

**Molecular Modeling Strategies for the
Therapeutic Intervention of Spinal Muscular
Atrophy with Progressive Myoclonic Epilepsy
(SMA-PME)**



By

Syeda Aniqah Bukhari

Master of Sciences in Bioinformatics

Fall 19-MSBI-NUST00000319238

Supervised by:

Dr. Ishrat Jabeen

Research Center for Modeling and Simulation (RCMS)

National University of Science and Technology (NUST)

Islamabad, Pakistan

November 2021

Molecular Modeling Strategies for the Therapeutic Intervention of Spinal Muscular Atrophy with Progressive Myoclonic Epilepsy (SMA-PME)

A thesis submitted in partial fulfilment of the requirement for the degree of
Master's in Bioinformatics.



By

Syeda Aniqah Bukhari

Master of Sciences in Bioinformatics

Fall 19-MSBI-NUST00000319238

Supervised by:

Dr. Ishrat Jabeen

Research Center for Modeling and Simulation (RCMS)

National University of Science and Technology (NUST)

Islamabad, Pakistan

November 2021

Certificate of Originality

I hereby declare that the research work presented in this thesis has been generated by me as a result of my own research work. Moreover, none of its contents are plagiarized or submitted for any kind of assessment or higher degree. I have acknowledged and referenced all the main sources of help in this work.

Syeda Aniqah Bukhari

Fall 2019-MS BI-4 00000319238

Dedication

I dedicate this work to my beloved Parents for their prayers, love, support, and most of all their huge believe in me and my dreams.

Thanks For enlightening me and making me who I am.

Acknowledgment

[Glory be to You; we have no knowledge except what you have taught us. Verily, it is You, the All-Knower, the All-Wise. (Surah Baqarah-2:32)]

Allah-there is no deity except Him, the Ever-Living, the Sustainer of [all] existence. We pay our all gratitude to Allah, the Almighty, from whom we seek help, and, in whose premises, there are all lives and matters. We are obligated to Him for all knowledge He has provided us in the completion of this project.

My sincere respect goes to my supervisor Dr. Ishrat Jabeen whose guidance, motivation and believe in me throughout this journey. She is a true leader and a great mind of her time. It was an honor to learn and work under her supervision. I am deeply indebted for her precious advice, time, efficient contribution, and supervision during my work. I could not have imagined having a better advisor and mentor for my study. I would also like to thank my guidance committee members: Dr. Zamir Hussain and Dr. Muhammad Tariq Saeed from RCMS, NUST for their encouragement and insightful comments. I would like to thank all the faculty members and lab assistants at RCMS, NUST for their guidance and support.

I am indebted to my father Syed Ashfaq Hussain Bukhari and my mother Syeda Naghma Ashfaq Bukhari from being the best parents, teacher and friend to me. I have no suitable word that can fully describe their everlasting love and care to me. I owe this and every success of my life to them and will be in debt for their hard work to make my dreams come true. I am really pleased to pay my special affections to my brothers Syed Omer

Bukhari and Syed Rauf ul Hassan Bukhari for their wisdom, continuous support, and prayers throughout my studies. They are my biggest supporters and without their motivations I wouldn't have the courage to stand in this world. I am proud of them individually for their journeys and I owe my sincere gratitude to them.

Last, but not the least, special thanks to my friend Ammara Naz, Rijja Hossain, Aiman Rauf, Maham Ahmad and Mehrosh Agha for creating such positive and learning environment in the lab. Their insightful ideas and discussions help me a lot during my research phase, their moral support made me to come up from the crises during this project. They have such a beautiful mind and heart and I adore each one of them. I am also thankful to my colleagues Mariya Ehsan, Fatima Iqbal, Aqsa Khalid, Humaira Ismatullah, and Sadaf Ejaz for the scientific discussions and guiding me in every possible way. Finally, I want to thank all those wished this memoir to reach its successful completion.

Table of Contents

List of Abbreviations	i
List of Tables	iii
List of Figures.....	iv
Abstract.....	iv
Chapter 1: INTRODUCTION.....	1
1 Spinal Muscular Atrophy with Progressive Myoclonic Epilepsy (SMA-PME).....	2
1.1 Reason for SMA-PME	2
1.2 Functions of Acid Ceramidase (aCDase)	3
1.3 Acid Ceramidase (aCDase) Dysfunctionality	4
1.4 Our Proposed Strategy	5
1.5 Objectives.....	5
Chapter 2: LITERATURE REVIEW.....	6
2.1 Evolution of SMA-PME.....	7
2.2 Signaling Pathway of ASAH1 Gene Expression.....	9
2.3 Ongoing Treatments and Research	11
2.4 Structural Analysis of aCDase	15
2.5 Structural Analysis of CREB	16
Chapter 3: METHODOLOGY.....	20
3.1 MODULE I (aCDase Structure-Based Methodology)	21
3.1.1 Data Collection	23
3.1.2 Structural Preparation of Sequences	23
3.1.3 Constraint Docking.....	23
3.1.4 De novo Protein Modeling.....	23
3.1.5 Molecular Dynamics Simulations	23
3.2 MODULE II (CREB Structure-Based Methodology)	25
3.2.1 Data Collection and Structural Preparation	25
3.2.2 Molecular Docking.....	29
3.2.3 Molecular Dynamics Simulations	30
3.2.4 Ligand-based Pharmacophore Analysis	31
3.2.5 Virtual Screening	33
Chapter 4: RESULTS.....	33
4.1 MODULE I (aCDase Structure-Based)	36
4.1.1 Structural preparation of sequence.....	36
4.1.2 De Novo Modeling.....	38

4.1.3	Molecular Dynamics Simulations (MDS).....	42
4.2	MODULE II (CREB Structure-Based).....	48
4.2.1	Molecular Docking Analysis	48
4.2.2	Molecular Dynamics Simulations	50
4.2.3	Ligand Based Pharmacophore.....	53
4.2.4	Virtual Screening	56
4.2.5	Biological Evaluation	56
Chapter 5:	DISCUSSION.....	64
Chapter 6:	CONCLUSION.....	67
Chapter 7:	REFERENCES	68
Chapter 8:	APPENDIX.....	75

List of Abbreviations

SMA-PME	Spinal Muscular Atrophy with Progressive Myoclonic Epilepsy
LSD	Lysosomal Storage Disorders
aCDase	Acid Ceramidase
SNP	Single Nucleotide Polymorphism
FD	Farber Disease
ORF	Open Reading Frame
CHO	Chinese Hamster Ovary
HSCT	Hematopoietic Stem Cell Transplant
BRM	Biological Regulatory Mechanistic
AD	Adenyl Cyclase
cAMP	Cyclic AMP
CREB	cAMP Element Binding Protein
CBP	CREB Binding Protein
CRTC	CREB Regulated Transcriptional Cofactor
SK1	Sphingosine Kinase 1 Phosphorylates
P13K	Phosphoinositide 3-Kinases
KLF-6	Kupple Like Factor 6
rh-aCDase	Recombinant Human aCDase
μPLI	Recombinant Micro Plasmin
μPLG	Human Micro Plasminogen
PICP	Procollagen Type I Carboxy-Terminal Propertied
NTN	N-Terminal Nucleophile Superfamily
NAAA	N-Acyletanolamine Acid Amidase
SP	Single Peptide
CRE	CREB Responsive Element
CAD	Constitutive Active Domain
PKA	Protein Kinase A
TAD	Transcriptional Activity Domain
BZIP	Dna Binding Domain
r-aCDase	Recombinant Acid Ceramidase
SUMO	Small Ubiquitin-Like Modifier

GSTs	Glutathione S-Transferases
MDS	Molecular Dynamics Simulation
RMSD	Root Mean Square Deviation
RMSF	Residue
KID	Kinase Inducible Domain
SMILE	Simplified Molecular-Input Line-Entry System

List of Tables

Table 3.1	Inducers of CREB and their activities.....	27
Table 4.1	Statistical parameters of the modelled protein before and after energy minimization (forcefield Amber99). The ids of constructed models remain similar to their respective sequences for analysis.....	39
Table 4.2	Statistical parameters of Ramachandran plot before and after the 50 nsec of MD simulations.....	46
Table 4.3	Crucial pharmacophore features of the finally selected model with respective radiuses and mutual distances.....	54
Table 4.4	Statistical evaluation of the pharmacophore model w.r.t internal dataset.....	56
Table 4.5	2D structures of hits obtained after virtual screening by Drug Bank database for repurposing against CREB protein.....	58
Table 4.6	2D structures of hits after =virtual screening by ZINC database with their natural origin.....	61
Table 4.7	2D structures of hits after virtual screening by FDA-approved EPI/SMA drug database for repurposing against CREB protein.....	62

List of Figures

Figure 1.1	Hydrolysis of ceramide by aCDase4
Figure 2.1	BRM of aCDase. (a) ASAH1 gene expression via cAMP/CREB mediated pathway. The green color box highlighted the CBP/CREB-CRTC complex, a crucial transcription factor for ASAH1 synthesis (b)ERK1/2 mediated ASAH1 gene expression. ER-a is responsible for activating CREB for ASAH1 gene expression (c) c-Met mediated KLF6 signaling; the transcription factor SP1 makes a complex with KLF6 and undergoes a pro- aCDase precursor. The precursor translocate towards cytoplasm and converts into functional aCDase.....10
Figure 2.2	Conversion of pro Acid ceramidase (aCDase) to Functional aCDase; Primary structure of aCDase exhibits α subunit and β subunit. The Cys-143 residue of the β subunit is responsible for the hydrolysis of ceramides.....16
Figure 2.3	Primary Structure of CREB: Q1 and Q2 domains are responsible for glycosylation. KID domain activates CREB by kinases like CBP/P300 at SER133 residue while bZIP domain makes a complex with CRTC and undergoes the transcriptional process.....18
Figure 3.1	Overall Workflow of Module I: r-aCDase,.....21
Figure 3.2	Overall workflow of Module II: CREB based.....25
Figure 3.3	Workflow of Molecular Docking. The biological data was retrieved from Protein Databank under the pdb id 2lxt. However, the chemical data of 13 ligands were collected from ChEMBL, PubChem and Literature. Then molecular docking was performed by GOLD software where binding pocket was identified which was the interphase of CREB and CBP. By taking the coordinates -3.0045 (X), 9.7060 (Y), and -4.9740 (Z) docking was performed. After that, the pose analysis and selection were made for correlation analysis and binding hypothesis formation.....29
Figure 3.4	Overall workflow of Molecular Dynamics Simulations of selected highly active ligand/CREB complex. This process was done by removing CBP from the docked complex and processed towards system building. After energy minimization MD production for 400nsec was made and the results were analyzed by means of trajectory analysis.....31

Figure 3.5	Overall workflow of Ligand based Pharmacophore Model Generation using stable conformer from MD simulation studies. The step was done by extracting the feature from the template and generation of confirmation from rest of the docked data. Based on features, model was generated and screened against the packed confirmation for internal set validation.....	32
Figure 3.6	Different steps of virtual screening; from ADMET toxicity to Hits identification. Number of compounds in blue color (ZINC database), pink color (DRUGBANK database) and orange (EPI/SMA database) color at the end of each step declares no of hits of the respected validation step.....	34
Figure 4.1	Sequence alignment of GST tag's sequence and 1DUG sequence. The red highlighted region indicates the perfect match between both sequences.....	36
Figure 4.2	(a) Ramachandran plot of stable GST model. About 149 residues are in the allowed region, with three outliers in the disallowed region. (b) Stabled GST tag structure after homology modeling. (c) RMSD vs. time plot after MD simulations suggested the stability of the protein's backbone after 50nsec.....	37
Figure 4.3	Ramachandran plot of Seq1-7 proteins (a-g) before and after energy minimization. The number of outliers (residues in the disallowed region) significantly decreased after energy minimization, resulting in refined 3D structure.....	41
Figure 4.4	RMSD plots for modeled proteins of seq1 to seq7 (a-g), respectively. The plots showed backbone convergence (RMSD) on the y-axis and simulation time steps (nsec) on the x-axis. Based on the least convergence and longer stability time, seq4 and seq7 proteins showed the most stable state.....	43
Figure 4.5	RMSF plots for modeled proteins of seq1 to seq7 (a-g), respectively. The plots showed fluctuation in residues (RMSF) on the y-axis and residue index on the x-axis. The region with a loop region indicating the most fluctuated behavior.....	45
Figure 4.6	Ramachandran plot of the selected recombinant proteins (left) and their 3D structures (right). (a) seq3 Ramachandran plot and stable modeled structure. The red color structure represents the aCDase, while yellow declared GST tag. (b) The ramachandran plot of seq4 with its stable modeled structure (red color shows	

	aCDase domain while purple color represents Histidine + SUMO (small ubiquitin modifier like protein). (c) The ramachandran statistics of seq7 with its stable structure. Blue structure represents Melittin peptide tag while red shows aCDase hydrolytic domain.....	48
Figure 4.7	All docked ligands (yellow) at the interphase of CBP (purple) and CREB (blue). CBP wrap at a 90° along with CREB for its structural stability.....	49
Figure 4.8	Correlation plot between pAC50 values and Gold fitness score (0.29). The Blue marked triangle indicated the highly active compound (0.1nM) with a pAC50 value of 10 and GOLD score of 38.59.....	50
Figure 4.9	Molecular dynamics simulation of the highly active compound (LIG_1). The graph shows the CoRMSD on the y-axis during the time frame of 400nsec on the x-axis. The simulation remained inconsistent for the first 200nsec and showed a stable behavior for the last 100nsec, suggesting the stability of the model.....	51
Figure 4.10	Protein-Ligand interactions of LIG_1/CREB complex. The y-axis shows the interaction fraction, while residues of CREB demonstrate on x-axis. The residue SEP133 showed most stable H-bond interaction throughout the simulation trajectory. While ARG131, TYR134, and ILE137 demonstrated hydrophobic interactions for 20-30% of the simulation time.....	52
Figure 4.11	Interacting residues pattern before and after MD. The brown-colored structure represents the ligand molecule. The sky blue color structure shows the CREB resides. The dotted blue colored line represents the H-bond interactions, while maroon dotted lines represent the hydrophobic interactions between the ligand and protein. Before MD, the binding residues were SEP121, ARG125 and ARG135. However, the ligand translocated its position to ARG131, SEP133, TYR134 and ILE137 after MD	53
Figure 4.12	Combination of 14 feature sites found in the template during pharm_LIG_1/CREB pharmacophore model generation	54
Figure 4.13	In the pharm_LIG_1/CREB pharmacophore model, the blue-colored sphere represents the Hydrogen bond acceptor while green colored sphere represents the hydrophobic features	55

Abstract

SMA-PME has been recognized as a rare autosomal recessive neuromuscular disorder. It is depicted as a childhood onset of proximal muscular atrophy progressive with myoclonic seizures. This disorder is caused by 13 SNP mutations in the *ASAH1* gene, which encodes acid ceramidase (aCDase) proenzyme. After activation aCDase hydrolyzes lysosomal ceramide into sphingosine and fatty acid, thus maintains lipid's cellular homeostasis. As a result of mutations, transcription of premature proenzyme occurs, which leads to activity loss of aCDase up to the one-third level. Therefore, due to enzyme insufficiency, there is ceramide accumulation and neural cell degradation such as SMA-PME. Until now, no cure for aCDase deficiency has been recognized, and the standard of treatment focuses on symptom management. However, recent proof of study on structural activation of aCDase has proposed as a source of recombinant therapy of Farber Disease. It shows that the catalytic domain of the aCDase can perform work when combined with some potent peptide (tag). Moreover, research on the *ASAH1* gene with biological regulatory pathways investigations highlights the role of transcription factors (CREB) in inducing gene expression.

In present study, two therapeutic options have been explored: 1. r-aCDase modeling- by optimizing the stability between the binding domain of aCDase and a tag peptide for. 2. *ASAH1* driven signaling pathway to induce the expression of TF (CREB) to enhance the production of aCDase. This research resulted into the three stable r-aCDase structures that can be used to improve the transport and efficacy of aCDase. Furthermore, it was also identified that most significant residue for CREB induction was ARG131, SEP133, TYR134 and ILE137. Most probable 3D binding conformation of the template after MD simulation was used for 3D predictive modeling. The built pharmacophore model exhibited 92% accuracy with two H-bond acceptors and two hydrophobic features. Lastly, an integrated virtual screening pipeline was generated to predict 19 FDA approved drugs as potential hits against CREB to induce *ASAH1* gene expression. Thus, the proposed hits can be utilized in future for further experimental validation.

Keywords: *ASAH1* gene; aCDase deficiency, Recombinant Replacement Therapy; De novo modeling CREB; Molecular Docking; Virtual screening.

CHAPTER 1

INTRODUCTION

1 Spinal Muscular Atrophy with Progressive Myoclonic Epilepsy (SMA-PME)

Spinal muscular atrophy with progressive myoclonic epilepsy or SMA-PME is considered a rare autosomal recessive syndrome [1]. It is a neuromuscular disorder characterized by childhood onset of proximal muscle weakness and atrophy, followed by myoclonic seizures. Further symptoms include progressive epilepsy, tremor, and sensorineural hearing disability [2]. The neuronal-muscular damage eventually leads to severe respiratory problems and the patient's death [3]. SMA-PME is an autosomal recessive disease where both parents transmit the disease [4].

Jankovic Rivera first described SMA-PME as the Jankovic-Rivera syndrome [5]. Mutations in the *ASAH1* gene are accountable for this disorder. SMA-PME has 24 confirmed cases associated with this gene [6]. Therefore, as a distinct clinical entity, SMA-PME is a part of the lysosomal storage disorders (LSD) distinguished by acid ceramidase (aCDase) deficiency [7].

1.1 Reason for SMA-PME

SMA-PME is triggered by mutations in the *ASAH1* gene that encode the acid ceramidase (aCDase) pro-enzyme. *ASAH1*, or the Human ceramidase gene, is located at chromosome 8's (8p21.3/22) short arm. It is 30 kb in length, containing 14 exons that range from 46 to 1200 bps long [8]. There are about 13 reported mutations (P26R, T42A, T42M, Y59*, V75A, Y137C, K125N, N173I, T179I, F199, G284X, R296X) that have the potency to cause SMA-PME [9]. These mutations are single nucleotide polymorphism in nature, i.e., a single change of amino acid can cause protein dysfunctionality [10]. As a result of these SNP, transcription of premature pro-enzyme occurs, which ultimately destabilizes the protein fold. This incident consequently leads to activity loss of aCDase up to the one-third level. Therefore, due to enzyme deficiency, ceramide accumulation occurs, causing neural cell degradation such as SMA-PME [11].

SMA-PME children, after some years of mind and body development, start experiencing pain in the lower limb, causing muscular distress and trouble in walking. The atrophy then spreads to other body muscles and causes difficulty in breathing.

