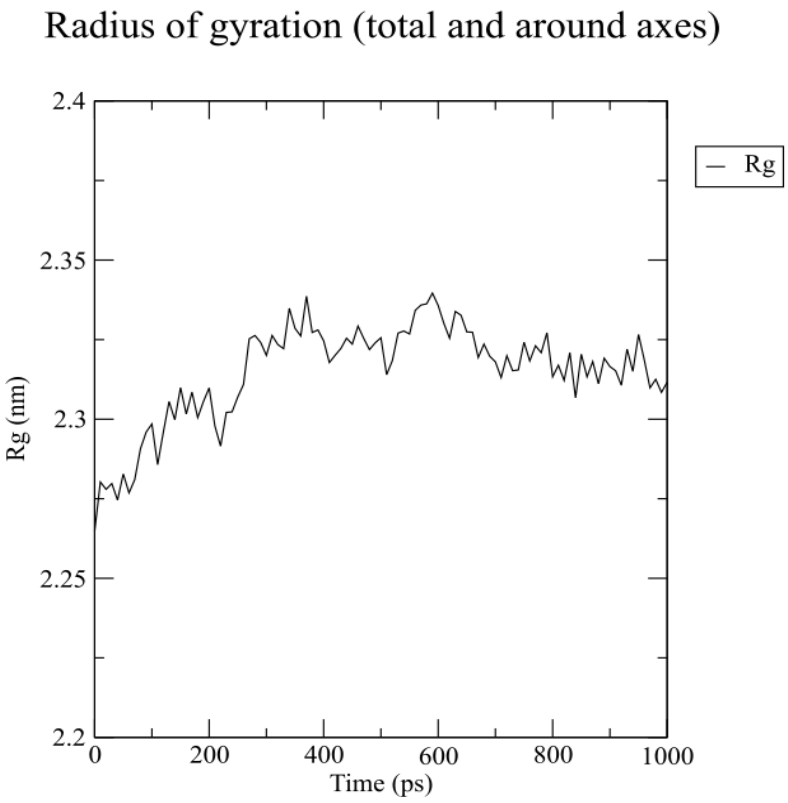

Figure 4.26 Radius of gyration plot of iRhom2 complete cytoplasmic N-terminus domain

The radius of the gyration plot for the iRhom2 cub-domain shows that at '0' ps, the initial value of Rg was 1.3 nm, but soon after, it can be observed variating around 1.2 ± 1.35 nm with the course of 1000 ps. Therefore, the behaviour of the cub-domain is relatively stable as compared to the complete cytoplasmic N-terminus domain. This result was not expected because the less compactable behaviour of the complete cytoplasmic N-terminus domain depends highly on the secondary structure of this domain, which mainly consists of a big loop and α-helices. The loop at the N-terminus and α protein behaviour of iRhom2 contributes to a higher radius of gyration and low compactness of complete cytoplasmic N-terminus domain [85]; therefore, the radius of gyration is higher for this domain is justified. The iRhom2 cub-domain shows a non-variant behaviour, which connects to the stability and compact folding of the domain. The secondary structure of the cub-domain consists of both α and β components, which leads to a lower radius of gyration as compared to complete cytoplasmic N-terminus domain.

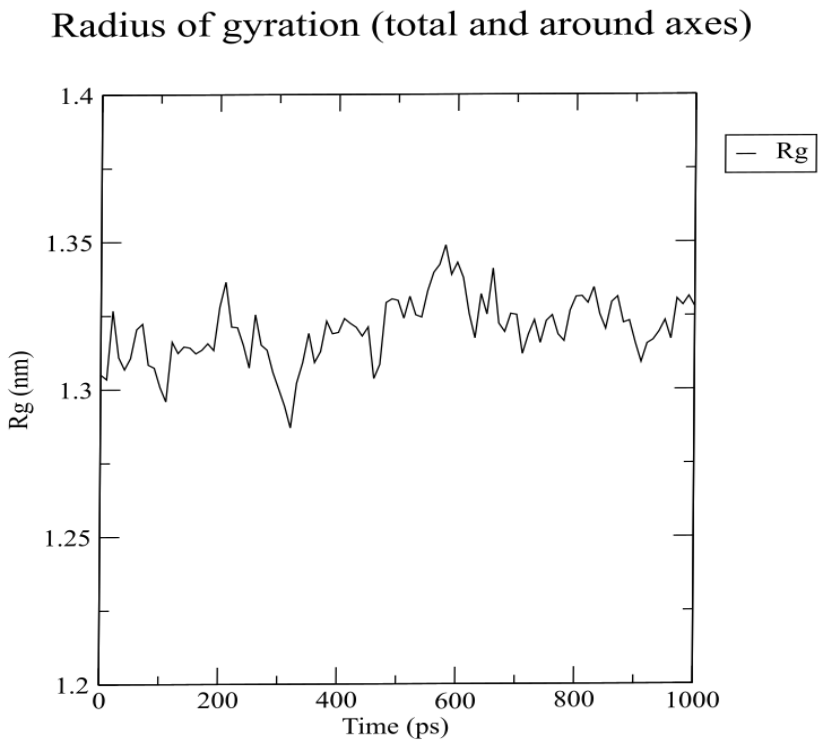

Figure 4.27 Radius of gyration plot of iRhom2 cub-domain

4.5.6 Structural stability w.r.t Root Mean Square Deviation (RMSD)

After the successful completion of production MD, the next step was to evaluate the whole system. The RMSD was calculated; GROMACS built-in utility ‘rms’ was used for this objective. Finally, the backbone was selected for both the group and least-square fits. The output was converted to ‘ns’ (nanoseconds) even the trajectory was written in ‘ps’ (picoseconds). This step was done to avoid dealing with the complicated trajectory values generated after RMSD calculation. The results were shown as a graph, which states RMSD relative to the energy minimized structure. Figure 4.28 and Figure 4.29 explains calculated RMSD for iRhom2 domains, the cytoplasmic N-terminus domain, and the cub-domain. The RMSD plot (Figure 4.28) of the cytoplasmic N-terminus domain shows that RMSD levels off at 0.1 nm (1 Å) at 0.0 ns, which indicates this domain was stable at the start; however, it can be observed that soon after reaching 0.15 nm, the RMSD value keep on increasing concerning time. During the time frame of 1 ns, the RMSD estimation was terminated at 0.4 nm. However, if the MD were run for more than one ns, the RMSD slop

may keep on increasing. The plot explains the unstable behaviour of iRhom2 complete N-terminus cytoplasmic domain.