Due to weakness, the body gets highly susceptible to pneumonia and other respiratory diseases. When the muscle wasting start, the patients experience seizures due to a high accumulation of ceramide in the neuron [3].

1.2 Functions of Acid Ceramidase (aCDase)

Acid ceramidase or N-acylsphingosine deacylase (EC 3.5.1.23) is a member of the hydrolase family, responsible for maintaining lipid metabolism inside the body [12]. aCDase is ubiquitously translated and present in lysosomes. Through the intra-lysosomal membrane, it regulates lipid breakdown [13]. Under normal circumstances, ASAH1 gene transcription is carried and produces inactive aCDase pro-enzyme. Then, this pro-enzyme gets activated by auto cleavage process at Cys-143 residue (Arg-415/Asp162 in assistance[13]), rendering a functional heterodimeric aCDase. This mature structure is comprised of α - subunit (13 kDa) and β -subunit (30 kDa) [14]. Functional aCDase utilized the same auto-cleaved Cys-143 at β -subunit for hydrolysis of ceramides [15]. After the post-transcriptional modification of the heterodimer, aCDase hydrolyzes lysosomal ceramide into sphingosine and free fatty acid as shown in figure 1.1. This biochemical reaction is carried out in the cytosol of the lysosome, where ATP is generated to do the fat metabolism [16]. An additional layer of information reveals that aCDase has another 'helper' protein, saposin-D, for its optimal activity. Saposins are small lysosomal proteins that are expected to be participating in membrane disruption for ceramide access to aCDase [14].

Human aCDase is an intracellular cysteine amidase responsible for maintaining cellular homeostasis [17], and the breakdown of ceramide is considered the last step of lysosomal sphingosine breakdown. Therefore, the disruption in this process affects many other metabolic activities like the sphingomyelin pathway, de novo biosynthesis pathway, glycolipids pathway, and sphingosine-1-phosphate pathway [18].

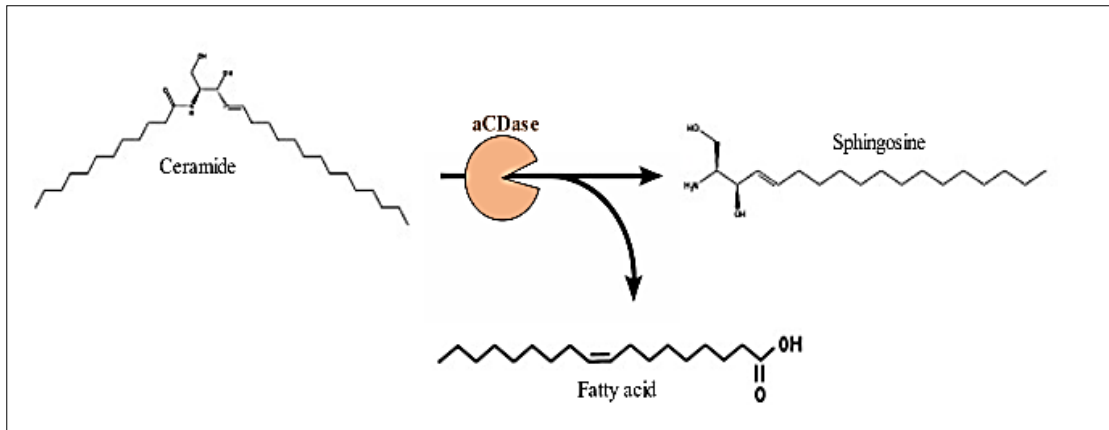


Figure 1.1: Hydrolysis of ceramide by aCDase.

1.3 Acid Ceramidase (aCDase) Dysfunctionality

aCDase works as the housekeeping enzyme. Increased expression of aCDase is liable for certain cancers, type-2 diabetes, and Alzheimer's disease. However, some mutations are accountable for lowering the level of aCDase inside the body. The decline in aCDase causes two severe Lipid Storage Disease (LSD); Farber disease (FD) and Spinal muscular atrophy with progressive myoclonic epilepsy (SMA-PME). FD triggered by mutations of the *ASAH1* gene, which are discrete from SMA-PME. The patient experiences joint pain, skin nodules formation, hoarseness, and life-threatening lung damage in this disease. However, SMA-PME discriminates this clinical entity in muscle wasting and seizures [6].

In SMA-PME, mutations in the *ASAH1* gene concededly result in premature aCDase [11]. There are about 13 SNP mutations responsible for aCDase dysfunctionality. The most crucial mutation is T42M, which is found among 12 SMA-PME patients' clinical reports. The premature aCDase lost its ability to convert into a mature functional heterodimer, thus unable to hydrolyze the cell's ceramide. The consequences are deficiency of aCDase inside the body and high ceramide level to the point where it causes neural cell degradation [19]. By considering the cited problem, the current research project has been designed to explore two therapeutic options, either as a recombinant enzyme product or induction of *ASAH1* gene expression to sustain or upregulate body aCDase level.

1.4 Our Proposed Strategy

As a result of mutations in ASAHI, transcription of premature pro-enzyme occurs, which ultimately destabilizes the protein fold. This event, therefore, leads to activity loss of aCDase up to the one-third level. Thus, as a result of enzyme inadequacy, there is ceramide accumulation which causes neural cell degradation. One of the consequences of such a drastic event is SMA-PME. Hence, we intend to introduce two therapeutic targeting options. One is r-aCDase models to check the tag's stability and effectiveness. While another one is to target the ASAHI driven signaling pathway- to induce TF (CREB) expression, enhancing the production of aCDase. The purpose of both targets will be upregulation of aCDase level inside the body as therapy.

1.5 Objectives

1. To develop 3D structural models of sequences of r-aCDase.
2. To develop and analyze tag's structural stability for effective transport of aCDase.
3. To modulate normal physiological function of CREB by.
 - i. Establishing the binding hypothesis of CREB.
 - ii. Developing the 3D predictive model for the prediction of the SMA-PME potential modulators.

CHAPTER 2
LITERATURE REVIEW

2.1 Evolution of SMA-PME

Jankovic-Rivera first described spinal muscular atrophy with progressive myoclonic epilepsy or SMA-PME in 1979 as Jankovic-Rivera syndrome. They investigated 3 Texan families and observed an indiscriminate early-onset sensitive myoclonus pattern, leading deliberately to progressive distal muscle weakness and wasting. They reported that a syndrome similar to Jankovic-Rivera syndrome had not been reported previously [20]. In 1989, Moser and his co-workers [21] carried out the metabolic analysis of inherited diseases and identified a lipid storage disease with characteristics of aCDase deficiency. In 1995, acid ceramidase purification and characterization were completed by Bernardo [22]. The isolation of aCDase from the urine was achieved via chromatography. The characteristics analysis of aCDase showed its importance in ceramide metabolism. Apart from this, Koch established a team [23] that isolated cDNA of aCDase from pituitary tissue and skin fibroblast culture. The cDNA sequence was found to be composed of 2312 bps, where 1185 bps of ORF (open reading frame) encoded for 395 amino acid polypeptides. The results declared that human aCDase shares approximately 90% of similarity with murine aCDase. This study was further utilized by Li and his co-workers in 1999 [24] for cloning the human *ASAH1* (h-*ASAH1*) gene. The characterization of the h-*ASAH1* gene revealed a 30kb length gene with 14 exons (46 to 1201bp). The promoter region comprised of GC rich content having activity at 475 bp. The study also detected a minor quantity of aCDase in heart and skeleton muscle.

In 2003, purification and characterization of recombinant aCDase were achieved [25]. The study was based on the human aCDase quantification in CHO cell lines (Chinese hamster ovary). The results suggested that the primary structural model of r-aCDase contained the same domain as humans. However, the de-glycosylation investigation mentioned in the same research implied that enzymes are enriched in mannose grouped oligosaccharides, which mainly contribute to the hydrolysis process. Till 2011, Jankovic-Rivera syndrome came under the umbrella of LSD. In 2012, Zhou [26] identified some mutations different from FD disease and declared this disease as Spinal Muscular Atrophy with Progressive Myoclonic Epilepsy (SMA-PME). Previously any case of SMA-PME was linked to the *SMN1* gene as a subtype

of SMA disease. Therefore, by SNP microarray analysis of the *ASAH1* gene, SMA-PME was affirmed as a disorder apart from any subtype syndrome. In 2014, Dymont and his team analyzed the genetic, clinical and biochemical mechanism of SMA-PME [3]. The different heterogeneous variants of aCDase were identified via exome sequencing analysis. In this study, a novel case of SMA-PME with distinguished motor neuron damage was reported. Because of this, the research highlighted the phenotypic effect of *ASAH1* mutation causing SMA-PME. In 2016, for the first time, Filosto depicted the association of the *ASAH1* gene with other types of SMA-PME [27]. Thus, expanding the spectrum of diseases caused by *ASAH1* mutations. This study is considered as part of diagnostic testing for non-5q SMA patients.

Fabian made a breakthrough in 2018 [28] by publishing different wet-lab studies for the therapeutic intervention of aCDase deficiency. Here, the authors highlighted Hematopoietic Stemcell Transplant (HSCT) utilization among FD patients to increase muscle mobility. In addition to this, HSCT with ex-vivo gene therapy improved the neurological symptoms. However, enzyme replacement therapy is standard for many LSD treatments where a recombinant enzyme can improve the deficiency of aCDase with better efficacy and transport properties. Though, the therapy is still lacking clinical trials for SMA-PME due to certain limitations. Recently, Van der Beek identified a biallelic variant of the *ASAH1* gene expanding the biochemical and phenotypic spectrum of SMA-PME [29]. This variant of *ASAH1* still has the same phenotype as SMA but didn't exhibit any PME symptoms. Based on the previous study, Mahmoud tested a new biomarker, C26-Ceramide, to diagnose SMA-PME [29]. The evaluated genetic spectrum of the Egyptian family declared three different variants, including c.1126A>G (p.Thr376Ala), exon-5 deletion, and c.1205G>A (p.Arg402Gln). However, a patient with c.1205G>A (p.Arg402Gln) variant exhibited symptoms of PME without seizures and epilepsy. In the end, a new term is suggested by Radhakrishnan, "SMA-plus" for the SMA-PME patient that exhibited phenotype associated with the non-SMN gene. SMN gene, which is responsible for the SMA phenotype, was distinguished and characterized separately. The author commented that though the genetics studies of SMA-PME have evolved, there is still a need for interventions for proper clinical settings for the treatment of SMA-PME [6].

2.2 Signaling Pathway of ASAH1 Gene Expression

In normal conditions, ASAH1 gene transcription is carried out by series of signaling pathways that induce aCDase production inside the cell. Therefore, biological regulatory mechanistic (BRM) reveals many transcription factors responsible for ASAH1 biosynthesis. The role of these transcription factors is to maintain the level of aCDase in the lysosome for functional ceramide metabolism.

cAMP/CREB mediated gene expression: At first, on the binding of neurotransmitter (ACTH), MC2R based GPCR- receptor activated, allowing the cascading of cAMP/PKA pathway [30]. The α subunit of the GPCR receptor is responsible for activating adenylyl cyclase (AD), which allows the conversion of ATP to cyclic AMP (cAMP). cAMP is accountable for the activation of PKA subunit. PKA phosphorylates CREB (cAMP Element Binding Protein) at Ser133 position converting it into SEP133. This event promotes CBP (CREB Binding Protein) and its paralogue p300 recruitment towards CREB at this residual position to carry out gene expression [31]. Attachment of CBP/p300 is crucial for the CREB to remain activated during the metabolic process. CREB-mediated ASAH1 gene expression is further stabilized by another transcriptional factor, CRTC (CREB regulated transcriptional cofactor). CRTC binds to the bZip domain of CREB, assessing CREB to the ASAH1 DNA during gene expression. The CBP/CREB/CRTC complex formation is vital for carrying out CREB's transcriptional activity [32], i.e., promotes ASAH1 gene expression by binding to multiple regions of the ASAH1 promoter [33].

ERK1/2 mediated gene expression: The β subunit of the GPCR receptor is responsible for ASAH1 gene expression via RAF-1 dependent ERK signaling. Here, downstream signaling of kinase proteins including c-SRC, RAS, RAF-1, MEK, and ERK1/2 occurs via phosphorylation. ERK1/2 activation induces Sphingosine kinase 1 (SK1) activity and promotes a rapid secretion of Transcription factor Sp1(SP1) in an ER α -dependent manner. Phosphorylation of ER α and its recruitment to the ASAH1 promoter results in acid ceramidase synthesis. ER α also phosphorylates CREB protein, hence increasing ASAH1 transcription [7]. Here CREB proves to be a crucial transcriptional factor in the upregulation of precursor of aCDase pro-enzyme.

KLF6 mediated gene expression: c-Met activation is responsible for the upregulation of PI3K (Phosphoinositide 3-kinases) inside the cell. PI3K activates AKT protein which is an essential second messenger. AKT, through various pathways, promotes the regulation of KLF-6 protein (Kupple like factor 6) [34]. This protein acts as a transcriptional factor of ASAH1 gene by combining with SP1 protein from the ERK1/2 pathway. Both protein, after complex formation, binds to the high GC content of the DNA to translate acid ceramidase, facilitating sphingosine degradation [35].

The end product of these regulatory pathways is a precursor of the pro-enzyme of aCDase. aCDase pro-enzyme undergoes post-translational modification for functional aCDase heterodimer as shown in figure 2.1 [13, 36]. This heterodimer regulates sphingolipid metabolism after ubiquitous expression in the lysosome [37].

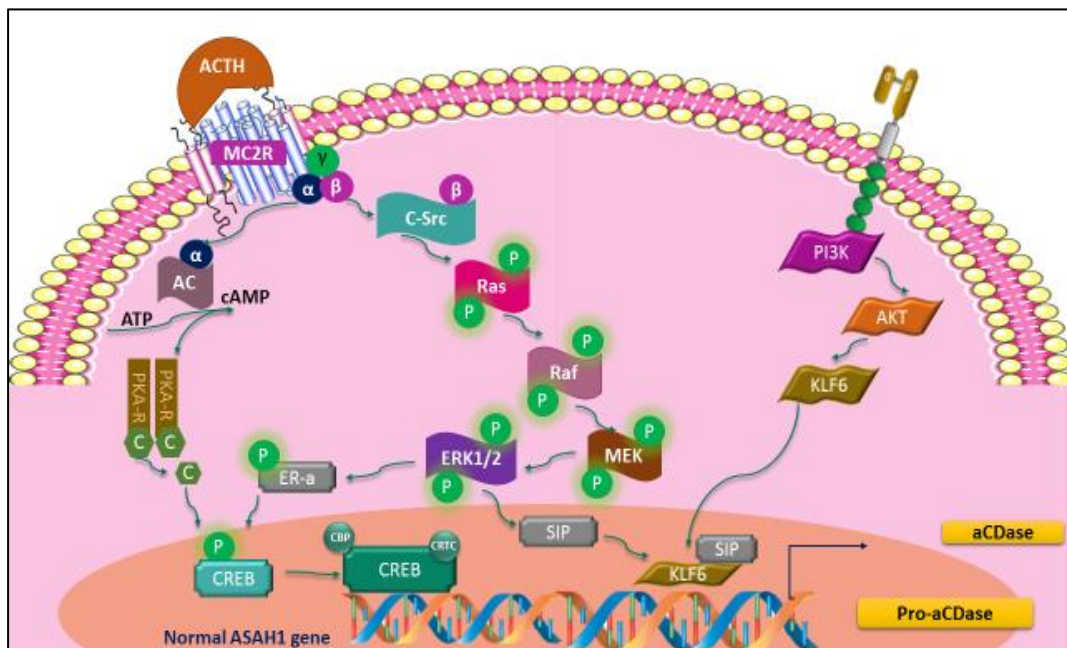


Figure 2.1: BRM of aCDase. (a) ASAH1 gene expression via cAMP/CREB mediated pathway. The green color box highlighted the CBP/CREB-CRTC complex, a crucial transcription factor for ASAH1 synthesis (b)ERK1/2 mediated ASAH1 gene expression. ER-a is responsible for activating CREB to do ASAH1 gene expression (c) c-Met mediated KLF6 signaling; the transcription factor SP1 makes a complex with KLF6 and undergoes a pro- aCDase precursor.

The precursor translocates towards cytoplasm and converts into functional aCDase.

**AC: Adenyl cyclase, PKA: protein kinase A, CREB: cyclic AMP-responsive element-binding protein (CREB), SP1: Transcription factor Sp1, KLF-6 KLF-6 protein (Kupple like factor 6)*

2.3 Ongoing Treatments and Research

Currently, no cure for SMA-PME has been reported. Therefore, various strategies were adopted for symptom management/treatment [38]. For example, anti-inflammatory medications with physiotherapy can help in solving muscle pain and mobility issues. On the other hand, anti-epileptic drugs prove beneficial for controlling seizures and tremors for a specific time. Furthermore, for patients with a respiratory disorder or pneumonia, ventilators can be consumed. In addition, a power motor wheelchair can help address the walking problem of the patient [39]. The main reason for this type of intervention is the poor diagnosis of disease in clinical practice [3].

2.3.1 *In silico* Research

Many *in silico* studies on diagnosing the disease have been reported, considering the complexity of the disease. However, due to limited understanding of disease mechanisms, bioinformatic approaches have been restricted for SMA-PME. Though the mutation identification part of the disease problem is very well understood now. Whole Exome Sequencing (WES) is an effective approach to identify mutations at the transcription level. On the other hand, many mRNA splicing-based mutations get lost in clinical settings due to limited data. So Kernohan in 2017 [40] demonstrated RNA sequencing from the patient's blood to recognize disease-causing mutations for non-hematological conditions. Here they performed transcriptome sequencing on the patient's leukocyte followed by Sanger sequencing. The results stated that SMA-PME patient has biallelic type ASAH1 mutations (c.504A>C (p.Lys168Asn), c.458A>G (p.Tyr153Cys)). This study provided a diagnosis for unsolved patients.

In 2019, mutation identification in the ASAHI gene was investigated by Rajput and his co-workers [41]. In this study, detailed analysis of mutations causing FD and SMA-PME was highlighted. Prediction of amino acid substitution stated that 16 mutations (Y36C, E138V, V97E, L182V, T222K, G235D, R226P, R254G, G235R, N320S, R333G, D331N, R333H, P362R, W185R, and G169R) were identified as more deleterious to cause the disease. Both FD and SMA-PME get distinguished based on silent mutations and symptom occurrence. On the other hand, structural analysis of aCDase is considered one of the milestones aiding future therapies. In 2018, research on aCDase and SMA-PME highlighted the mechanism of the disease based on its structural association. Molecular docking and molecular simulation studies have shown the importance of Cys-143 residue converting pro-aCDase to functional aCDase. β subunit of the enzyme is responsible for its hydrolase activity while α subunit is crucial for the stability of the whole complex.