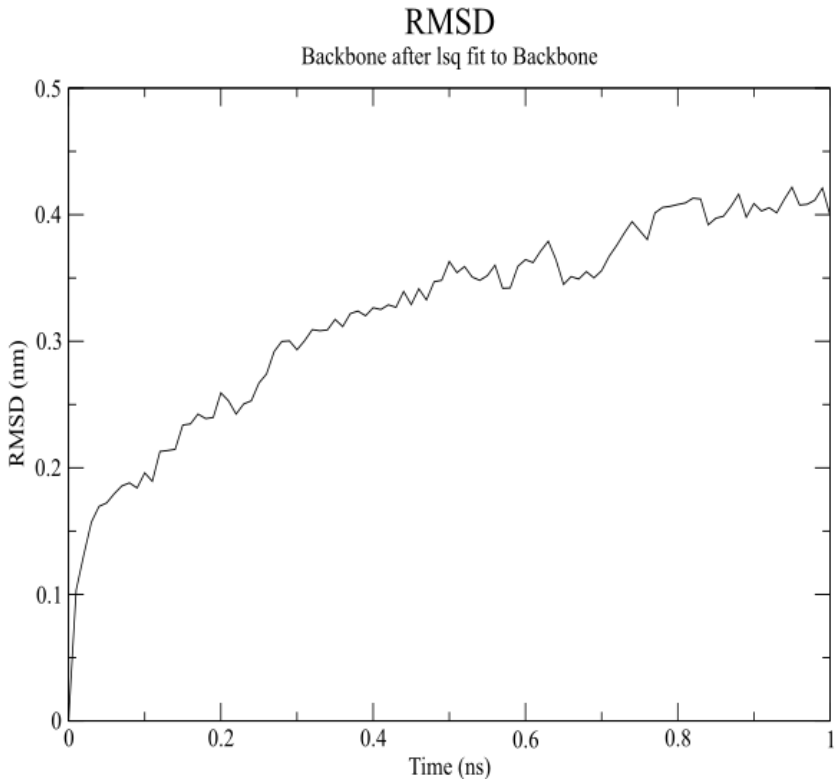

Figure 4.28 RMSD plot of iRhom2 complete cytoplasmic N-terminus

The RMSD estimation plot (Figure 4.29) for the iRhom2 cub-domain shows the RMSD value levels off to 0.1 nm at 0.0 ns, as the complete cytoplasmic N-terminus domain did at the beginning of the estimation process. Later at around 0.4 ns, the RMSD was 0.2 nm, the cub-domain structure shows a stable behaviour expected from this domain. However, after passing the 0.4 ns time frame, a shift and increased fluctuation in RMSD value can be observed. Finally, the RMSD value reached slightly above 0.25 nm, the highest point of this estimation. The slop stabilised around 0.25 nm from 0.5 to 0.8 ns, then the value dropped to 0.2 nm, and after that, it kept on increasing above 0.25 nm. Thus, it is concluded that the geometry of cub-domain model is not well structured however, the final slope of cub-domain is stabler as compared to the complete N-terminus domain [11], [22].

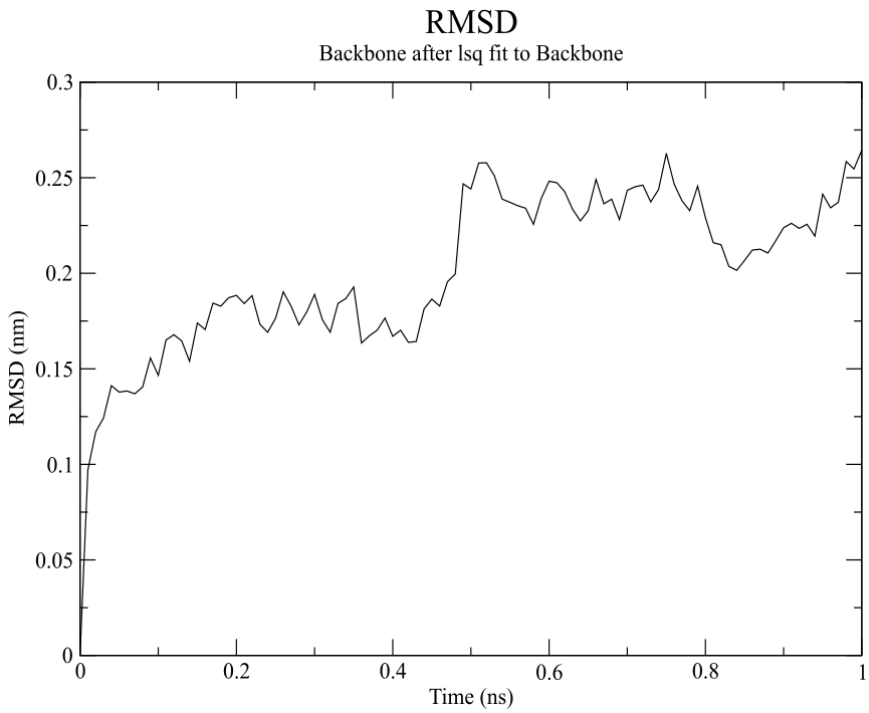

Figure 4.29 RMSD plot of iRhom2 cub-domain

However, for the stability profiling of iRhom2 domains, we cannot completely rely on RMSD estimation, as the value of RMSD does not draw reliable conclusions unless the MD simulation is run for a significantly longer time [93].

4.6 Ramachandran plot evaluation after MD

To further compare both iRhom2 domains, Ramachandran plots were generated using procheck. Ramachandran plot also shows 72.5% core residues, 20.7% residues in the allowed region, 3.4% in the generously allowed region and 0.2% in the disallowed region (Figure 4.30). These results by PROCHECK [63] explains that this domain structure has more steric clashes after MD simulation as compared to the raw N-terminus cytoplasmic domain of iRhom2 and an unstable geometrical assembly. The result of this MD align with the findings of *Houser et al* [36], i.e. wild-type iRhom2 in mammals exhibits unstable behaviour.