2.3.2 In-vitro/ In-vivo Experimentations

On the optimistic side, biological assay studies show an extensive analysis for the treatment of this disease. These studies are enriched in sequencing the ASAHI gene, mutation identification, diagnosis in the clinical setting, and therapies for aCDase deficiency. However, the therapeutic interventions were performed for other LSD but can be based on the treatment of SMA-PME. For example, for understanding SMA-PME and molecular assessment of ASAHI gene, the characterization of SNP sites inside the gene was investigated. For this purpose, characterization and identification of splicing exon elements and perturbing mutant sites were identified. Bashyam cloned the ASAHI gene and investigated the missense mutation having the ability to cause the disorder. This study provides the diagnostic basis for diseases related to the ASAHI gene [42].

Talking about therapeutic assessment for LSD, the work done on gene therapy for FD disease involved inducing aCDase to reduce the ceramide accumulation in the target tissues. aCDase mediated gene therapy in a knock-out model of mouse showed positive results. Lentiviral mediated neonatal aCDase delivery was made for assessing the sphingolipids' metabolism. In addition, the model organism showed enhanced enzyme activity. However, due to the over-expression of the enzyme in the targeted tissue, this approach highlights cancer's endangerment. Thus, the utility of this

therapy is limited for SMA-PME patients [43]. The second therapeutic option was Hematopoietic stem cell transplantation (HSCT). For this purpose, a multipotent HSCT study was carried out in other LSD where cells get implanted in bone marrow or peripheral blood to improve mobility inside the body [44]. The hurdle of this treatment to be effective for SMA-PME patients is that it only works for individuals who do not have significant neurologic involvement. Nevertheless, this therapy has been doing well in clinical trials for FD disease.

Along with this, another study reveals that ex vivo gene therapy monitored by transplant characterizes a development over HSCT alone [45]. Thus, it may deliver a longer-lasting beneficial benefit than customary HSCT [14]. In addition to their endogenous gene manifestation, the transduced cells prompt enzyme derived from the therapeutic trajectory, which, in concept, increased enzyme manufacture, lysosomal action, and prospective cross-correction [46]. Although this investigation is for an altered type of SMA, it is conceivable that a related gene therapy method may also be promising for patients with the SMA-PME phenotypes [8].

Although acknowledged for the pre-clinical trials, enzyme replacement therapy might be the future cure of SMA-PME. According to molecular analysis of aCDase, targeting the enzyme's catalytic site is vital for recombinant replacement therapy can be made [47]. Enzyme replacement therapy or ERT with recombinant human aCDase (rh-aCDase) is currently improving and characterizes a promising treatment for aCDase deficiency in many LSDs [8]. In the past, similar studies on different enzymes have been executed. First recombinant insulin was produced by Michael and his coworkers in 1992 from bacterial and yeast systems. In this study, mammalian pancreatic extraction was carried out for mimicking enzyme properties for recombinant technology. Extracted insulin chain and calibrated proteins like cytochrome c, pancreatic peptide, chymotrypsinogen A, and bovine serum were utilized for the stable formation of r-insulin [48]. In 2003, clinical trials for recombinant micro plasmin (μ Pli) were carried for neuroprotection. μ Pli is the modified and purified form of (μ Plg) or Human micro plasminogen. The vector utilized methanol oxidase as a tag for the stability of μ Pli. As μ Pli lacked a lysine binding site during the recombination process, in vivo model showed more potency comparable to native plasmin [49]. Furthermore, much work has been done on osteoporosis, emphasizing the production of recombinant PICP (Procollagen type I

carboxy-terminal propertied). As a native protein, PICP was utilized with HIS and Myc tag for efficacy. The results simplified that the production of recombinant PICP based antibodies showed an elevated bone matrix formation [50].

Correspondingly, rh-aCDase clinical trials have been done for cystic fibrosis and Faber disease. The in-vivo and in-vitro proof of studies provides evidence of elevation in ceramide metabolism on rh-aCDase intervention. The human acid ceramide cDNA was injected in Chinese Hamster cell lines (CHO) for recombinant enzyme production. This recombinant aCDase could antidote the disease by reducing the infiltration of macrophages and ceramide level inside the cell [4]. In another study for aCDase domain analysis, a recombinant sequence for rh-aCDase was prepared. The highlight of this study was the utilization of melittin signal peptide as a tag for stability for stability purposes. The sequence for rh-aCDase retrieved from this study can be employed for further recombinant-based research of aCDase. However, this study lacks the targeted administration of aCDase inside the patient [14]. Therefore, studies on embattled CNS administration of enzymes to circumvent this restriction [51]. The usage of cocktail therapy with CNS-targeting moieties is currently assessed as a suitable method for enzyme distribution to CNS [37]. However, recent evidence of experimentations showed delivery of rh-aCDase utilizing onco-retrovirus in a disease similar to SMA-PME. Taking this study into consideration, the delivery of synthetic optimized rh-aCDase can be improved [2].

Besides peptide-based therapy, induction of ASAH1 via biological regulatory networking (BRN) is the current focus of research. Many cancer studies on ASAH1 gene expression with cellular signaling cascades announce the role of transcription factors (TF) to induce aCDase production [33]. Therefore, research to investigate the induction of ASAH1 employing different TF has been carried out. The most collaborative study is the ASAH1 transcription mediated by ACTH/cAMP/CREB signaling pathway [32, 52]. In another study, ERK1/2/MAPK signaling cascade also results in ASAH1 transcripts inside the cell. This pathway also ends up in making interaction with CREB. So CREB is the vital target for the upregulation of ASAH1 gene [53].

2.4 Structural Analysis of aCDase

N-Acylsphingosine Amidohydrolase 1 or ASAH1 gene is responsible for the translation of the ceramidase family. One of the members of this family is acid ceramidase which shows optimal activity at 3.8-4.3 pH inside the lysosome. In 1995 this enzyme was first identified and isolated from rats [12]. The purified glycosylated aCDase (from urine [22]; 50kDa) belongs to the N-terminal Nucleophile superfamily (NTN) of hydrolyses [30]. The enzyme's precursor entails N-Acyletanolamine Acid Amidase (NAAA) and acid ceramidase precursor (C89.001) regions. Post-translational modifications of NAAA region undergo alternative splicing in the form of proteolytic processing [54]. This process consists of the auto-cleavage at Cys-143 reindeer 395 amino acids based single peptide (SP). SP constitutes two subunits of α (1-21 amino acids) and β subunits (22-395) responsible for heterodimer formation inside the lysosome (figure 2.2). The α subunit of heterodimer comprises a helical globular domain made by five α -helixes which are joined together by a linker of 36 amino acids [55]. On the other hand, the β -subunit contains central aligned anti-parallel sheets (two in number) flanked on each side of six helixes. The α subunit is non-glycosylated, while the β subunit has an N-glycosylated terminal for hydrolysis. For stability of the heterodimeric structure, both subunits remain at the distance of 894 \AA^2 where they form two disulfide bonds at positions C388-C392 and C388-C392, respectively. However, the N-terminal linker at α -helix wraps around the β -subunit to complete the dimerization of aCDase [56, 57].

The structure of in-active pro-aCDase and active heterodimer remains the same after the autocleavage process. But the difference lies in the conformational changes of these two states at the α and β subunits junction. At the C-terminal of α subunit, inactive residues 141 and 142 disordered themselves, allowing the helix (135 to 140 residues) to move at a 60° angle [58]. This conformational change reveals Tyr 137 of β subunits, uncovering Cys143 at its base and exposing the hydrophobic binding cavity. The residues Cys143, Arg 159, Asp 162, Glu 225, Asn 320, and Arg 333 make the catalytic domain of aCDase [59]. The enzyme catalysis reveals that Cys 143 of β sheets plays a vital role in hydrolysis, making it a hydrolase domain. At the same time, the helixes and linkers are responsible for stabilizing the heterodimeric structure [33].

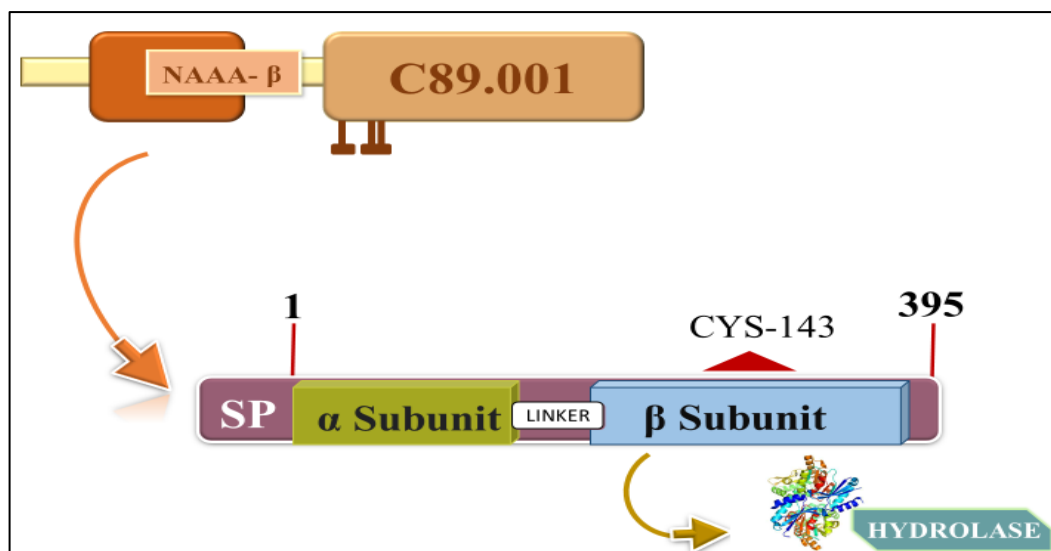


Figure 2.2: Conversion of pro Acid ceramidase (aCDase) to Functional aCDase; Primary structure of aCDase exhibits α subunit and β subunit. The Cys-143 residue of the β subunit is responsible for the hydrolysis of ceramides.

* SP: Signal peptide, N-Acyletanolamine Acid Amidase (NAAA), C89.001: acid ceramidase precursor.

2.5 Structural Analysis of CREB

cAMP Responsive Element Binding Protein (CREB) is a crucial transcriptional factor in maintaining many signaling pathways. This 43 kDa protein transcribe by CREB1 gene and binds to CRE (CREB responsive element) sequence on DNA for gene expression. Exon sequencing of the human genome suggests that CREB is responsible for the gene regulation of about 4000 genes [60]. About its structure, CREB constitutes four domains, each with the specific task for CREB activity. Q1 and Q1 domains of CREB are glycosylation domains required for its post-translational modification [61]. Q1 domain is located at the N-terminal in the primary structure of CREB (figure 2.3). Residue S40 in the Q1 domain is accountable for O-glycosylation at the time of protein forming. Q2 domain is responsible for the hormonal simulation activation of CREB. Q2 is also called Constitutive Active Domain (CAD) due to its interaction with TATA-binding protein-associated factor II. Residue S241 of Q2 domain undergoes phosphorylation, while T228 and S260 are crucial for O glycosylation [62].

However, CREB is activated by KID or kinase inducible domain, stimulated by PKA (protein kinase A) at SER 133 residue. CBP (CREB Binding Protein) and p300 regulate the cellular function by binding at this phosphor-SER133 or SEP133. About 50% of CREB phosphorylation inside the cell occurs at this residue [63]. CBP/p300 has a KIX domain that constitutes 90 conserved residues (from *C. elegans* to *Drosophila melanogaster*), which allow CREB to remain activated during its functioning. The KIX domain comprises three helical structures joined together by a loop region. The helix-A is involved in the interaction with KID domain, making the hydrophobic patch for CREB activation. The helix-B and helix-C wrapped around the patch by 90°, making it stable for a longer time. The interacting residues of CBP-KIX to CREB-KID domain are from TYR649-LEU672, forming the dimeric structure for gene expression. The KIX domain also interacts with ARG130, ARG131 and TYR134 in the vicinity of SEP133 [64].

On the other hand, other kinases like AKT, CaMKt's, MSK1, PKC, CamKII, CK I, and II phosphorylate CREB at SER129, SER143, SER142, TYR143, and ARG145 residues [65]. However, residues SER109, SER111, SER114, SER117, and SER118 are responsible for inhibiting CREB. On the other hand, acetylation occurs at LYS11, LYS94, and LYS136 regions of CREB. Q2 and KID are collectively called Transcriptional Activity Domain (TAD), responsible for initiating the transcription machinery and stabilizing it. Q2 is located beside KID domain, separating it from bZIP (DNA Binding Domain) [66]. bZIP or dimerization domain is responsible for dimerizing the CREB at GC rich site. It utilizes Mg^{+2} ions, which enhance the activity of CREB up to 25 folds [67]. bZIP binds to the palindromic region of DNA (TGACGTCA), modulating the targeted gene expression. This domain is unstable, so a cofactor CRTC (cAMP-regulated transcriptional co-activator (CRTC)) binds at the K225 and K334 residues to sumoylate the domain. CRTC utilizes Ca^{+2} for more vigorous complex formation with bZIP. After protein formation, whole transcription machinery dispersed and transformed into its native state [68].

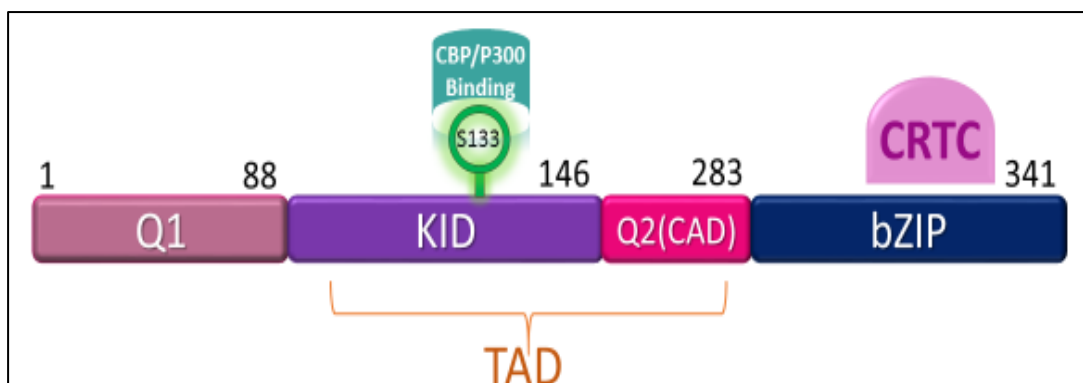


Figure 2.3: Primary Structure of CREB: Q1 and Q2 domains are responsible for glycosylation. KID domain activates CREB by kinases like CBP/P300 at SER133 residue while bZIP domain makes a complex with CRTC and undergoes the transcriptional process.

With time, various cancer studies have been performed to detect the induction of CREB in cell lines. AC_{50} of these inducers describes the activity concentration at which induction happens [69]. In ChEMBL, there are about 7 inducer with AC_{50} values of 1700 nM, 2200nM, 4300nM, 8200nM, 13500nM, 13900nM and 38300nM associated with CREB protein. These values were retrieved from the biological assay performed on Hep G2 human cell lines [70]. However, induction of CREB activity has been tested via 17β estradiol at 0.1 nM. CREB formed stable interactions with CBP protein at this concentration and performed gene regulation [71]. In another study, $1\alpha,25$ -dihydroxycholecalciferol (calcitriol) induced maximum activity of CREB at 100 nM concentration. Calcitriol here up-regulated the binding of CREB protein at its respective CRE sites for cellular processes [72].

A lot of current researches on malignancies involve CREB upregulation. For example, one such investigation utilized retinoic acid to increase CREB/CBP gene regulation inside gastric cancer cells. At the concentration of 100000 nM, retinoic acid increased CREB induction up to 94% inside the GC cells and caused cell proliferation for further study [73]. In another cancer study, treatment of yessotoxin (YTX) at 30nM in K-526 cell lines increased the phosphorylation of CREB up to 15% [74]. However, vandetanib showed upregulation of phosphorylated CREB in neuroblastoma cell lines. The phosphorylated CREB has SER133 residue at the phosphorylated stage (SEP) [75].

In contrast, some in vivo experimentation highlighted the role of antidepressants like reboxetine (Rbx) and fluoxetine (Flx) in the induction of the CREB pathway. Both of these drugs were injected into model organisms (rats) to analyze SER-133 based CREB phosphorylation in one investigation. The study outcomes revealed that Rbx induced early hyperphosphorylation of CREB in the hippocampus, while Flx activated CREB gene expression via ERK2 signaling pathway [76].

CHAPTER 3

METHODOLOGY

3.2.1 MODULE I (r-aCDase)

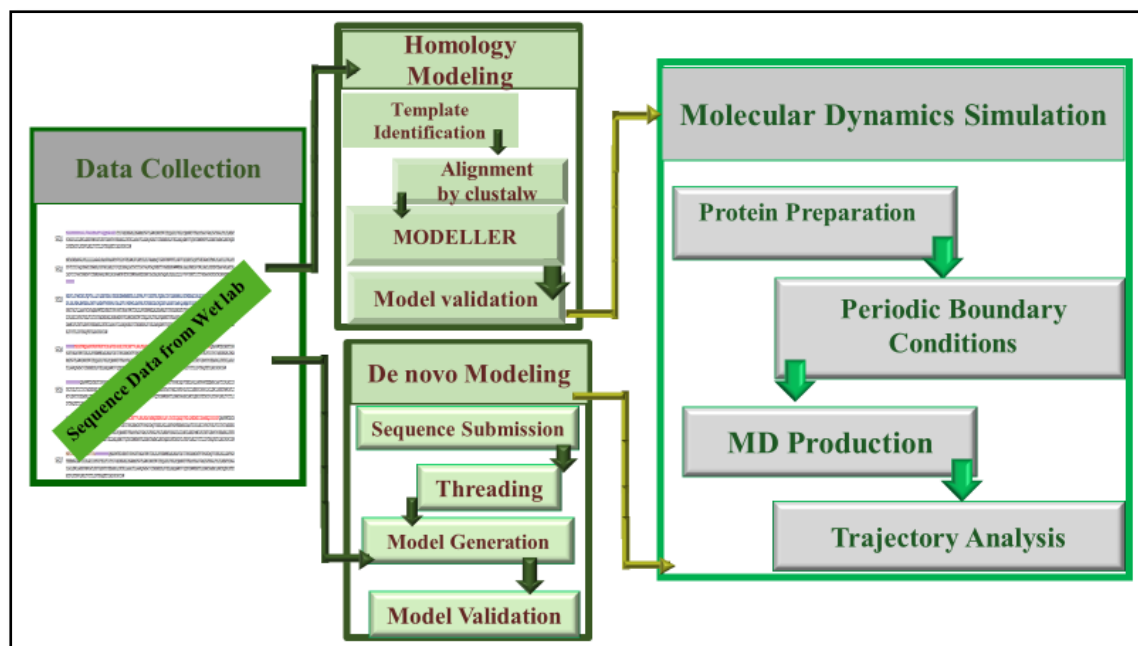


Figure 3.1: Overall Workflow of Module I: (r-aCDase)

3.1.1 Data Collection

The recombinant acid ceramidase sequences (r-aCDase) data weren't available on any online repositories like NCBI, Pubmed, and Protein Data Bank. Therefore, different molecular biology-based companies were contacted for r-aCDase sequences as their research and clinical product. The peptidyl sequence from these companies is the recombinant of different tag proteins with native aCDase. The recombinant aCDase provided by Aviva System Biology Company (<https://www.avivasysbio.com/>) and MyBioSource, Inc., (<https://www.mybiosource.com/>) was found under the product catalog of OPCD01495 and MBS2030445, respectively. The recombinant sequence consists of the N-terminal His tag and β chain of aCDase (Cys143- Trp 395 residues). The His tag is responsible for the higher stability of the r-aCDase. It also helps in increasing purification and yield during the chromatography process [77]. In protocol for synthesis, the source organism for the recombinant process is *E. coli*, while specie reactivity is suitable for *Homo sapiens*. The final product has 33kDa of molecular weight. Likewise, MyBioSource, Inc., has a similar product where they utilize C-6xHIS tag and aCDase

(22-395 amino acids). However, the source organism here is the yeast for the optimization of the protocol.

Toddy Khuc company works as a franchisee of MyBioSource, Inc., provides the r-aCDase product for different organisms as their source host. Under the product id MBS1355633, it offers three r-aCDase sequences. To synthesize 53.7kDa of r-aCDase, they utilize acid ceramidase sequence (Uniport accession # Q13510) with different tags and source hosts. The first sequence includes N-terminal 6xHis-SUMO-tagged protein specific for E. coli as the source host. SUMO-Tag is crucial for protein folding of recombinant aCDase [78]. While second and third sequences constitute N-terminal 10xHis-tagged and C-terminal Myc-tagged with the source organism of baculovirus and mammals. Myc is an antibiotic-based tag protein beneficial here for the subcellular localization of r-aCDase [79].

Another molecular company name Bio-Techne Brands (<https://www.biotechne.com/>), provides services regarding recombinant h-ASAHI protein with the catalogue identity of 4494-AH. The sequence of this protein is taken from Uniport with accession# Q02083 (Met1-Lys359 residues). However, there is a silent mutation at amino acid 151 where there is an isoleucine residue in place of valine. The final recombinant product contains a C-terminal 10 His tag, purified from the Chinese Hamster Ovary cell line. On the other hand, Abnova Taiwan Corporation (<http://www.abnova.com/>) consumes a GST tag in their recombinant aCDase synthesis, which is considered a high-affinity tag for recombinant protein activity [80]. Human ASAHI gene products from 1-395 amino acids have been utilized for product synthesis. They optimized the protocol for an in-vitro wheat germ expression system that preserves the normal biological function of human aCDase.

Cusabio (<https://www.cusabio.com/>) specializes in the production and research of recombinant antibodies and enzymes. For r-aCDase, the product is available under the id number CSB-EP619774HU. The product constitutes of N-terminal 6xHis-SUMO-tagged joined with aCDase via RT linker. Lastly, one sequence is retrieved from the literature from a study on structural activation of aCDase. The author cloned a recombinant r-aCDase by combining N terminal melittin signal peptide followed by 10xHis tag and ceramidase (UniProt Q13510). In addition,

there is resubstituting of isoleucine at position 93 in replacement of valine [65]. The retrieved sequences are displayed in Appendix 1.

3.1.2 Structural Preparation of Sequences

The next step is the conversion of collected sequences from different companies to stable structural protein models. The sequences were divided into tag and aCDase. PDB structure of aCDase was available under the id of 5u7z. For the tag part of recombinant protein, BLAST of the sequences was performed. Unfortunately, most sequences showed 100% alignment for HIS and SUMO tags, but their structure was unavailable. So the sequences having these tags remained unprepared at this step. However, myc tag (PDB:1A93) has a 3D structure of 100% identity and 100% coverage in homo sapiens. The structures that were retrieved from PDB get energy minimized by the Amber99 force field.

On the other hand, for GST tag, homology modeling via MODELLER software was performed [81]. For template selection, 1DUG-*Homosapiens* (100% identity and 91% coverage) was chosen. By copying the coordinates of templates, 100 models for both tags were generated. After checking all models' statistics before and after energy minimization (Amber99), Ramachandran plot and ERRAT score were made. In the end, model 68 for the GST tag was nominated for simulation. The energy minimized models from homology modeling were then stabilized by molecular dynamics simulation (MDS) via GROMACS software at 50 nano-secs. For this step, the tutorial of Lysosomes in Water by GROMACS was followed. Here, topology files of the protein prepared for solvation and ions addition. The defined box was then energy minimized and equilibrated at the OPLS-AA/L force field, and MD production was made for 50nsec. This RMSD stabled 3D structure then processed further bond-making between tag and aCDase.

3.1.3 Constraint Docking

With the purpose of bond formation between tag and aCDase, constraint docking via GOLD-Heme software was implemented. Type of interactions (either Hydrogen bond or Covalent bond) between tag protein and aCDase were reviewed in the literature. This step is first done via MOE software, but protein destabilized after the energy minimization process. In GOLD, residues for both aCDase and tag as ligand

were specified. The type of bonding and constraint weight was designated. The hurdle found here was that both proteins were at distant proximity, so bond formation at this step was a challenging task. Cluspro [82] was used to overcome this impediment, which is responsible for stable protein-protein docking. The resulting recombinant protein destabilized by making a long unstable loop even before the energy minimization step.

3.1.4 De novo Protein Modeling

De novo modeling solved the previously stated complications by avoiding destabilizing bonds between tags and aCDase and predicting the complete recombinant structure of retrieved sequences. Furthermore, those structures which remained unprepared at the first step were also constructed by this approach. This step was done by trRosetta software from the Yang Lab [82], responsible for forming a structure-based prediction model via deep learning. The submitted sequence queued in the cluster system where multiple sequence alignment (MSA) of the respective sequence was made. Based on MSA, homologue templates were detected by the threading technique. The predicted models were constructed on direct energy minimization with specific built-in restrained algorithms, like the orientation of residues distribution and inter amino acid distance. In the end, after complete atomic structure refinement, a 3D protein model was created. After the formation of predicted models, the statistics before and after the energy minimization were calculated. The constructed models were allotted with IDs of their respective sequences for analysis. The Ramachandran plot and ERRAT score indicated the seven final models for each sequence then subjected to molecular dynamics simulations for further stability.

3.1.5 Molecular Dynamics Simulations (MDS)

Structural stability and refinement of modeled proteins were further elucidated by Molecular Dynamics Simulation (MDS) using Maestro-SCHRODINGER 2021 [83]. This step was done by first preparing and refining the protein at 7.4 pH. The protein was then energy minimized at the OPLS4 force field. For the next step, the system was built by utilizing desmond using TIP3P (water built) in a cubic boundary box with a distance of 5Å. These explicit water molecules were added for protein solvation. The system was then neutralized by adding salts (Na⁺/ Cl⁻) and minimized for the readjustment of protein in the solvated environment. Lastly, in the molecular

dynamic section, time steps of 50nsec/100nsec were given at 300K temperature and 1.013 bar pressure. The stability of energetically favorable proteins were investigated based on protein RMSD and protein RMSF.

3.2.2 MODULE II (CREB Structure-Based Methodology)

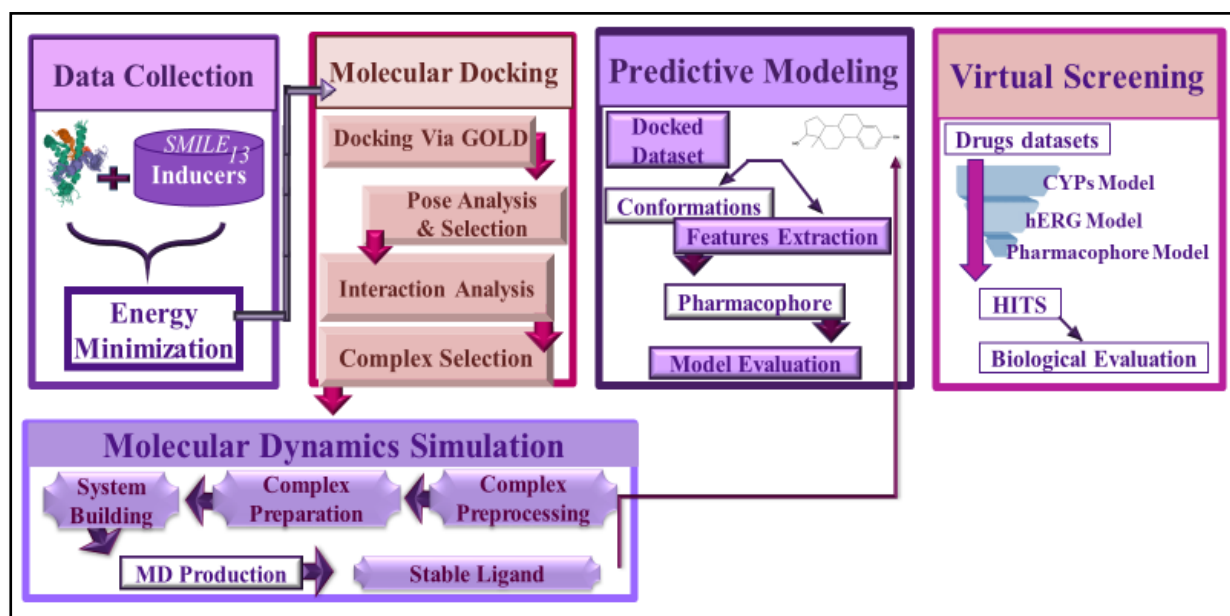


Figure 3.2: Overall workflow of Module II: CREB Based

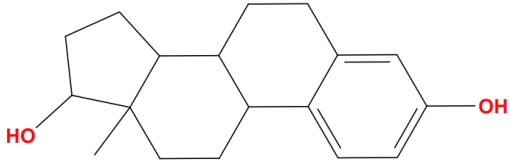
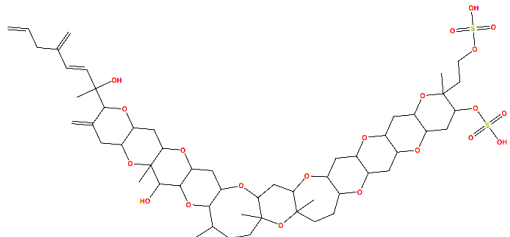
3.2.1 Data Collection and Structural Preparation

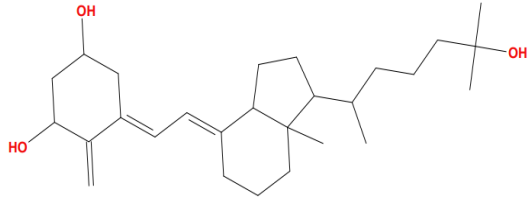
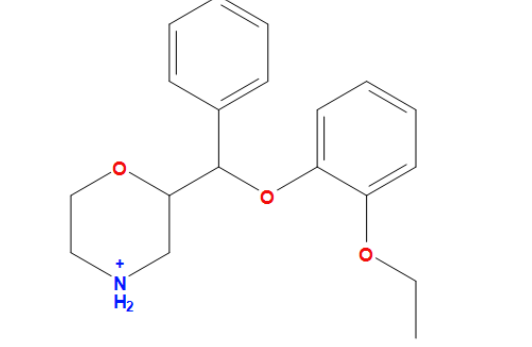
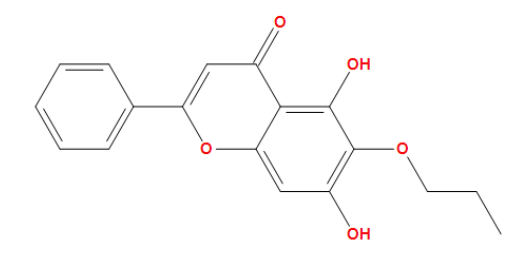
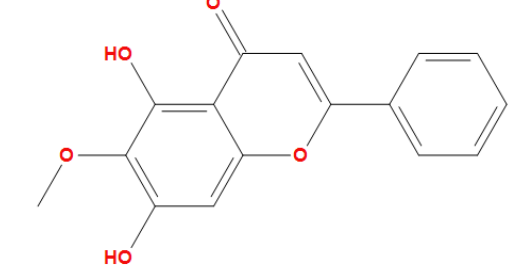
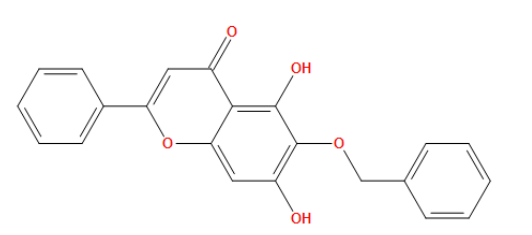
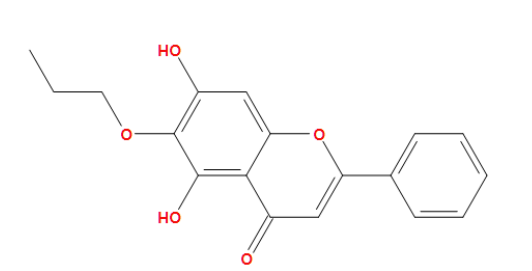
For the second target, CREB protein structure was retrieved from Protein Data Bank. There were different structures available for CREB in the form of complexes. The DNA binding domain (bZIP) structures were PDB: 5ZK1, PDB: 5ZK0, and PDB: 4HTM. However, the current project focuses on the kinase inducible domain (KID) of CREB protein for its induction. For this purpose, two structures, PDB: 2LXS and PDB: 2LXT of KID, were retrieved in the complex form. The PDB Id 2LXT was selected because it lacks mutation in the KID domain and constitutes phosphorylated SER133 residue, which is crucial for activating CREB protein [84]. This homo sapien based NMR structure comprised three entities which making a complex of 16.36kDa. Chain A of the structure forms CREB Binding Protein (CBP) accountable for activating CREB in our body. The second entity, chain B, is Histone-Lysine-N-methyltransferase, which is a positive regulator of CBP. The last entity or chain C is comprised of our protein of interest, the KID domain of CREB (ASP116-PRO149).

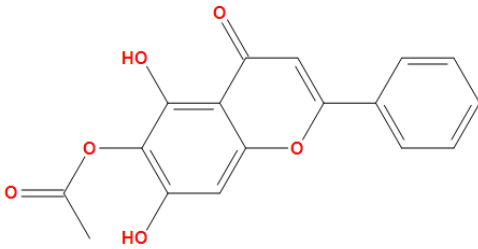
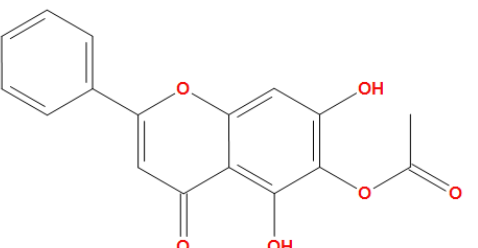
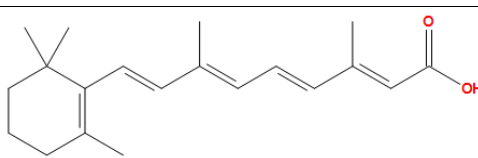
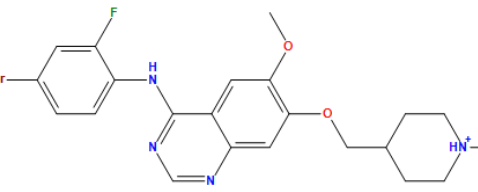
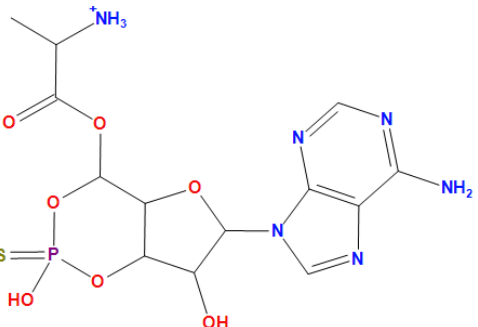
Additionally, chain C consists of a phosphorylated residue at SER133. After phosphorylation, the residue gets modified into SEP at position 133. The researches showed that the presence of CBP is necessary for CREB to remain in a stable conformation. So, the PDB structure 2LXT is first pre-processed by removing chain B to expose the interphase of chain A (CBP) and chain C (CREB). Then, the structure was energy minimized through Amber99 forcefield. The energy minimized structure was then subjected to molecular docking studies.

A structurally diverse data set of 13 inducers with ChEMBL id and experimentally determined activity values (concentration of activation; AC_{50}) against CREB was retrieved from ChEMBL databank and different biological assays [85]. The ligands from AC_{50} 0.1nM to 2200nM were defined as actives. Ligands with $AC_{50} < 2200$ nM were defined as inactive. This data set was then utilized for docking and MD studies. In addition, the simplified molecular-input line-entry system (SMILE) code of each ligand was assessed for making their 3D structure by CORINA software. These 3D structures were then energy minimized by MMFF94 in MOE. Table 3.1 displays the AC_{50} values of inducers associated with CREB.

Table 3.1: Inducers of CREB and their activities.