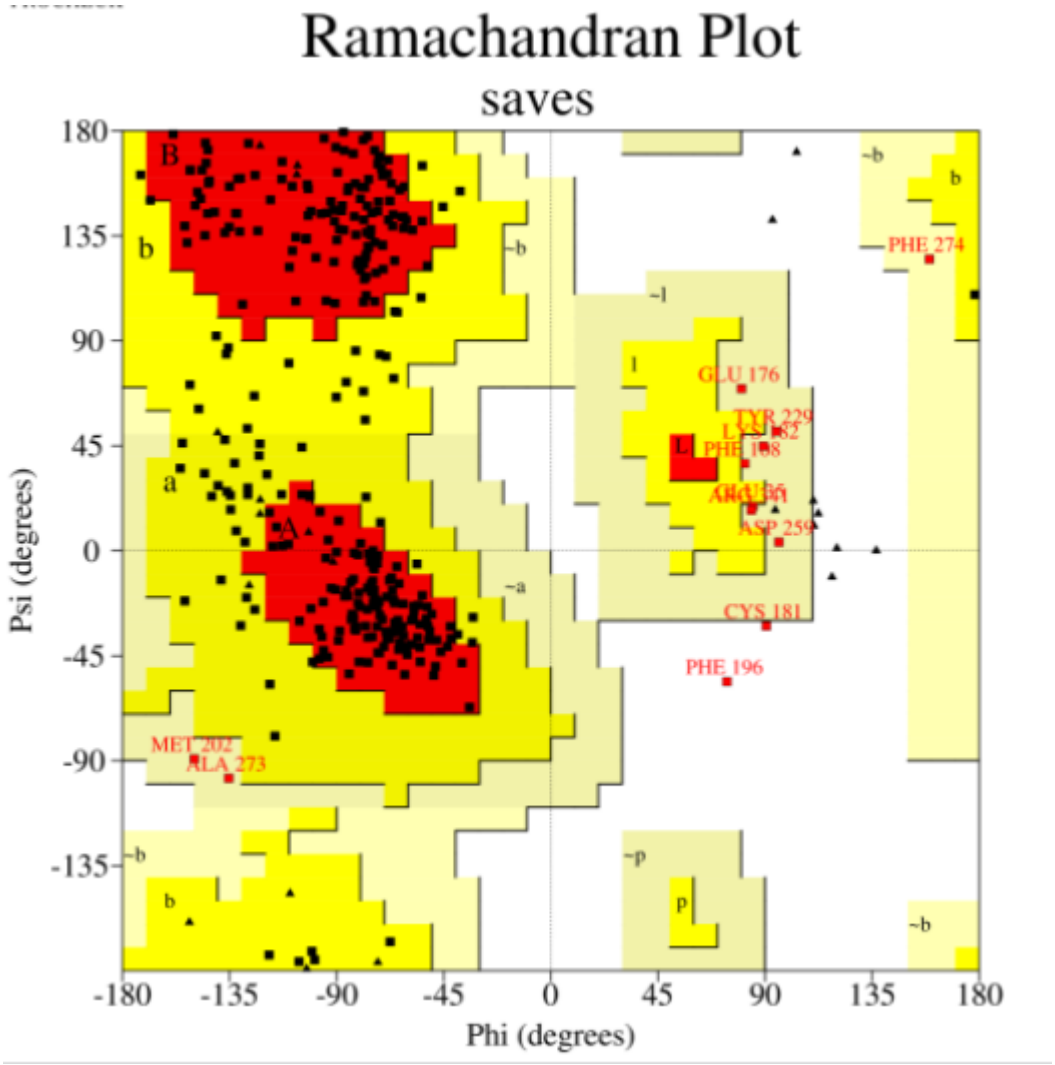

Figure 4.30 Ramachandran plot of complete cytoplasmic N-terminus iRhom2 domain (1-351) after MD simulation

PROCHECK server shows five errors, two warnings and only one pass evaluation out of eight evaluations. Ramachandran plot also shows 79.4% core residues, 19.0% residues in the allowed region, 1.6% in the generously allowed region and 0% in the disallowed region (Figure 4.31). These results by PROCHECK [63] explains that this domain structure has more steric clashes as compared to cub-domain before MD simulation and a did not have a desirable geometrical assembly. This outcome also negates *Hosur et al.* [36] theory and align with the findings of *Siggs at al.* [34]

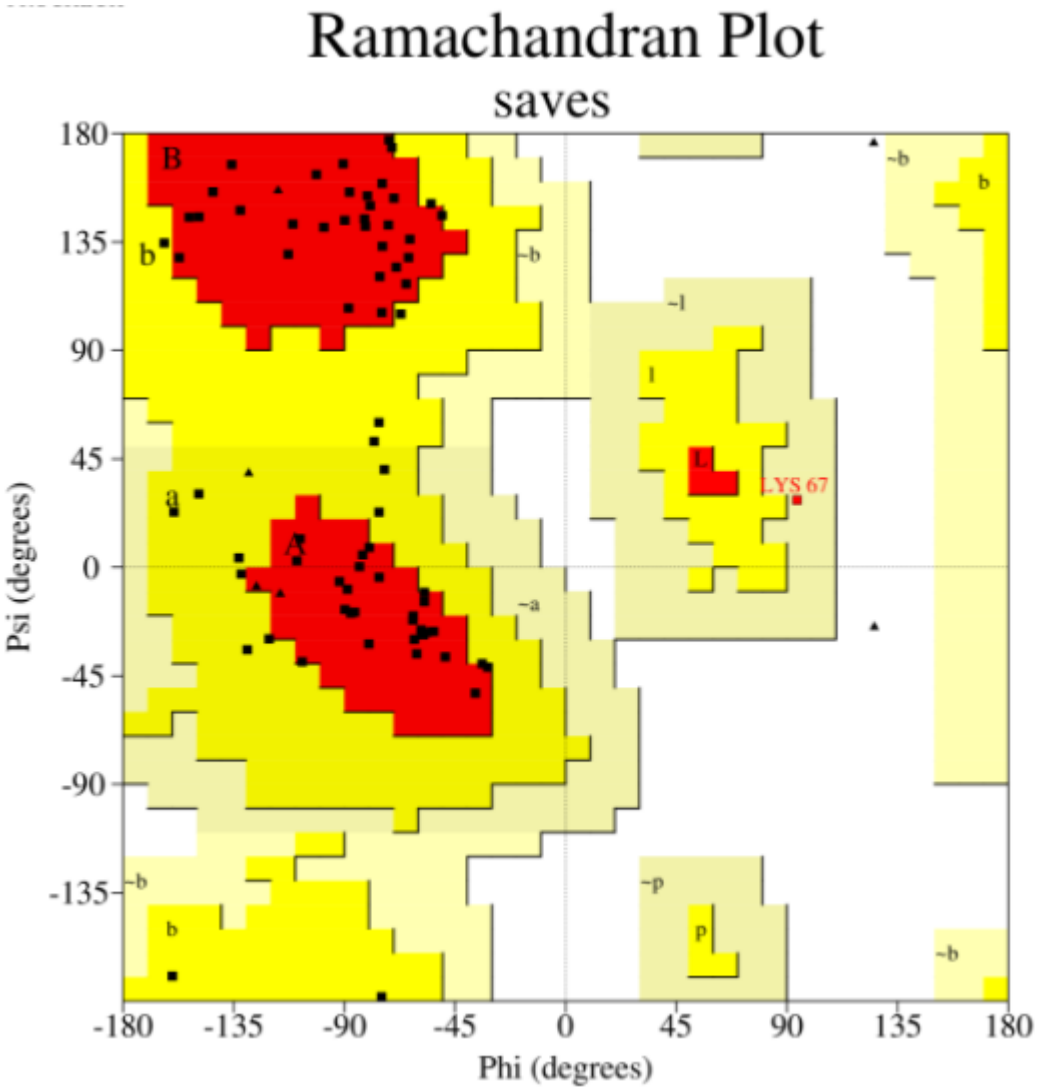

Figure 4.31 Ramachandran plot of cub-mutated iRhom2 domain (268-351) after MD simulation