Inducer	Ligand IDs	Molecular Weight (g/mol)	AC_{50} (nM)	pAC_{50} ($\log_{10}(1/ac_{50})$)
	LIG_1	272	0.1	7
	LIG_2	1143	0.03	7.52

	LIG_3	416	0.1	7
	LIG_4	345	0.3	6.52
	LIG_5	254	1.7	5.76
	LIG_6	284	2.2	5.6
	LIG_7	314	4.3	5.36
	LIG_8	254	8.2	5.08

	LIG_9	446	13.5	4.86
	LIG_10	254	13.9	4.85
	LIG_11	475	100	4
	LIG_12	300	100	4
	LIG_13	254	38.3	4.4168015

3.2.2 Molecular Docking

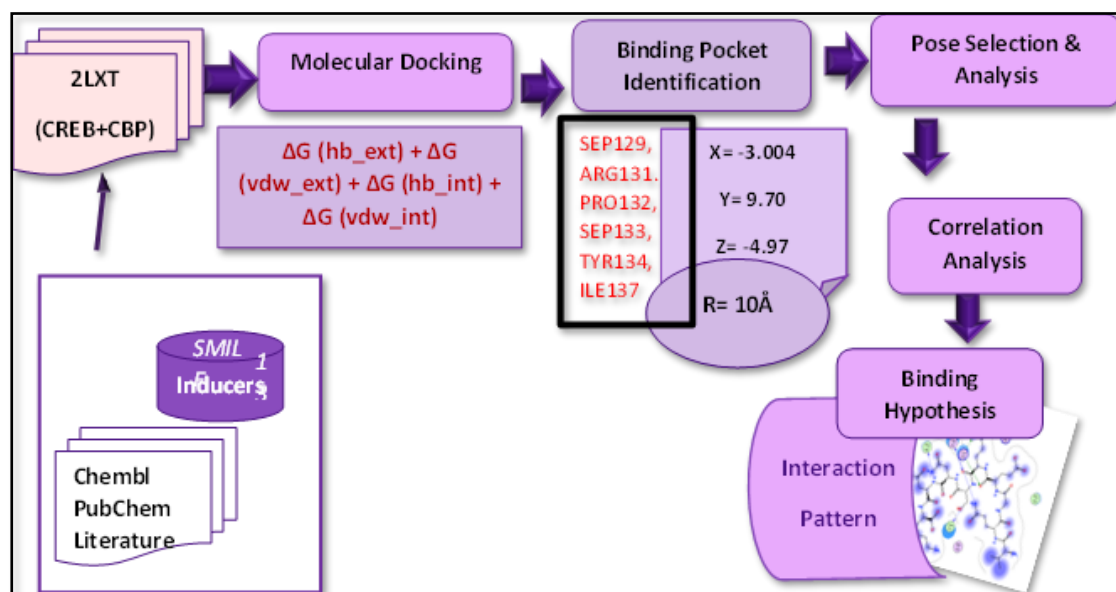


Figure 3.3: Workflow of Molecular Docking. The biological data was retrieved from Protein Databank under the pdb id 2lxt. However, the chemical data of 13 ligands were collected from ChEMBL, PubChem and Literature. Then molecular docking was performed by GOLD software where binding pocket was identified which was the interphase of CREB and CBP. By taking the coordinates -3.0045 (X), 9.7060 (Y), and -4.9740 (Z) docking was performed. After that, the pose analysis and selection were made for correlation analysis and binding hypothesis formation.

3.2.2 Ligand-Protein Docking

The energy minimized data sets of CREB inducers and preprocessed CBP-CREB protein was used for molecular docking step to form the interacting complexes. By utilizing GOLD-Heme suite 5.3.0 [86], 13 ligands with their unique IDs and pAC₅₀ values were docked in the catalytic domain of CREB protein. The binding pocket was defined by taking the interacting interphase of CBP with CREB. The interacting residues of the CREB binding site were ARG131, PRO132, SEP133, TYR134, ARG135, LYS136, and ILE137, as reported in previous docked studies [65]. In contrast, the residues of the CBP cavity site were TYR649, TYR650, HIS651, LEU652, LEU653, ALA654, GLU655, LYS656, ILE657, TYR658, LYS659, ILE660, GLN661, LYS662, GLU663, LEU664, GLU665, GLU666, LYS667,

ARG668, ARG669, SER670, ARG671 and LEU672 [65]. The ligands were considered flexible to the binding pocket. The binding coordinates around the binding cavity (20Å) were -3.0045 (X), 9.7060 (Y), and -4.9740 (Z). The slow protocol was followed for accuracy and high precision of the software. Initially, 100 Genetic Algorithm runs were generated for each ligand. However, due to the software selection of early termination on repetition of a conformer, some ligands showed less conformation.

GOLD fitness score, which combines certain energy values, was utilized to select the top 10 poses. For each ligand, the highest valued pose was selected. The AC_{50} values were converted into pAC_{50} using the formula $pAC_{50} = -\log (AC_{50} \text{ in M})$. A correlation (R) plot between inducers' pAC_{50} value (y-axis) and their GOLD fitness score (x-axis) was calculated. The R²-value (0.29) showed that a weak correlation might be due to unstable interacting patterns, i.e., the most active ligand, **LIG_1** (0.1nM) showed a weak binding pattern. So, to stabilize the interaction and 3D conformation of the ligand, molecular dynamic simulations were performed. This stable conformation was then used for 3D predictive model formation.

3.2.3 Molecular Dynamics Simulations

After docking analysis, the highly active compound w.r.t pAC_{50} value was selected for feature analysis. So to stabilize the interacting complex, molecular dynamic simulations (MDS) were executed using Maestro-SCHRODINGER software [87]. For this purpose, ligand and protein complex was prepared by removing the CBP from the interphase of CREB in MOE [88]. Then, the complex was prepared in Desmond-SCHRODINGER software at pH 7.4 [89]. The workplace structure was preprocessed by adding hydrogen and refined by optimizing the H-bond assessment at pH 7.4. A default force field of OPLS4 was selected for energy minimization to remove any strong van der Waals interactions that may cause structure distortion within the complex. The complex was then solvated in a TIP3P water modeled cubic box of 5Å^o, where Na⁺ or Cl⁻ (salt concentrations = 0.15M) were added to neutralize the system [90]. After equilibrating the protein-ligand complex, MDS production was made at the standard temperature (300K) and pressure (1.01325 bar) for 400 nano sec. The resultant CαRMSD and protein-ligand interaction plot highlighted complex

stability with its interacting residues for that ligand. An abstract idea of MD simulation is shown in the figure below.

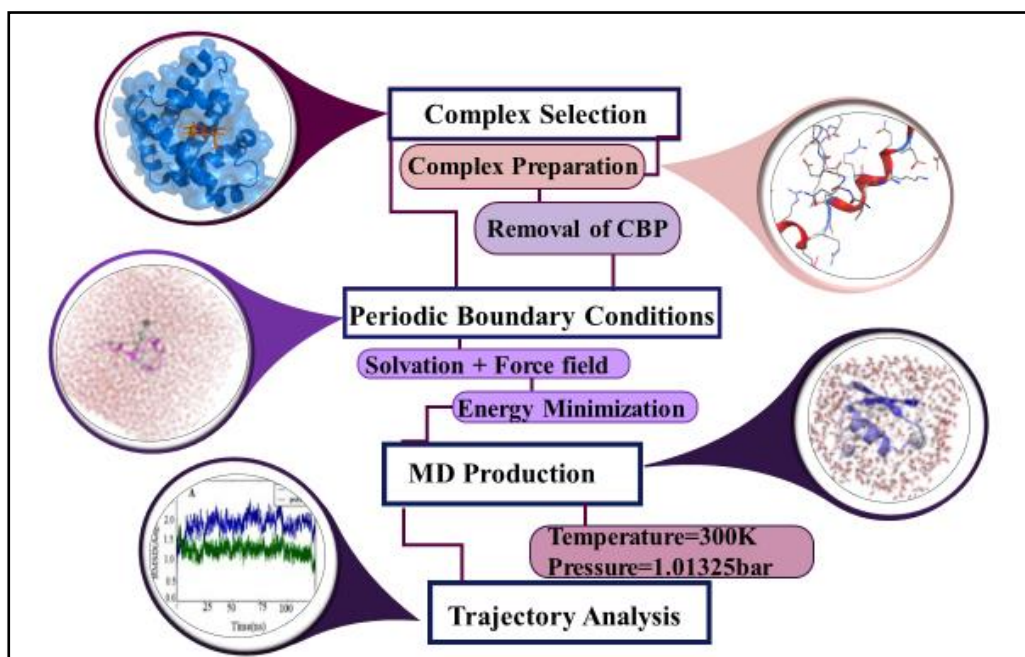


Figure 3.4: Overall workflow of Molecular Dynamics Simulations of selected highly active ligand/CREB complex. This process was done by removing CBP from the docked complex and processed towards system building. After energy minimization MD production for 400nsec was made and the results were analyzed by means of trajectory analysis.

3.2.4 Ligand-based Pharmacophore Analysis

After MD simulation, the stable interacting conformation of highly active ligand i.e., LIG_1 was achieved. This ligand after md (**mdLIG_1**) was then utilized for the feature extraction via pharmacophore modeling. According to the International Union of Pure and Applied Chemistry (IUPAC), a pharmacophore model is based on those electronic and steric features elucidates for the interaction between the biological entities during a biological response [91]. Based on features like Hydrophobic (Hyd), Donor and Acceptor (Don&Acc), Hydrogen bond Donors (Don), Aromatic rings (Aro), and Hydrogen bond Acceptors (Acc), the model discriminates the actives ligand from inactive. Therefore, with the help of a stable active ligand, the pharmacophore model was build using the EHT option in Pharmacophore Editor-MOE.

Ligand-based pharmacophore was built on five ligands as actives and seven ligands as inactives. The MD stable ligand (**mdLIG_1**) (0.1nM) which was retrieved from the docked dataset to stabilize the interacting pattern and conformation, was used as a template and removed from the docking result database. The rest of the docked ligand database (top poses) was then utilized to form the best conformation via the stochastic search algorithm in MOE. The resultant database (*conf_out.mdb*) was then packed via the SVL command for the Pharmacophore query. In Pharmacophore Editor, the template was loaded in the 3D window of MOE [88], and the EHT feature generation protocol was selected. The features observed from the model were then applied for hit identification. The overall procedure of pharmacophore modeling is shown in figure 3.6.

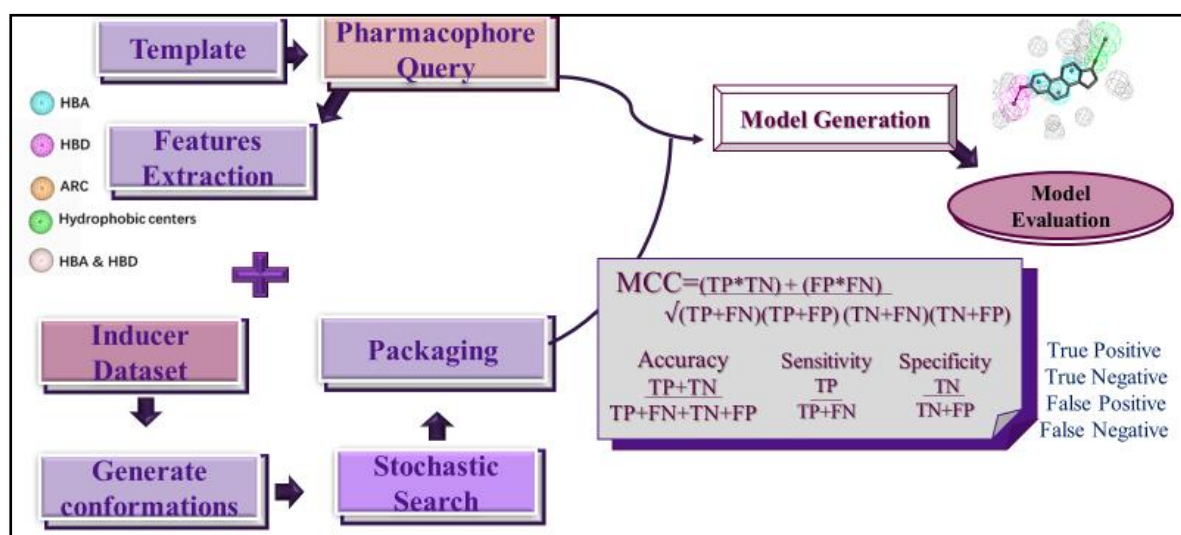


Figure 3.5: Overall workflow of Ligand based Pharmacophore Model Generation using stable conformer from MD simulation studies. The step was done by extracting the feature from the template and generation of confirmation from rest of the docked data. Based on features, model was generated and screened against the packed confirmation for internal set validation.

After the formation of features with their specified radius, hit identifications were made. The packed conformational database (*packeddocked_conf.mdb*) was given as input in the pharmacophore search. The algorithm searched for all possible hits with features similar to the input ones. By evaluating the number of actives and inactives in the internal hits, the respected pharmacophore model validation was made. This step was crucial to predict the reliability of a model. Therefore, a confusion matrix algorithm was used to classify the dataset. Those ligands which are actives ($AC_{50} < 2200nM$) and predicted active

by the pharmacophore model were stated as True Positives (TP). Similarly, True Negatives (TN) were those inactive ligands ($AC_{50} > 2200\text{nM}$) that were predicted as inactive by model. While those that are active in nature but do not appear in hits, i.e., predicted as inactives, were False Negatives (FN). Likewise, False Positives (FP) were those ligands that were inactive but recognized as actives by the model.

After classification of ligands, evaluations were made on certain statistical parameters such as Specificity (true negative rate), Sensitivity (true positive rate), Accuracy (overall prediction rate), and Precision (positively predicted values). However, the decision was clarified based on Matthews Correlation Coefficient (MCC), which correlated the actual and predicted value by taking all classified variables (TP, TN, FP, and FN). The model was predicted best by its prediction statistics and utilized further for the virtual screening of different libraries.

3.2.5 Virtual Screening

The best-featured model was used for the virtual screening. For this purpose, three different libraries were established and extracted. The first library was prepared by taking data for all online FDA-approved medicine that has been used for epilepsy and spinal muscular atrophy. This library consisted of 48 drugs with antiepileptic, anticoagulative, and stabilizing muscle control properties. The second library was extracted from Drug Bank Database [92], an online repository for all FDA-approved drugs (10631). In contrast, the third one is retrieved from the commercially available natural compounds database, i.e., the ZINC Database [93]. These databases were then individually screened against CYP's model to evaluate the metabolic activities of the compounds. The filter is commercially available at the Online Chemical database server (ochem.eu), consists of a further 5 CYP's models such as CYP2C19, CYP2C9, CYP1A2, CYP2D6, and CYP3A4 to screen out the compound, effectively. The compounds screened from the model were stated as the non-inhibitors of CYP.

The non-inhibitors obtained from the above screening then passed through the in-house hERG pharmacophore model, a cardiotoxicity prediction model prepared by Dr. Saba Munawar [94]. The non-hits of the model were declared to be nontoxic against the hERG channel. The selected compounds were then screened against the best-featured model to identify the most potent external hits. Finally, the hits from the CREB

pharmacophore model proposed as effective and safer leads against CREB. The overall procedure is shown in figure 3.7.

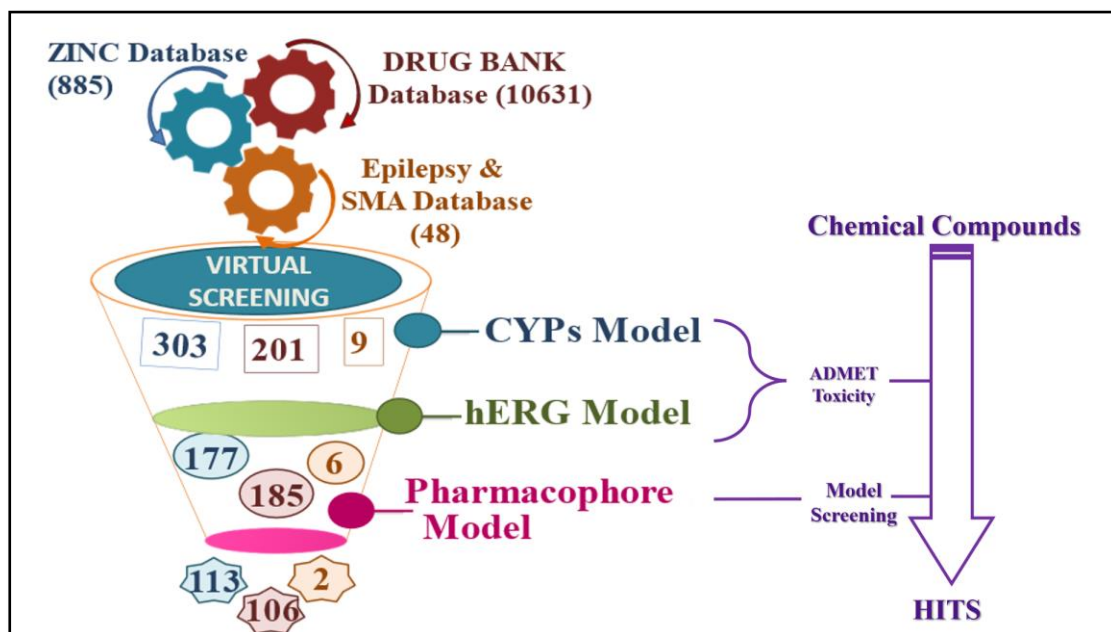


Figure 3.6: Different steps of virtual screening; from ADMET toxicity to Hits identification. Number of compounds in blue color (ZINC database), pink color (DRUGBANK database) and orange (EPI/SMA database) color at the end of each step declares no of hits of the respected validation step.

CHAPTER 4

RESULTS

4.1. MODULE I (aCDase Structure-Based)

4.1.1 Structural preparation of sequence

During the structural preparation of sequences, finding of homologous template was a difficult task. For tags like HIS, myc and SUMO no structural template was available for copying of coordinates in homology modeling. However, one homologous template was available for GST. So, this tag was utilized for its homology modeling. For this purpose, the sequence of the GST tag was submitted at BLAST to find the local region similarity with other PDB sequences. The results showed that 1DUG had 100% identity and 91% coverage. This protein was then used as a template to copy the coordinates for homology. Figure 4.1 shows the sequence alignment for modeling. The alignment score was 98, which suggested 1DUG as the best template for homology modeling via MODELLER.

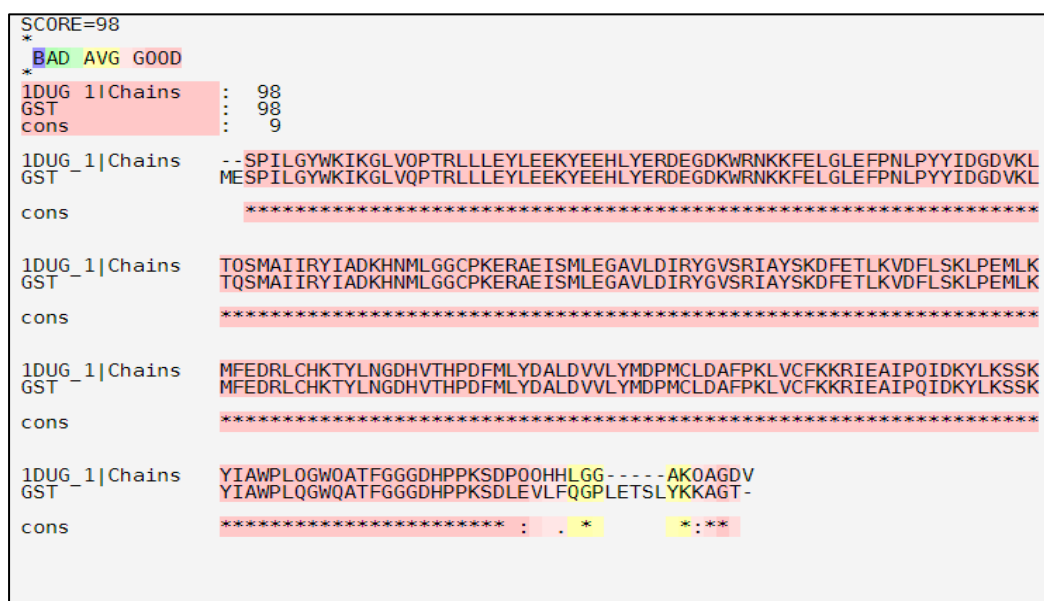


Figure 4.1: Sequence alignment of GST tag's sequence and 1DUG sequence. The red highlighted region indicates the perfect match between both sequences.

After completion of homology modeling, 100 homologous structures were generated. Based on the ERRAT score and Ramachandran plots, model no 68 was selected for Molecular Dynamic Simulation. For this purpose, MD simulation was performed for 50nsec via GROMACS. The simulation results highlighted the stability of the structure with an RMSD of 5.2Å with respect to time. The model's statistics were determined via the Ramachandran plot. This plot is the combination of different

regions such as allowed region (red color), generously allowed region (yellow color) and disallowed region (white color). Here large number of protein residues located in allowed and generously allowed region. However, the statistics showed the three residues in the disallowed region thus stated as outliers. These residues are ALA228, Gln68 and Lys 219. The least number of outliers in the plot indicated a stable structure formation (figure 4.2).

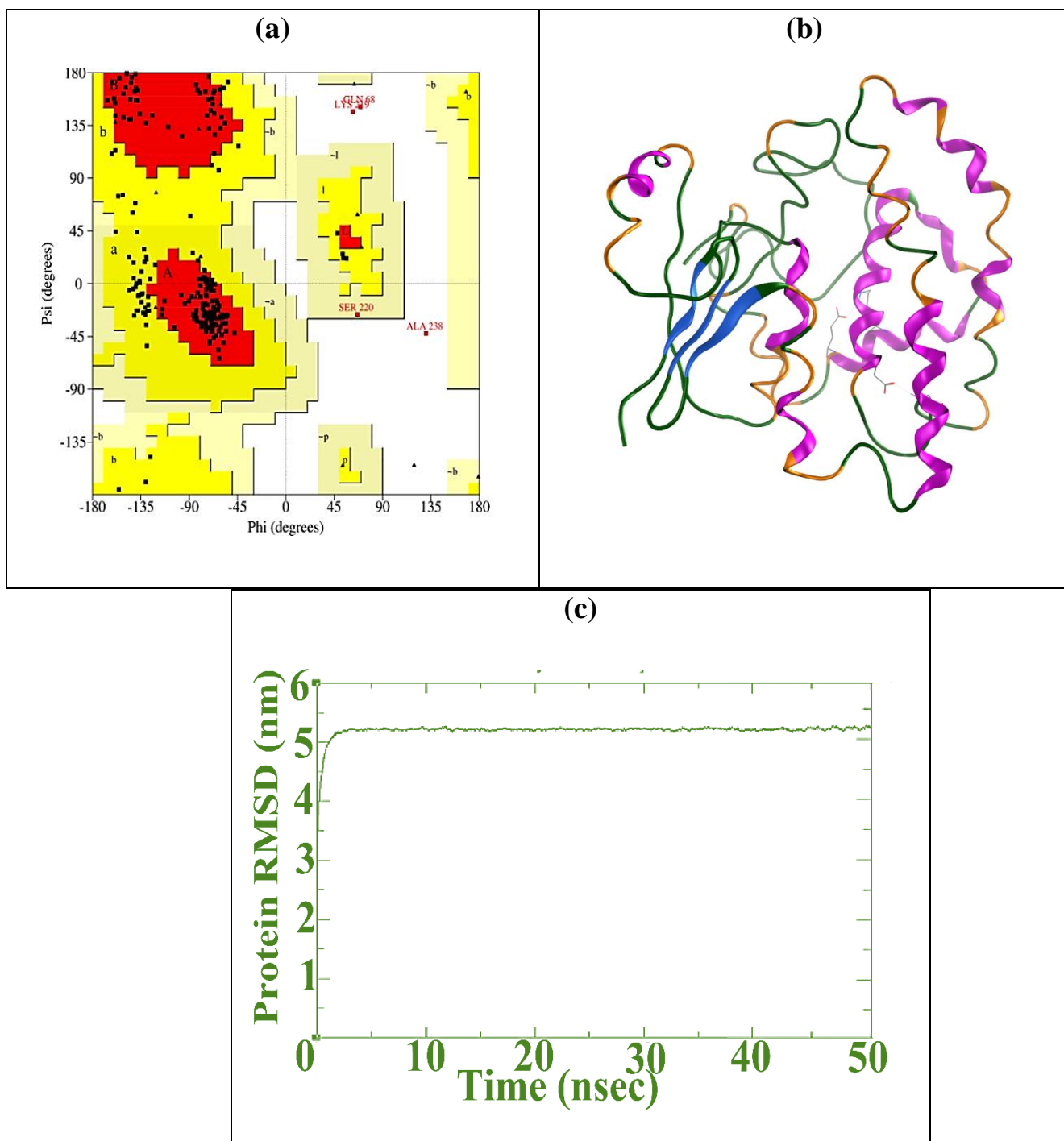


Figure 4.2: (a) Ramachandran plot of stable GST model. About 149 residues are in the allowed region, with three outliers in the disallowed region. (b) Stabled GST tag structure after homology modeling. (c) RMSD vs. Time plot after MD simulations suggested the stability of the protein's backbone with the time of 50 nsec.

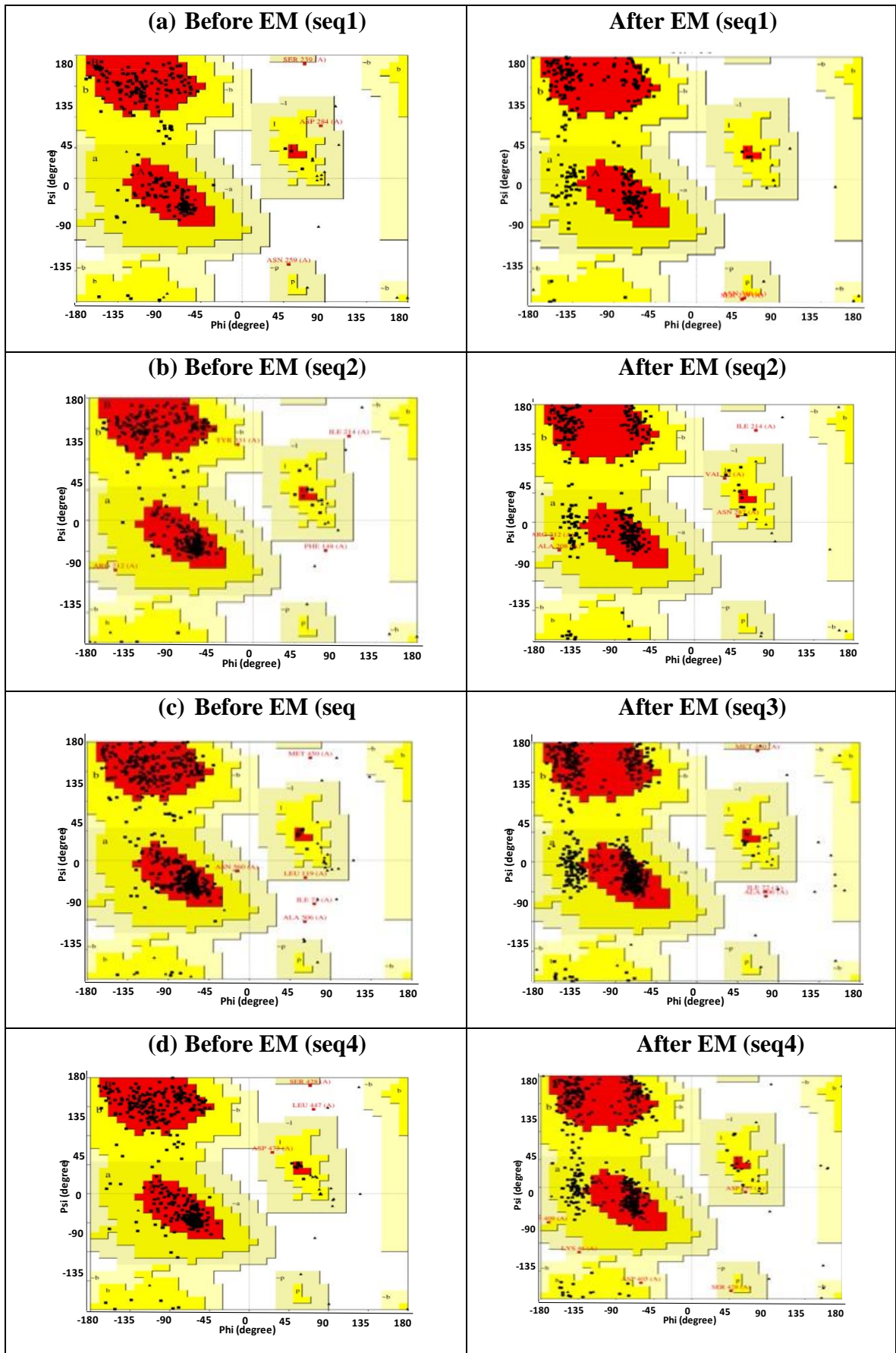
4.1.2 De Novo Modeling

Due to the unavailability of a homologous template for some tags and unstable structure after protein-protein docking, trRosetta was utilized for protein structure formation. The web-based server, trRosetta, significantly improves protein prediction by targeting the co-evolutionary data via deep learning and predict the inter-residue orientation along with distance [95]. Briefly, five models were generated for seven sequences, respectively. Based on the highest TM value, the confidence score of the predicted model, the top model was selected. (Table 2). In the next phase, model evaluation, ERRAT scores were calculated to analyze the quality factor of protein atomic interactions. The highest ERRAT score was shown by seq2 (93.90), seq3 (94.72), and seq7 (92.36) (Table 2). These values indicated the formation of stable pairwise interactions between bonded atoms [96]. The models were further evaluated using the Ramachandran plot to visualize the stable residues (allowed region and additionally allowed region) and the unstable residues/outliers (disallowed region). Here based on ψ - ϕ torsion angles [97], the outlier behavior was observed, mainly for seq5 (4 residues in the disallowed region), seq6 (5 residues in the disallowed region) and seq (7 residues in the disallowed region), respectively (Table 4.1). The models were then subjected to energy minimization using the Amber99 force field in MOE and re-evaluated via Ramachandran plot.

After the energy minimization, as the modeled protein achieved stable confirmation, the residual number of the outlier declined, and the graphs showed better statistics than before (figure 4.3).

**Table 4.1: Statistical parameters of the modelled protein before and after energy minimization (forcefield Amber99).
The ids of constructed models remain similar to their respective sequences for analysis**

	Tm-Score from trRossetta	Before MD simulations				After MD simulations			
		allowed residues	additionally allowed	generously allowed	disallowed	allowed residues	additionally allowed	generously allowed	disallowed
seq1	0.83	227	27	2	1	176	73	2	0
seq2	0.78	281	34	2	2	226	88	4	1
seq3	0.54	504	42	2	3	397	151	0	3
seq4	0.77	337	36	1	2	280	131	5	0
seq5	0.83	302	35	1	4	233	107	1	1
seq6	0.75	372	43	3	5	291	123	6	3
seq7	0.81	303	37	4	9	243	98	10	2



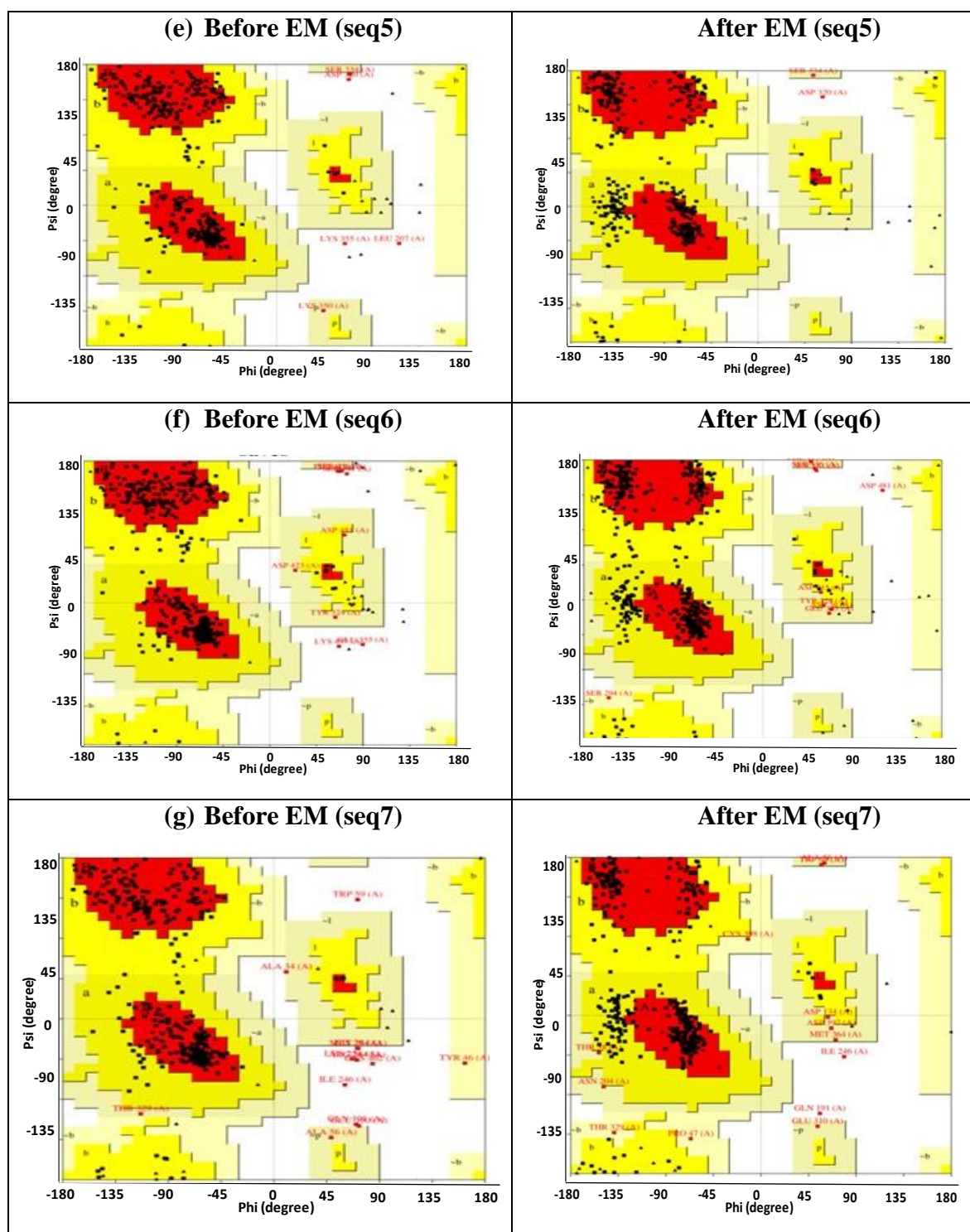


Figure 4.3: Ramachandran statistics of Seq1-7 proteins (a-g) before and after energy minimization. The number of outliers (residues in the disallowed region) significantly decreased after energy minimization, resulting in refined 3D structures.

4.1.3 Molecular Dynamics Simulations (MDS)

The modeled protein was further subjected to molecular dynamics simulations (MDS) utilizing Maestro-Schrodinger for further optimization. All structures stabilized at 50nsec except seq5 and seq6, which showed structural stability at 100nsec due to the presence of a loop between tag and aCDase. The stability of structures was measured in terms of the deviation of backbone (α -carbon) structure with reference to its initial confirmation [98]. For this purpose, the root mean square deviation (RMSD) plot was visualized for all structures. RMSD (eq1) measured the average change in dislocation of atoms in a protein with a simulation time frame. It was measured by

$$RMSD = \sqrt{\frac{1}{N} \sum_{i=1}^N (r'(t_x) - r'(t_{ref}))^2} \dots \dots \dots \text{Eq1}$$

Where N and r' were selected atoms and position of modeled protein, t_{ref} was time reference of the first-time step, and t_x was the total simulation trajectory time frame. Figure 4.4 (a-g) shows the RMSD of 3D models of the respective sequences. For seq1, the structure showed stability at 15nsec with 4.5A° RMSD. After 20nsec, the backbone exposure of protein with respect to RMSD remained between 4.8A° to 5.6A°. Similarly, for seq2 and seq3, the trajectory continued to evolve till 30nsec with 4.0A° RMSD and 12ns with 10.5A° RMSD, respectively. Both modeled structures showed a stable state at 50nsec time-step with RMSD values of 4.8A° and 9.0A°, separately. For seq4, the structure showed steady behavior with backbone convergence to 4.2A° to 4.8A° from 20nsec to 50nsec simulating time. As modeled structures of seq5 and seq6 comprised a loop, simulation time steps of 100nsec were required for its complete stability. The α -carbon deviation for seq5 remained between 3.0A° to 4.5A° while the seq6 modeled structure showed 12.0A° RMSD. The backbone structure deviation was least shown by the seq7 modeled protein with 3.2A° RMSD at 20sec time step, which was further stabilized at 50nsec with an RMSD of 3.5A°. Considering the less deviation from the original (reference) structure and prolonged steady simulation time i.e. $(r'(t_x) - r'(t_{ref}))^2$, seq4 and seq7 proteins declared to have stable trajectories throughout system evolution.