Discussion

iRhoms are inactive family members of Rhomboid proteins [9]. Like rhomboids, iRhoms are known to be conserved in most mammals [41]. The function of iRhom2 is highly linked with its name ‘inactive rhomboid’, as it is responsible for the regulation of TNF-α converting enzyme (ADAM17), which is involved in the cleavage of TNF-α and EGFR ligands at the cell membrane [94]. The cleavage of TNF-α leads to the maturation of ADAM17 OR TACE, which promotes the release of amphiregulin which contributes to cell proliferation, inflammation and promotes controlled cellular healing under normal circumstances. After the cleavage of TNF-α and EGFR ligands, iRhom2 degrades in ER and remains unstable [1].

However, the deletion of 286 amino acids of iRhom2 cytoplasmic N-terminus leads to cub-mutation (observed in mice). This mutation leads towards the stability of cub-mutated iRhom2 [22]. After the cleavage of TNF-α and EGFR, iRhom2 leads to the hyperactivation of the EGFR pathway, causing abnormal growth of cells, increased healing (super healing), different types of cancer and many other fatal diseases (Table 1.3) [22], [32]. Thus, this mutant version of iRhom2 could be proved as a potential target.

Another study reported some facts contradictory to the findings of [22]. During the same year, [11] challenged and tested the concept of iRhom2 stability due to the cub-mutation in mice. This study was done using KO studies on mice and concluded that cub-mutation in iRhom2 does not lead to the hyperactivation of the EGFR pathway. However, this study reported the downregulation of the EGFR pathway. In addition, the cub mutant iRhom2 embryonic fibroblasts release less amphiregulin than wild-type iRhom2, and that the cub mutant form of iRhom2 is less able than wild-type iRhom2 to bind to TACE and promote its maturation.

This contradiction still exists today, and not many studies have been performed to clarify

this research question. iRhom2 is a potential target in many drug discovery processes due to its vital role in regulating the EGFR pathway [25], [32], [95]; therefore, the hypothesis of this research is based on solving the contradiction about the stability of iRhom2.

This hypothesis was studied using different *in-silico* techniques, including global alignment (Needleman-Wunsch) of mice iRhom2 with human iRhom2 because the research question of this study was based on mice iRhom2. Furthermore, the iRhoms are highly conserved in mammals, so the results of global alignment showed high sequence identity among both mammalian proteins. Therefore, the hypothesis of this study can be tested on human iRhom2 [11], [22], [32].

The crystal structure of iRhom2 has not been solved yet [3]; therefore, different 3D protein modelling approaches (discussed earlier in chapter 3) were used to model iRhom2. I-TASSER (fold-recognition based) and Robetta (ab-initio based) servers were mainly used for modelling. However, the structures modelled by I-TASSER were not of good quality. The evaluations done on structures predicted using I-TASSER by SAVES server indicates that protein sequences with less than 18% sequence similarity do not result with correctly folded (accurate geometries. However, the iRhom2 domain structures predicted by Robetta resulted in better-folded structures, with the least stereochemical clashes. This result indicates that ab-initio and de-novo-based modelling techniques can predict more accurate structures for complex transmembrane proteins with the least sequence similarities. Furthermore, the predicted structures of iRhom2 domains by the Robetta server resembled the theoretical prediction of the iRhom2 structures mentioned in the literature (Figure 4.6) (Figure 4.7).

For stability profiling, MD simulation was performed using GROMACS. The MD simulation was run for one ns, which was the primary time frame to run MD for simple proteins in a solvent. [75]. The results of the MD simulation helped to draw a clear idea about the structural stability of iRhom2. The pre-processing for the MD simulation and production MD results aligned with the hypothesis based on the research of *hosur et al*. [22]. The energy minimization and equilibration of both iRhom2 domains of interest showed expected behaviour. The complete cytoplasmic N-terminus domain showed

fluctuations throughout the one ‘ns’ run; moreover, the unstable behaviour of this domain can be observed during production MD. The RMSD and radius of gyration results suggest that this domain was not folded correctly; the density of this domain was low, and the RMSD value was higher than normal. Thus, the domain was losing its compactness throughout one ‘ns’ MD simulation. The loss of compactness leads to the deduction that wild-type iRhom2 is unstable, degraded in ER after converting an EGFR ligand ‘Pro AREG’ to mature AREG. [22], [32]. The unstable behaviour of wild-type iRhom2 was also observed during the structural evaluations by the SAVES server; moreover, the MD simulation further improves the reliability of the results.

The MD simulation results of the cub-domain align with the theory of *Hosur et al.* [22]. The study of *Hosur et al.* emphasizes the structural stability of cub-mutated iRhom2 and the excess release of AREG. The NPT and NVT equilibration indicated that the structure of the cub-domain predicted by Robetta was equilibrated sooner than the complete cytoplasmic N-terminus. The production MD run results provided a clearer understanding of the stability of cub-mutated iRhom2. The plots of the radius of gyration and RMSD showed a relatively consistent slop than the complete cytoplasmic N-terminus domain. The plot of the radius of gyration was consistent, which indicates the compacts of the cub-domain structure remained intact during the complete MD run. The compactness of the cub-domain structure leads to availability; however, the slop of RMSD did not remain stable until the end; this part indicates that there might be few steric clashes left while modelling the structure of this domain, or the geometry is not correct.

The MD results do not fit in with the theory of [34]. This study reported that the loss of 268 amino acids from the N-terminus of iRhom2 leads to the gain of function mutation in iRhom2. This study provided an insight into the importance of these 268 amino acids. Moreover, this part of iRhom2 is a reported sight for many other important point mutations and interactors as well, like tylosis (Hs: I186T, P189L, D188N/Y) (Ms: P159L) and FRMD8 (200-300) [10], [32], [41]. The active sites responsible for the normal function of iRhom2 might be located between these 268 amino acids; therefore, the loss of these amino acids leads to the uncontrolled ability of iRhom2 to attach with TACE and thus hyperactivates EGFR [22], [32].