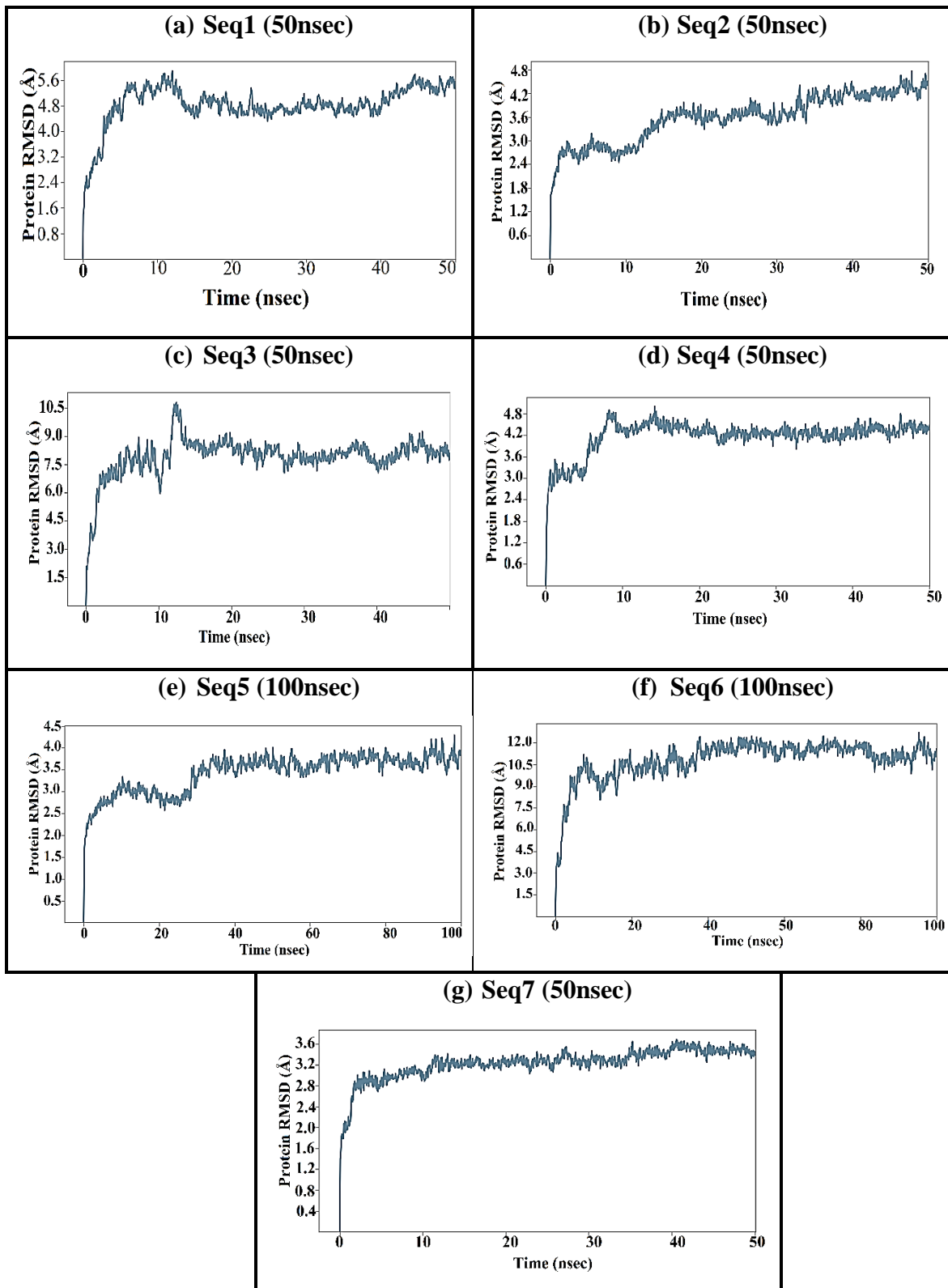
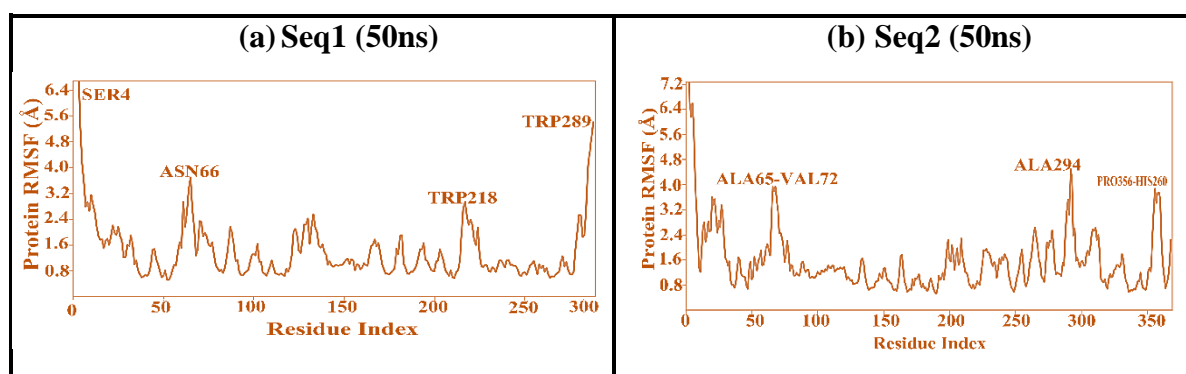


Figure 4.4: RMSD plots for modeled proteins of seq1 to seq7 (a-g), respectively. The plots showed backbone convergence (RMSD) on the y-axis and simulation time steps (nsec) on the x-axis. Based on the least convergence and longer stability time, seq4 and seq7 proteins showed the most stable state.

To strengthen the results from the RMSD plot, the modeled residues' root-mean-square fluctuation (RMSF) was also investigated. Figure 4.5 (a-g) shows the RMSF plot for all modeled proteins. The spikes showed the fluctuations among the residues during the period of simulation. The N and C terminal showed more fluctuating behavior than the rest of the protein due to the presence of a loop structure [99]. The RMSF plot for seq1 characterized the mobility of some single residues like SER4, ASN66, TRP218, and TRP289. The SER4 and TRP289 showed fluctuation of 5.6\AA and 6.4\AA due to terminal atom positions. However, ASN66 and TRP218 showed a minor RMSF at 3.2\AA and 4.1\AA , respectively. ALA294, ALA65-VAL72, and PRO356-HIS260 showed the same behavior in seq2 at 4.2\AA RMSF. In seq3, the change in residual position reached up to 7.0\AA to 8.0\AA for LEY49, GLU117, and MET169. For seq4, the highest fluctuation remained between 3.2\AA to 4.2\AA for three residues only (ARG61, PRO121, ASN255), leaving the whole structure stable below 2.4\AA RMSF. For seq5 and seq6, the fluctuations were shown by THR17-PRO26 and HIS5-ALA18 region, which are associated with a larger loop in both proteins, respectively. Lastly, in seq7, the whole residual stability was shown at 2.5\AA RMSD with only two peaks for PRO112 and ASN204. Based on the residual mobility compared to its position, seq4 and seq7 showed less fluctuated behavior in all modeled proteins.



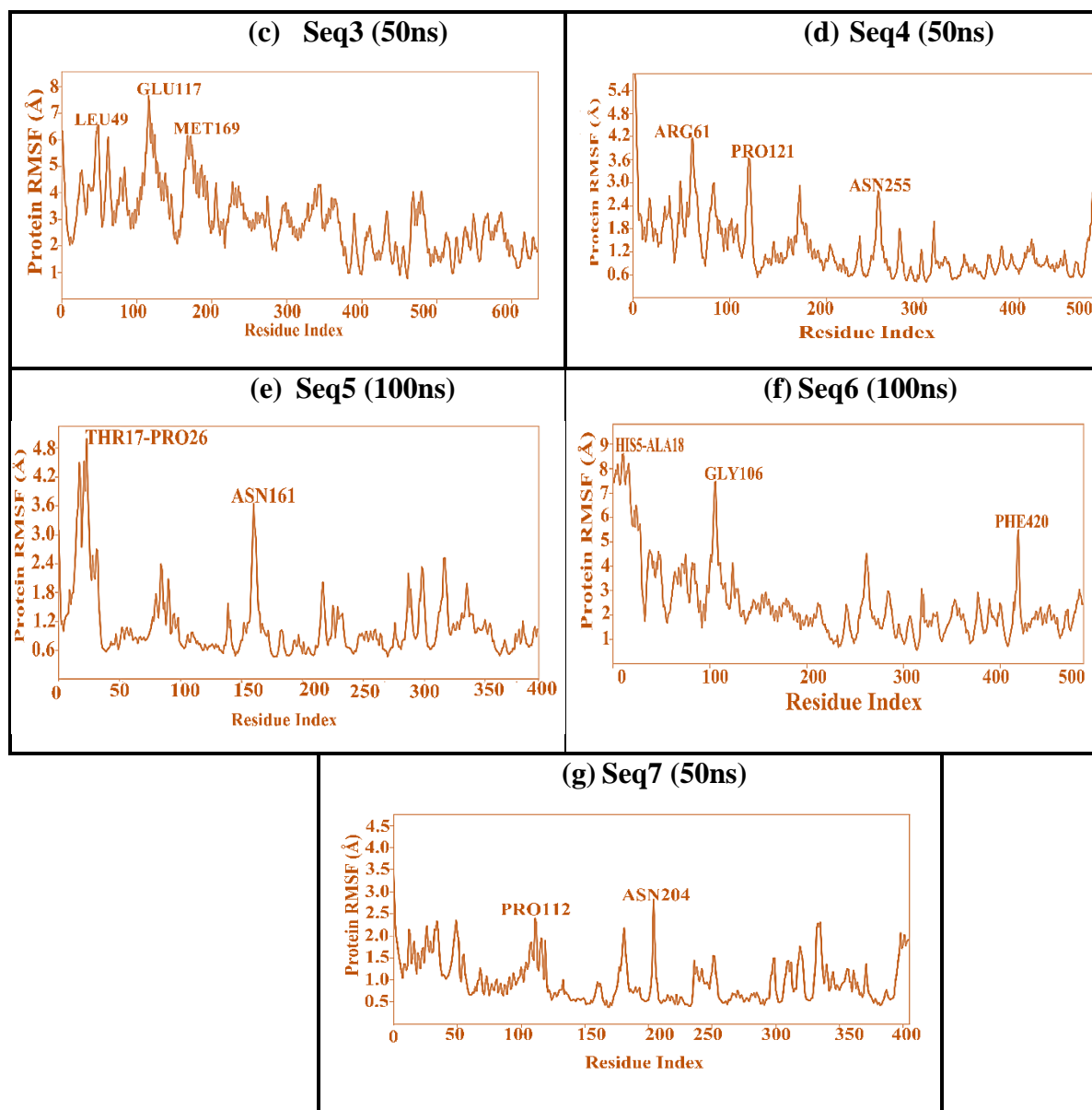


Figure 4.5: RMSF plots for modeled proteins of seq1 to seq7 (a-g), respectively. The plots showed fluctuation in residues (RMSF) on the y-axis and residue index on the x-axis. The region with a loop region indicating the most fluctuated behavior.

Table 4.2 Statistical parameters of Ramachandran plot before and after the 50 nsec of MD simulations

	Before MD simulations				After MD simulations			
	allowed residues	additionally allowed	generously allowed	disallowed	allowed residues	additionally allowed	generously allowed	disallowed
seq1	176	73	2	0	200	46	3	2
seq2	226	88	4	1	281	33	3	2
seq3	397	151	0	3	462	83	2	4
seq4	280	131	5	0	349	60	3	4
seq5	233	107	1	1	284	53	4	1
seq6	291	123	6	3	334	80	5	4
seq7	243	98	10	2	295	52	4	2

All stable structures were then visualized in MOE and were energy minimized after cleaning the structure from solvents. The energy minimization at Amber99 forcefield left most of the tag structure in the loop form, making it unstable. This step was further investigated by the Ramachandran plot to identify the outlier regions of the simulated proteins (Table 4.2). The statistics for modeled proteins improved after MD simulations. The residues in the allowed region increased significantly after simulations. Some of the outliers which were appeared in plots were lessened after energy minimization. The results showed that seq3, seq4, and seq7 were the most energetically favorable recombinant proteins (figure 4.6 (a,b,c)).

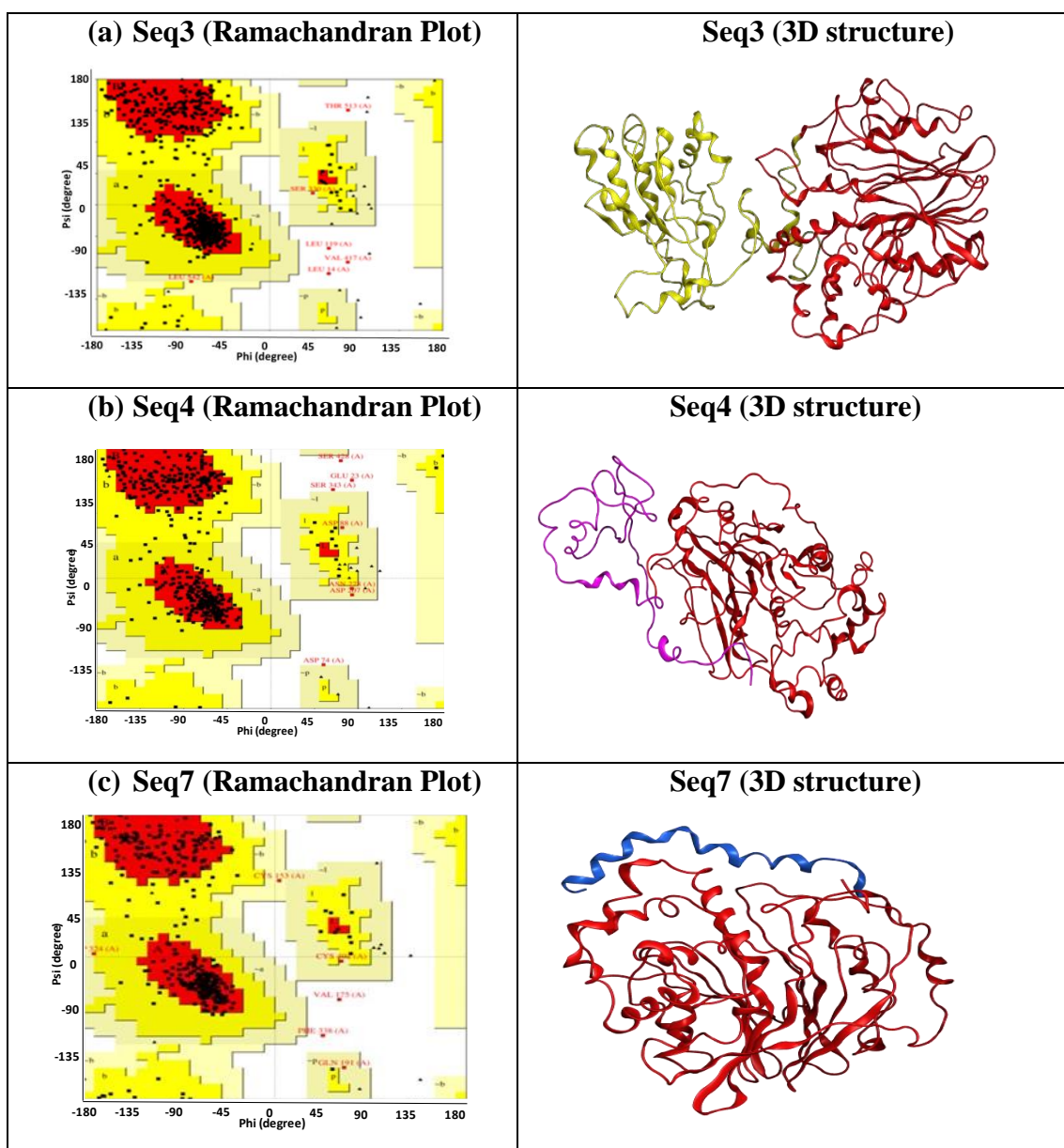


Figure 4.6: Ramachandran plot of the selected recombinant proteins (left) and their 3D structures (right). (a) seq3 Ramachandran plot and stable modeled structure. The red color structure represents the aCDase, while yellow declared GST tag. (b) The ramachandran plot of seq4 with its stable modeled structure (red color shows aCDase domain while purple color represents Histidine + SUMO (small ubiquitin modifier like protein). (c) The ramachandran statistics of seq7 with its stable structure. Blue structure represents Melittin peptide tag while red shows aCDase hydrolytic domain.

So as per the analysis of RMSD plots, RMSF plots, and structural analysis, it was concluded that the order of recombinant aCDase according to its stability is as follow:

Seq7>Seq4>Seq3

Here it can be assumed that these recombinant proteins were energetically stable, and the tag used in these proteins (melittin peptide, HIS+sumo tag, GST tag, respectively) should be effective to respond when fused with aCDase as a treatment.

4.2. MODULE II (CREB Structure-Based)

4.2.1 Molecular Docking

The energy minimized CBP-CREB protein and 15 ligands (involving two in-vivo) then undergo molecular docking via GOLD-Heme suite 5.3.0. The area of 20Å was selected with coordinates of -3.0045 (X), 9.7060 (Y), and -4.9740 (Z) around the binding interphase region of CBP and CREB (Chapter 3, Methodology). The ligands bind at the interphase of CBP and CREB validate the accuracy of the docking procedure (figure 4.7).

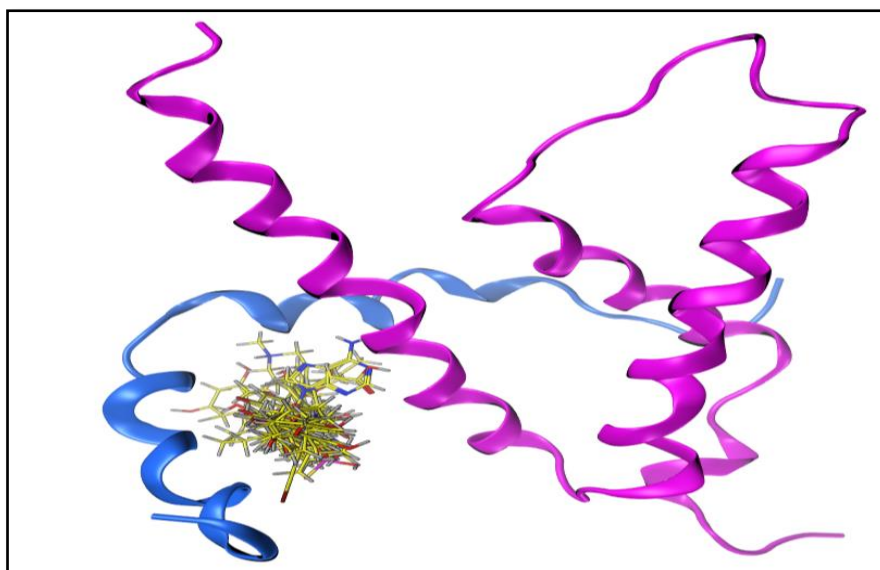


Figure 4.7: Docked ligands (yellow) in the interphase of CBP (purple) and CREB (blue). CBP wrap at a 90° along with CREB for its structural stability.

For each ligand, 100 poses were generated with the high precision protocol. The resultant GOLD fitness score (eq 2) [100] was calculated by considering all the energy values as follow:

$$\text{GOLD Fitness} = \text{Score}(\text{hb_ext}) + \text{Score}(\text{vdw_ext}) + \text{Score}(\text{hb_int}) + \text{Score}(\text{vdw_int}) \dots \text{Eq2}$$

The GOLD fitness score is the sum of all external H-bond and Vander Waal energies of interacting ligand and protein and ligand's internal energies such as torsion strain energy, intramolecular H-bond energy and Vander Waal energy between the ligands atoms. However, the fitness score for each ligand was in the continuous range.