This experiment provided an insight into the complex structural conformation of iRhom2. However, it is beyond the scope of this study to predict a good quality and correct 3D structure of complete wild-type and cub-cub-mutated iRhom2. Initially, the goal was to predict the complete 3D structures of wild-type and cub mutated iRhom2; however, the results of complete structure predictions by threading and fold recognition lacked the expected structural accuracy.

This result is due to the limitation of ab-initio based protein modelling techniques, which are not efficient enough to predict 3D structures of large proteins. The ab-initio protein modelling is preferred to predict small proteins due to the exhaustive process of ab-initio modelling. [54] Therefore, the outcome of this study leads to the direction of 3D structure prediction of only crucial iRhom2 domains. The methodological choices made for this study were constrained by the lack of studies available on this topic. The absence of the crystal structure of mice and human iRhom2 impacted the reliability of 3D structures predicted by Robetta.

The MD was run for ‘1 ns’ for the structures predicted by Robetta by following the GROMACS MD tutorials for ‘protein in solvent.’ [75]. The results of MD simulation were impacted by this methodological choice; the plots of RMSD showed an unexpected disrupt in the slop at the middle and end of both iRhom2 domains. Thus, we got an answer about the stability of the cub and complete cytoplasmic N-terminus domain of iRhom2. However, this outcome left a void in answering this research question more efficiently. The contradiction among the studies of *Hosur et al* [36]. and *siggs et al* [34]. is resolved now. The cub-mutated iRhom2 can be used as a potential target against many carcinomas and other diseases linked with cub-mutated iRhom2. However, more studies and knowledge are needed on the structural analysis of iRhom2 to establish a stronger theory for the stability of cub-mutated iRhom2.

Conclusion and Future Perspectives

This research was aimed to resolve the conflict about the stability of the iRhom2 structure in *homo sapiens*. The studies about the stability of mice iRhom2 were considered a backbone of this study due to the lack of research on iRhom2 in *homo sapiens*. The KO studies in mice iRhom2 concluded that the loss of the first 268 amino acids from the N-terminus leads to a gain of function mutation known as cub-mutation. This mutated protein can perform many functions that unstable or wild-type iRhom could not perform, including hyperactivation of the EGFR pathway, which leads to the induction of different types of cancers. However, another study disagrees with this theory and concluded that the loss of 268 amino acids did not hyperactivate the EGFR pathway but downregulates it. The contradiction among both of these studies needed to be solved, for the better understanding of role of iRhom2 in regulating The EGFR pathway and the induction of different types of cancers in *homo sapiens*. To accomplish this aim, the 3D structure prediction of important iRhom2 domains was performed. Robetta predicted the 3D structures of iRhom2, the cub-domain (268-350) and the complete cytoplasmic N-terminus domain. The structure evaluations by SAVES provided an initial insight into the structural stability of both domains. The cub-domain showed better statistics and the least outliers. This research clearly illustrated that the first 286 amino acids of iRhom2 are extremely critical for the proper functioning of wild-type iRhom2.

MD Simulations for both iRhom2 domains were performed to understand the structural stability of these domains furthermore. The MD simulation of these domains concluded that the cub-domain of iRhom2 had appeared more stable than the complete cytoplasmic N-terminus domain when both domains were compared side by using the same MD simulation parameters.

This research project effectively resolved the conflict about the stability of cub-mutated iRhom2. However, based on the results, future research can further enhance the picture of the structural stability of iRhom2 by increasing the time frame of MD simulations using high computational resources. Furthermore, to understand the functional stability of

iRhom2, both domains can be interacted with AREG via the protein-protein interaction technique. This additional step can further validate the hyperactivation of the EGFR pathway by observing the binding affinity of AREG with both of these domains.

References

[1] M. Freeman, “Rhomboids: 7 years of a new protease family,” *Seminars in Cell and Developmental Biology*, vol. 20, no. 2. Elsevier Ltd, pp. 231–239, 2009.

[2] E. Erez, D. Fass, E. B.-N. (London), : “How intramembrane proteases bury hydrolytic reactions in the membrane: Membrane protein biophysics,” *pascal-francis.inist.fr,2019*.

[3] M. Freeman, “The Rhomboid-Like Superfamily: Molecular Mechanisms and Biological Roles,” *Annu. Rev. Cell Dev. Biol.*, vol. 30, no. 1, pp. 235–254, Oct. 2014.

[4] A. Tichá, B. Collis, and K. Strisovsky, “The Rhomboid Superfamily: Structural Mechanisms and Chemical Biology Opportunities,” *Trends Biochem. Sci.*, vol. 43, no. 9, pp. 726–739, Sep. 2018.

[5] N. Bergbold and M. K. Lemberg, “Emerging role of rhomboid family proteins in mammalian biology and disease,” *Biochimica et Biophysica Acta - Biomembranes*, vol. 1828, no. 12. Elsevier, pp. 2840–2848, Dec-2013.

[6] W. W. Luo *et al.*, “iRhom2 is essential for innate immunity to DNA viruses by mediating trafficking and stability of the adaptor STING,” *Nat. Immunol.*, vol. 17, no. 9, pp. 1057–1066, Aug. 2016.

[7] W. W. Luo and H. B. Shu, “Emerging roles of rhomboid-like pseudoproteases in inflammatory and innate immune responses,” *FEBS Letters*, vol. 591, no. 20. Wiley Blackwell, pp. 3182–3189, Oct-2017.

[8] M. Freeman, C. Adrian," Tumor necrosis factor signaling requires iRhom2 to promote trafficking and activation of TACE," Cell, vol. 145, no. 1, pp. 79–91, Jan. 2012.

[9] M. Y. Lee, K. H. Nam, and K. C. Choi, “iRhoms; Its functions and essential roles,” *Biomol. Ther.*, vol. 24, no. 2, pp. 109–114, 2016.

[10] M. Zettl, C. Adrain, K. Strisovsky, V. Lastun, and M. Freeman, “Rhomboid family pseudoproteases use the ER quality control machinery to regulate intercellular signaling,” *Cell*, vol. 145, no. 1, pp. 79–91, Apr. 2011.

[11] O. M. Siggs, A. Grieve, H. Xu, P. Bambrough, Y. Christova, and M. Freeman, “Genetic interaction implicates iRhom2 in the regulation of EGF receptor signalling in mice,” *Biol. Open*, vol. 3, no. 12, pp. 1151–1157, 2014.

[12] X. Li *et al.*, “iRhoms 1 and 2 are essential upstream regulators of ADAM17-dependent EGFR signaling,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 112, no. 19, pp. 6080–6085, May 2015.

[13] W. Lee *et al.*, “IRhom1 regulates proteasome activity via PAC1/2 under ER stress,” *Sci. Rep.*, vol. 5, no. 1, pp. 1–14, Jun. 2015.

[14] A. Schweinitz *et al.*, “Design of novel and selective inhibitors of urokinase-type plasminogen activator with improved pharmacokinetic properties for use as antimetastatic agents,” *J. Biol. Chem.*, vol. 279, no. 32, pp. 33613–33622, 2004.

[15] K. Almholt *et al.*, “Reduced metastasis of transgenic mammary cancer in urokinase-deficient mice,” *Int. J. Cancer*, vol. 113, no. 4, pp. 525–532, 2005.

[16] H. Zou, S. M. Thomas, Z. W. Yan, J. R. Grandis, A. Vogt, and L. Y. Li, “Human rhomboid family-1 gene RHBDF1 participates in GPCR-mediated transactivation of EGFR growth signals in head and neck squamous cancer cells,” *FASEB J.*, vol. 23, no. 2, pp. 425–432, 2009.

[17] Z. Zhou *et al.*, “Human rhomboid family-1 suppresses oxygen-independent degradation of hypoxia-inducible factor-1a in breast cancer,” *Cancer Res.*, vol. 74, no. 10, pp. 2719–2730, May 2014.

[18] H. Yuan *et al.*, “RHBDF1 regulates APC-mediated stimulation of the epithelial-to-mesenchymal transition and proliferation of colorectal cancer cells in part via the Wnt/$β$-catenin signalling pathway,” *Exp. Cell Res.*, vol. 368, no. 1, pp. 24–36, Jul.

2018.

[19] W. Bin Zhang *et al.*, “Glutamine ameliorates lipopolysaccharide-induced cardiac dysfunction by regulating the toll-like receptor 4/mitogen-activated protein kinase/nuclear factor-$κ$B signaling pathway,” *Exp. Ther. Med.*, vol. 14, no. 6, pp. 5825–5832, Dec. 2017.

[20] D. N. Barnette, T. J. Cahill, M. Gunadasa-Rohling, C. A. Carr, M. Freeman, and P. R. Riley, “iRhom2-mediated proinflammatory signalling regulates heart repair following myocardial infarction,” *JCI insight*, vol. 3, no. 3, Feb. 2018.

[21] P. D. A. Issuree *et al.*, “iRHOM2 is a critical pathogenic mediator of infammatory arthritis,” *J. Clin. Invest.*, vol. 123, no. 2, pp. 928–932, Feb. 2013.

[22] V. Hosur, K. R. Johnson, L. M. Burzenski, T. M. Stearns, R. S. Maser, and L. D. Shultz, “Rhbdf2 mutations increase its protein stability and drive EGFR hyperactivation through enhanced secretion of amphiregulin,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 111, no. 21, 2014.

[23] D. Blaydon, S. Etheridge, … J. R.-T. A. J. of, and undefined 2012, “RHBDF2 mutations are associated with tylosis, a familial esophageal cancer syndrome,” *Elsevier*.

[24] S. Saarinen *et al.*, “Analysis of a Finnish family confirms RHBDF2 mutations as the underlying factor in tylosis with esophageal cancer,” *Springer*.

[25] A. Ellis, J. M. Risk, T. Maruthappu, and D. P. Kelsell, “Tylosis with oesophageal cancer: Diagnosis, management and molecular mechanisms,” *Orphanet Journal of Rare Diseases*, vol. 10, no. 1. BioMed Central Ltd., p. 126, 29-Sep-2015.

[26] S. K. Maney *et al.*, “Deletions in the cytoplasmic domain of iRhom1 and iRhom2 promote shedding of the TNF receptor by the protease ADAM17,” *Sci. Signal.*, vol. 8, no. 401, Nov. 2015.

[27] A. G. Grieve, H. Xu, U. Künzel, P. Bambrough, B. Sieber, and M. Freeman, “Phosphorylation of iRhom2 at the plasma membrane controls mammalian TACE-dependent inflammatory and growth factor signalling,” *Elife*, vol. 6, Apr. 2017.

[28] L. Yang *et al.*, “iRhom2 Uncv mutation blocks bulge stem cells assuming the fate of hair follicle,” *Arch. Dermatol. Res.*, vol. 308, no. 7, pp. 503–510, 2016.

[29] U. Künzel, A. G. Grieve, Y. Meng, B. Sieber, S. A. Cowley, and M. Freeman, “FRMD8 promotes inflammatory and growth factor signalling by stabilising the iRhom/ADAM17 sheddase complex,” *Elife*, vol. 7, Jun. 2018.

[30] I. Oikonomidi *et al.*, “ITap, a novel irhom interactor, controls TNF secretion by policing the stability of iRhom/TACE,” *Elife*, vol. 7, Jun. 2018.

[31] I. Oikonomidi *et al.*, “ITap, a novel irhom interactor, controls TNF secretion by policing the stability of iRhom/TACE,” *Elife*, vol. 7, Jun. 2018.

[32] I. Dulloo, S. Muliyil, and M. Freeman, “The molecular, cellular and pathophysiological roles of irhom pseudoproteases,” *Open Biol.*, vol. 9, no. 3, 2019.

[33] T. Maretzky *et al.*, “iRhom2 controls the substrate selectivity of stimulated ADAM17-dependent ectodomain shedding,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 110, no. 28, pp. 11433–11438, Jul. 2013.

[34] O. M. Siggs, A. Grieve, H. Xu, P. Bambrough, Y. Christova, and M. Freeman, “Genetic interaction implicates iRhom2 in the regulation of EGF receptor signalling in mice,” *Biol. Open*, vol. 3, no. 12, pp. 1151–1157, Dec. 2014.

[35] R. Idrees, S. Fatima, J. Abdul-Ghafar, A. Raheem, and Z. Ahmad, “Cancer prevalence in Pakistan: Meta-analysis of various published studies to determine variation in cancer figures resulting from marked population heterogeneity in different parts of the country,” *World J. Surg. Oncol.*, vol. 16, no. 1, Jul. 2018.

[36] V. Hosur, K. R. Johnson, L. M. Burzenski, T. M. Stearns, R. S. Maser, and L. D.

Shultz, “Rhbdf2 mutations increase its protein stability and drive EGFR hyperactivation through enhanced secretion of amphiregulin,” *Natl. Acad Sci.*

[37] X. Li *et al.*, “iRhoms 1 and 2 are essential upstream regulators of ADAM17-dependent EGFR signaling,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 112, no. 19, pp. 6080–6085, May 2015.

[38] W.-W. Luo *et al.*, “iRhom2 is essential for innate immunity to RNA virus by antagonizing ER- and mitochondria-associated degradation of VISA,” *PLOS Pathog.*, vol. 13, no. 11, p. e1006693, Nov. 2017.

[39] M. K. Lemberg and C. Adrain, “Inactive rhomboid proteins: New mechanisms with implications in health and disease,” *Semin. Cell Dev. Biol.*, vol. 60, pp. 29–37, 2016.

[40] U. Sahin *et al.*, “Distinct roles for ADAM10 and ADAM17 in ectodomain shedding of six EGFR ligands,” *J. Cell Biol.*, vol. 164, no. 5, pp. 769–779, Mar. 2004.

[41] Y. Christova, C. Adrain, P. Bambrough, A. Ibrahim, and M. Freeman, “Mammalian iRhoms have distinct physiological functions including an essential role in TACE regulation,” *EMBO Rep.*, vol. 14, no. 10, pp. 884–890, Oct. 2013.

[42] Y. Christova, C. Adrain, P. Bambrough, A. Ibrahim, and M. Freeman, “Mammalian iRhoms have distinct physiological functions including an essential role in TACE regulation,” *EMBO Rep.*, vol. 14, no. 10, pp. 884–890, Oct. 2013.

[43] H. Zou, S. M. Thomas, Z. Yan, J. R. Grandis, A. Vogt, and L. Li, “ Human rhomboid family-1 gene RHBDF1 participates in GPCR-mediated transactivation of EGFR growth signals in head and neck squamous cancer cells ,” *FASEB J.*, vol. 23, no. 2, pp. 425–432, Feb. 2009.

[44] S. Stadlmann, J. Pollheimer, … K. R.-W. repair and, and undefined 2006, “Response of human peritoneal mesothelial cells to inflammatory injury is regulated by interleukin-1b and tumor necrosis factor-a,” *Wiley Online Libr.*

[45] S. Stadlmann, K. Renner, J. Pollheimer, … P. M.-C. biochemistry and, and undefined 2006, "Preserved coupling of oxidative phosphorylation but decreased mitochondrial respiratory capacity in IL-1β-treated human peritoneal mesothelial cells," *Springer*.

[46] D. N. N. Barnette, T. J. J. Cahill, M. Gunadasa-Rohling, C. A. A. Carr, M. Freeman, and P. R. R. Riley, "iRhom2-mediated proinflammatory signalling regulates heart repair following myocardial infarction," *JCI insight*, vol. 3, no. 3, Feb. 2018.

[47] C. B.-N. reviews M. cell biology and undefined 2005, "ADAMs: key components in EGFR signalling and development," *nature.com*.

[48] S. S. Cao and R. J. Kaufman, "IRhoms: ERADicating the Messenger in Growth Control Signaling," *Dev. Cell*, vol. 20, no. 4, pp. 414–416, 2011.

[49] C. Adrain, K. Strisovsky, M. Zettl, L. Hu, M. K. Lemberg, and M. Freeman, "Mammalian EGF receptor activation by the rhomboid protease RHBDL2," *EMBO Rep.*, vol. 12, no. 5, pp. 421–427, May 2011.

[50] E. J. Greenblatt, J. A. Olzmann, and R. R. Kopito, "Derlin-1 is a rhomboid pseudoprotease required for the dislocation of mutant α-1 antitrypsin from the endoplasmic reticulum," *Nat. Struct. Mol. Biol.*, vol. 18, no. 10, pp. 1147–1152, Oct. 2010.

[51] X. Huang, "On global sequence alignment," *Bioinformatics*, vol. 10, no. 3, pp. 227–235, Jun. 1994.

[52] R. C. Edgar and S. Batzoglou, "Multiple sequence alignment," *Current Opinion in Structural Biology*, vol. 16, no. 3. Elsevier Current Trends, pp. 368–373, 01-Jun-2006.

[53] J. Moult, J. T. Pedersen, R. Judson, and K. Fidelis, "A large-scale experiment to assess protein structure prediction methods," *Proteins Struct. Funct. Genet.*, vol. 23, no. 3, pp. ii–iv, Nov. 1995.

[54] J. Lee, S. Wu, and Y. Zhang, “Ab initio protein structure prediction,” in *From Protein Structure to Function with Bioinformatics*, Springer Netherlands, 2009, pp. 3–25.

[55] E. P. Carpenter, K. Beis, A. D. Cameron, and S. Iwata, “Overcoming the challenges of membrane protein crystallography,” *Current Opinion in Structural Biology*, vol. 18, no. 5. Elsevier, pp. 581–586, Oct-2008.

[56] D. E. Kim, D. Chivian, and D. Baker, “Protein structure prediction and analysis using the Robetta server,” *Nucleic Acids Res.*, vol. 32, no. WEB SERVER ISS., p. W526, Jul. 2004.

[57] Y. Zhang, “I-TASSER server for protein 3D structure prediction,” *BMC Bioinformatics*, vol. 9, no. 1, p. 40, Jan. 2008.

[58] K. Ginalski and L. Rychlewski, “Protein Structure Prediction of CASP5 Comparative Modelling and Fold Recognition Targets Using Consensus Alignment Approach and 3D Assessment,” in *Proteins: Structure, Function and Genetics*, 2003, vol. 53, no. SUPPL. 6, pp. 410–417.

[59] R. Sato and T. Ishida, “Protein model accuracy estimation based on local structure quality assessment using 3D convolutional neural network,” *PLoS One*, vol. 14, no. 9, p. e0221347, 2019.

[60] C. Colovos and T. O. Yeates, “Verification of protein structures: Patterns of nonbonded atomic interactions,” *Protein Sci.*, vol. 2, no. 9, pp. 1511–1519, 1993.

[61] J. U. Bowie, R. Lüthy, and D. Eisenberg, “A method to identify protein sequences that fold into a known three-dimensional stucture,” *Science (80-.).*, vol. 253, no. 5016, pp. 164–170, 1991.

[62] R. A. Laskowski, M. W. MacArthur, D. S. Moss, and J. M. Thornton, “PROCHECK: a program to check the stereochemical quality of protein structures,” *J. Appl. Crystallogr.*, vol. 26, no. 2, pp. 283–291, Apr. 1993.

[63] R. A. Laskowski, J. A. C. Rullmann, M. W. MacArthur, R. Kaptein, and J. M. Thornton, “AQUA and PROCHECK-NMR: Programs for checking the quality of protein structures solved by NMR,” *J. Biomol. NMR*, vol. 8, no. 4, pp. 477–486, 1996.

[64] R. A. Laskowski and G. J. Swaminathan, “Problems of Protein Three-Dimensional Structures,” *Compr. Med. Chem. II*, vol. 3, pp. 531–550, Jan. 2007.

[65] G. Vriend, “WHAT IF: A molecular modeling and drug design program,” *J. Mol. Graph.*, vol. 8, no. 1, pp. 52–56, 1990.

[66] A. Hospital, J. R. Goñi, M. Orozco, and J. L. Gelpí, “Molecular dynamics simulations: Advances and applications,” *Advances and Applications in Bioinformatics and Chemistry*, vol. 8, no. 1. Dove Medical Press Ltd, pp. 37–47, 2015.

[67] J. W. Ponder and D. Caset, “FORCE FIELDS FOR PROTEIN SIMULATIONS,” 2003.

[68] M. Rueda *et al.*, “A consensus view of protein dynamics,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 104, no. 3, pp. 796–801, Jan. 2007.

[69] M. A. González, “Force fields and molecular dynamics simulations,” *Collect. SFN*, vol. 12, pp. 169–200, 2011.

[70] C. I. Bayly *et al.*, “A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules,” *J. Am. Chem. Soc.*, vol. 117, no. 19, pp. 5179–5197, 1995.

[71] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus, “CHARMM: A program for macromolecular energy, minimization, and dynamics calculations,” *J. Comput. Chem.*, vol. 4, no. 2, pp. 187–217, Jun. 1983.

[72] W. R. P. Scott *et al.*, “The GROMOS biomolecular simulation program package,”

J. Phys. Chem. A, vol. 103, no. 19, pp. 3596–3607, May 1999.

[73] M. J. Robertson, J. Tirado-Rives, and W. L. Jorgensen, "Improved Peptide and Protein Torsional Energetics with the OPLS-AA Force Field," *J. Chem. Theory Comput.*, vol. 11, no. 7, pp. 3499–3509, Jun. 2015.

[74] H. J. C. Berendsen, J. P. M. Postma, W. F. Van Gunsteren, A. Dinola, and J. R. Haak, "Molecular dynamics with coupling to an external bath," *J. Chem. Phys.*, vol. 81, no. 8, pp. 3684–3690, 1984.

[75] J. A. Lemkul, "Free Energy Calculations: Methane in Water GROMACS Tutorial Lysozyme in water Based on the tutorial created."

[76] M. J. Abraham *et al.*, "Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers," *SoftwareX*, vol. 1–2, pp. 19–25, Sep. 2015.

[77] T. V Pogorelov and M. Hallock, "Introduction to Molecular Mechanics with MOE: Why GFP needs a beta-barrel shell?," 2016.

[78] W. Humphrey, A. Dalke, and K. Schulten, "VMD: Visual molecular dynamics," *J. Mol. Graph.*, vol. 14, no. 1, pp. 33–38, Feb. 1996.

[79] A. D. Astuti and A. B. Mutiara, "Performance Analysis on Molecular Dynamics Simulation of Protein Using GROMACS," Dec. 2009.

[80] A. Amadei, A. B. M. Linssen, and H. J. C. Berendsen, "Essential dynamics of proteins," *Proteins Struct. Funct. Bioinforma.*, vol. 17, no. 4, pp. 412–425, 1993.

[81] S. Patodia, A. Bagaria, and D. Chopra, "Physical Chemistry & Biophysics Molecular Dynamics Simulation of Proteins : A Brief Overview," vol. 4, no. 6, pp. 4–7, 2014.

[82] B. P. Van Eijck, "Pressure calculation in molecular dynamics simulations of

molecular crystals," *Mol. Simul.*, vol. 13, no. 3, pp. 221–230, Sep. 1994.

[83] M. Hutt, T. Kulschewski, and J. Pleiss, "Molecular modelling of the mass density of single proteins," *J. Biomol. Struct. Dyn.*, vol. 30, no. 3, pp. 318–327, Jul. 2012.

[84] "Energy Minimization - an overview | ScienceDirect Topics." [Online]. Available: https://www.sciencedirect.com/topics/biochemistry-genetics-and-molecular-biology/energy-minimization. [Accessed: 30-Mar-2021].

[85] M. Y. Lobanov, N. S. Bogatyreva, and O. V. Galzitskaya, "Radius of gyration as an indicator of protein structure compactness," *Mol. Biol.*, vol. 42, no. 4, pp. 623–628, Aug. 2008.

[86] E. C. Meng, E. F. Pettersen, G. S. Couch, C. C. Huang, and T. E. Ferrin, "Tools for integrated sequence-structure analysis with UCSF Chimera," *BMC Bioinformatics*, vol. 7, no. 1, p. 339, Jul. 2006.

[87] A. Bateman *et al.*, "UniProt: The universal protein knowledgebase in 2021," *Nucleic Acids Res.*, vol. 49, no. D1, pp. D480–D489, Jan. 2021.

[88] W. R. Pearson, "An introduction to sequence similarity ('homology') searching," *Curr. Protoc. Bioinforma.*, vol. 0 3, no. SUPPL.42, 2013.

[89] A. A. Kermani, "A guide to membrane protein X-ray crystallography," *FEBS Journal*. Blackwell Publishing Ltd, 2020.

[90] A. Sailapathi, S. Gunalan, K. Somarathinam, G. Kothandan, and D. Kumar, "Importance of Homology Modeling for Predicting the Structures of GPCRs," in *Homology Molecular Modeling - Perspectives and Applications*, IntechOpen, 2021.

[91] R. Lüthy, J. U. Bowie, and D. Eisenberg, "Assessment of protein models with three-dimensional profiles," *Nature*, vol. 356, no. 6364, pp. 83–85, 1992.

[92] H. J. C. BERENDSEN, W. F. VAN GUNSTEREN, H. R. J. ZWINDERMAN, and

R. G. GEURTSEN, “Simulations of Proteins in Water,” *Ann. N. Y. Acad. Sci.*, vol. 482, no. 1, pp. 269–286, Dec. 1986.

[93] W. Schreiner, R. Karch, B. Knapp, and N. Ilieva, “Relaxation estimation of RMSD in molecular dynamics immunosimulations,” *Comput. Math. Methods Med.*, vol. 2012, 2012.

[94] D. McIlwain, P. Lang, T. Maretzky, … K. H.-, and undefined 2012, “iRhom2 regulation of TACE controls TNF-mediated protection against Listeria and responses to LPS,” *science.sciencemag.org*.

[95] R. satyaveni Geesala *et al.*, “Loss of RHBDF2 results in an early-onset spontaneous murine colitis Pathogenesis of Cystic Fibrosis View project Anti Breast cancer agents View project Loss of RHBDF2 results in an early-onset spontaneous murine colitis,” *Artic. J. Leukoc. Biol.*, vol. 105, no. 4, pp. 767–781, Apr. 2019.