To analyze the interacting binding pattern with a specific GOLD fitness score, a correlation plot (R) was made by taking the GOLD fitness score on the x-axis and pAC₅₀ on the y-axis (figure 4.8). The resultant graph showed a correlation of -0.4916 (R²=0.29), which suggested a weak correlation due to abnormal behavior of interacting patterns. Generally, it is assumed that the highly active compound should have a stable binding pattern, thus a high GOLD fitness score [88]. However, in the plot, it can be observed that all actives compounds had GOLD fitness scores lower than the inactives score in the data. The highly active compound (0.1nM) showed GOLD score of 38.5988 with pAC₅₀ value of 10. This behavior suggested an unstable binding pattern due to the presence of CBP during the docking procedure. As the

selected ligands were only compatible with CREB. So, due to the presence of CBP, the ligand docked inside the interphase but lost its stable interaction during the protocol. The residual binding pattern further strengthened this hypothesis, i.e., the actives docked ligands showed strong H-bond interaction with TYR650, HIS651 and ALA654 of CBP, which changed its binding position to SER121, ARG125, and ARG135 residues of CREB. Therefore, stabilization of binding interaction and conformation was necessary for further processing. So, to stabilize the interactions and conformation, a highly active ligand was selected for MD simulations.

$$(R = -0.49, R^2 = 0.29)$$

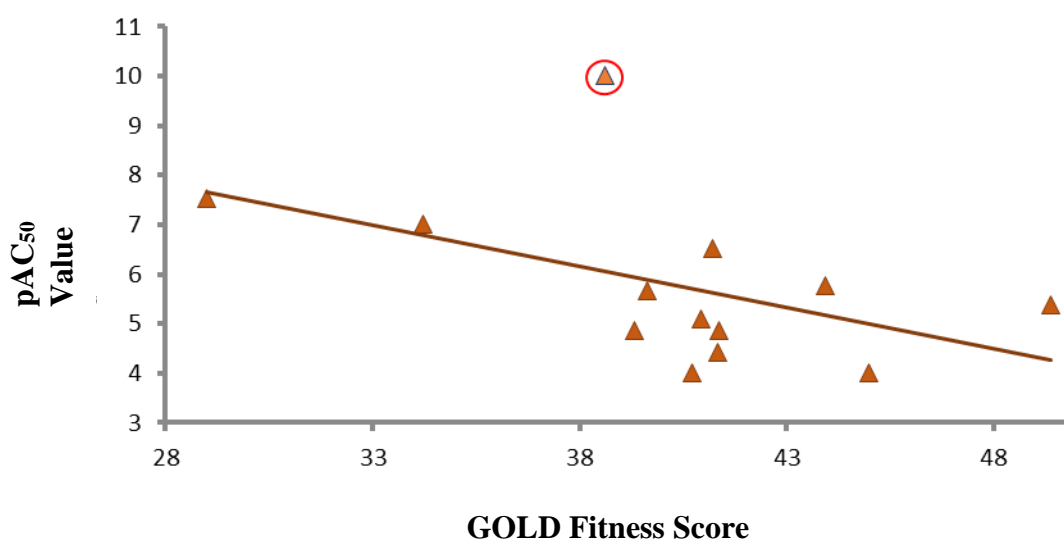


Figure 4.8: Correlation plot between pAC₅₀ values and Gold fitness score (0.29). The circle marked triangle indicated the highly active compound (0.1nM) with a pAC₅₀ value of 10 and GOLD score of 38.59.

4.2.2 Molecular Dynamics Simulation

Molecular dynamics simulation of highly active docked ligand was done to explore the binding stability and conformational flexibility of CREB-ligand complexes. For this purpose, CBP was removed from the docked complexes, and MD simulation was performed by using Desmond SCHRODINGER software. The simulation job for highly active ligand (LIG_1) was executed for 400nsec. The resultant trajectory of CaRMSD (eq3) [101] showed the fluctuations between the CREB-ligand complexes for a particular time but get stable towards the end.

C α RMSD measured the average deviation in atomic disruption of the Carbon- α backbone with simulation time frame. It was measured by

$$C\alpha RMSD = \sqrt{\frac{1}{N} \sum_{i=1}^N (r'(t_x) - r'(t_{ref}))^2} \dots \dots \dots Eq3$$

The RMSD (A $^\circ$) v/s time (nsec) showed the unstable behavior of both ligand and CREB at the start. The RMSD elevated to 9 A $^\circ$ for CREB and 54 A $^\circ$ for ligand in the first 100nsec. However, around time steps of 200nsec, CREB and ligand showed a drastic decline in RMSD to 6 A $^\circ$ and 10 A $^\circ$, respectively. After 220nsec, both CREB and ligand again showed an elevated behavior till 270nsec. At 300nsec, the complex demonstrated stable behavior with a minimum deviation of 4.5A $^\circ$ (for CREB) and 12A $^\circ$ (for ligand). This pattern remained stable for a time frame ($r'(t_x)$) of 100nsec further, declaring the premium stability of the docked structure.

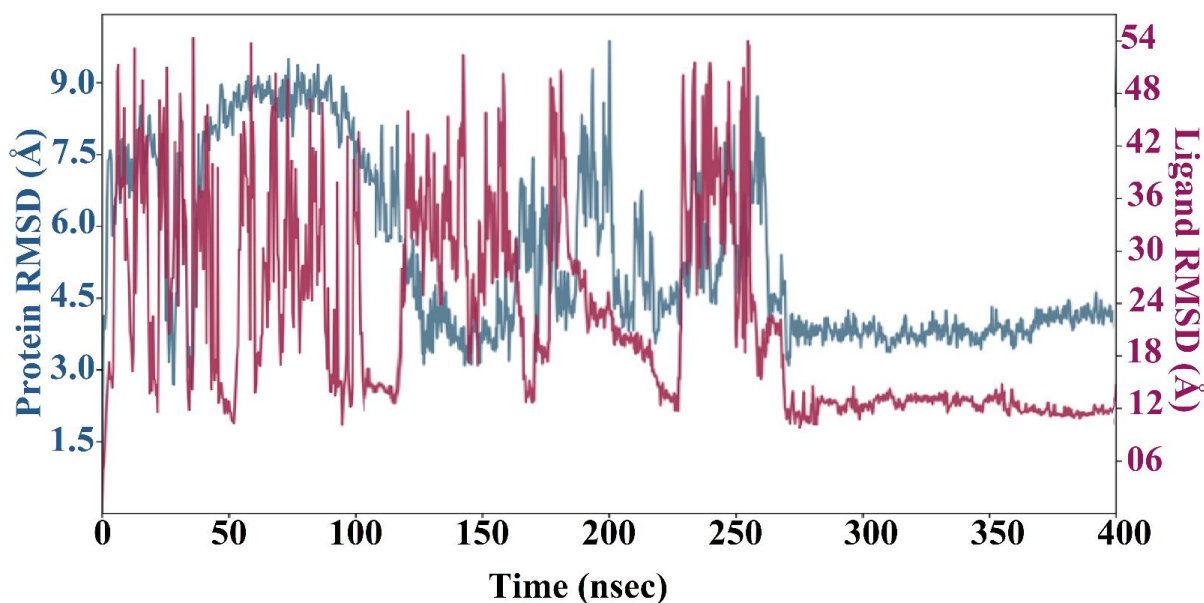


Figure 4.9. Molecular dynamics simulation of the highly active compound (LIG_1). The graph shows the C α RMSD on the y-axis during the time frame of 400nsec on the x-axis. The simulation remained inconsistent for the first 200nsec and showed a stable behavior for the last 100nsec, suggesting the stability of the model.

After analyzing the stability of the protein, the interactions pattern was monitored. For this purpose, Protein-ligand interacting contacts for **LIG_1/CREB** complex were visualized throughout the simulation time. Figure 4.10 explains the stacked bar graph over the progression of trajectory. SEP133 with an interaction fraction of 0.7 showed

strong H-bond interaction (IF), suggested that this interaction remained stable for 70% of the simulation time. The H-bond was demonstrated of Donor-hydrogen-acceptor-atom (D-H...A) type and found between terminal –H of the ligand and –O atom in the side chain of SEP133. Similarly, other residues such as ARG131, TYR134, and ILE137 showed Hydrophobic-IF between the ranges of 0.2-0.3 declared stability for 20%-30% of simulation time. Here, the π -H-bond was formed between the benzene ring of the ligand with the H-atom of ARG131 and ILE137. However, TYR134 was responsible for making a strong π - π bond via benzene ring of the ligand.

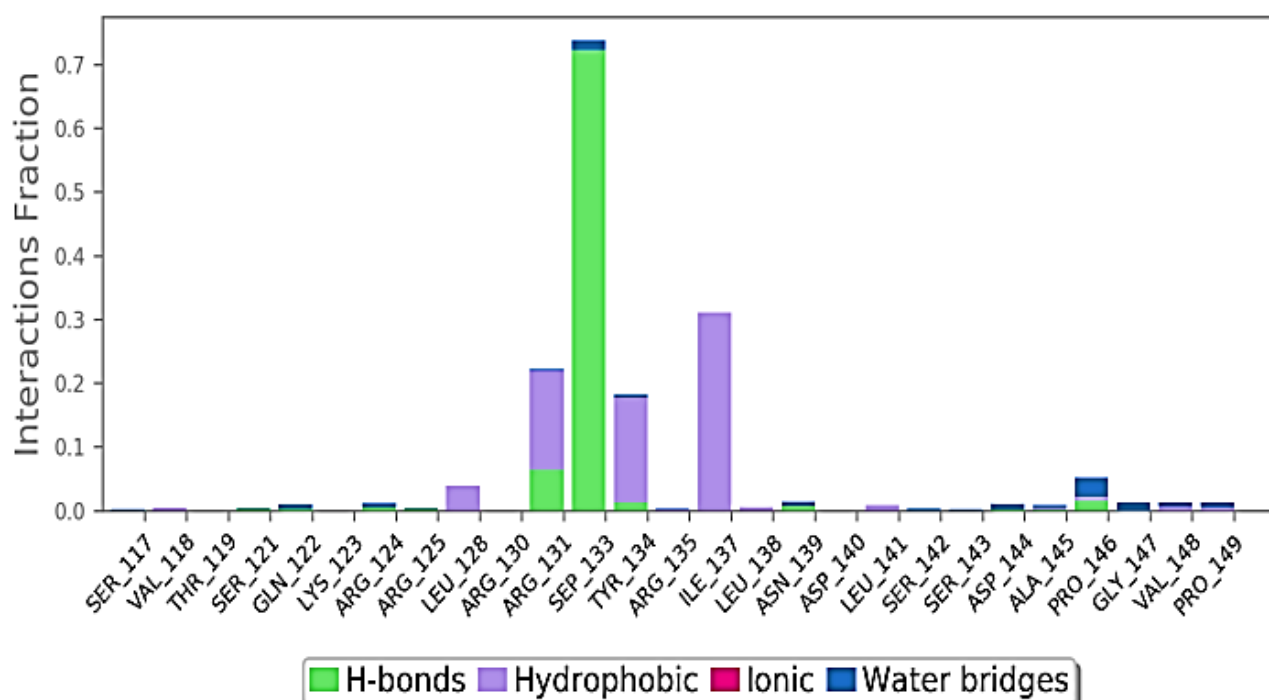


Figure 4.10: Protein-Ligand interactions of LIG_1/CREB complex. The y-axis shows the interaction fraction, while residues of CREB demonstrate on x-axis. The residue SEP133 showed most stable H-bond interaction throughout the simulation trajectory. While ARG131, TYR134, and ILE137 demonstrated hydrophobic interactions for 20-30% of the simulation time.

By comparing the interacting pattern before and after molecular dynamics simulations of LIG_1/CREB complex, it is revealed that the ligand translocated its position during MD simulation when an external force field was (OPLS4) applied. The interacting residues during the docking studies were SER121, ARG125, and ARG135, stabilized to the actual binding cavity of CREB. The resultant stable interacting pattern was observed among ARG131, SEP133, TYR134, and ILE137 residues, which were the reported

residues in the literature (figure 4.11 (a) and (b)). The ligand after MD achieved its stable conformation and persisted for 100nsec time step at the end. Here, the ligands atom remained enact with the binding cavity residues and showed an entropic role in the binding interaction. After MD, the final ligand structure was subjected to 3D feature extraction based on its new stable 3D conformation.

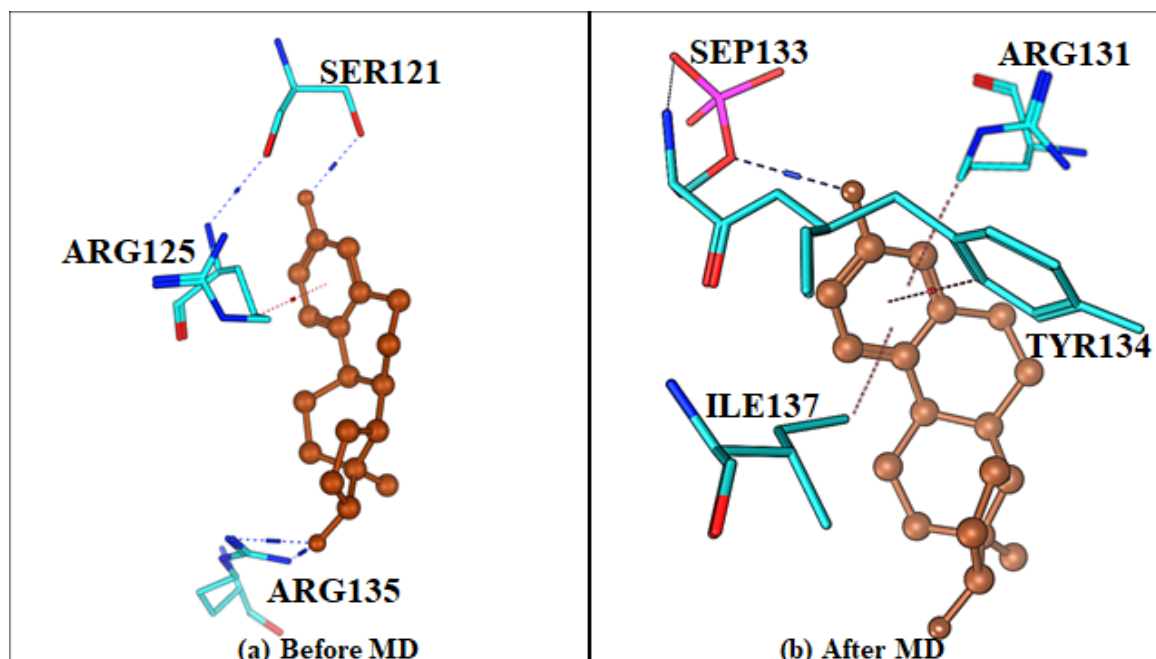


Figure 4.11 Interacting residues pattern before and after MD. The brown-colored structure represents the ligand molecule. The sky-blue color structure shows the CREB resides. The dotted blue colored line represents the H-bond interactions, while maroon dotted lines represent the hydrophobic interactions between the ligand and protein. (a)Before MD, the binding residues were SEP121, ARG125, and ARG135. (b) The ligand translocated its position to ARG131, SEP133, TYR134, and ILE137 after MD.

4.2.3 Ligand Based Pharmacophore

After MD, the stable conformational ligand was removed from the complex and used as a template for pharmacoinformatic guided feature extraction protocol. For making a 3D predictive model, the ligand was loaded in the pharmacophore query of MOE. After iterative selection of the ligand's functional group, pharmacophore was built by combining unique features of the ligand. The template had 14 combinations of features which were subdivided into Hydrophobic (Hyd), H-bond Acceptor (Acc),

H-bond Donor (Don), Donor and Acceptor (Don&Acc), and Aromatic (Aro) (figure 4.12).

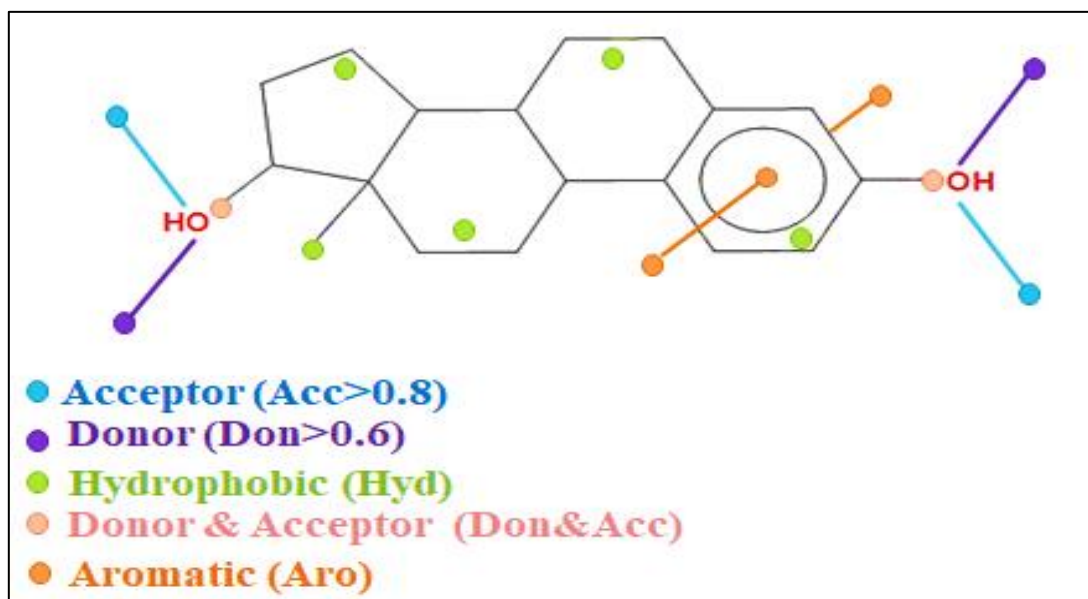


Figure 4.12: Combination of 14 feature sites found in the template during pharm_LIG_1/CREB pharmacophore model generation.

Multiple pharmacophore models were generated using these properties. The pharmacophore model (**pharm_LIG_1/CREB**) was built with four features sites, including 2 Acceptor (Acc) and 2 Hydrophobic (Hyd). The radius for each acceptor site was maintained at 1.9 Å while it was 0.9 Å for the hydrophobic feature (figure 4.13). Table 4.3 indicates the features of the pharmacophore model along with their respected radius. The selected features were similar to the sites of interactions during MD simulations with CREB residues (SEP133, ARG131, TYR134, and ILE137). This behavior strengthened the stability of the ligand and interacting features to be promising.

Table 4.3: Crucial pharmacophore features of the finally selected model with respective radiuses and mutual distances.

	F1_Accp	F2_Accp	F3_Hyd	F4_Hyd	Radius for model building
F1_Accp	-----	10.78	7.70	5.81	1.9Å°
F2_Accp	10.78	-----	6.78	11.20	1.9Å°
F3_Hyd	7.70	6.78	-----	5.01	0.9Å°
F4_Hyd	5.81	11.20	5.01	-----	0.9Å°

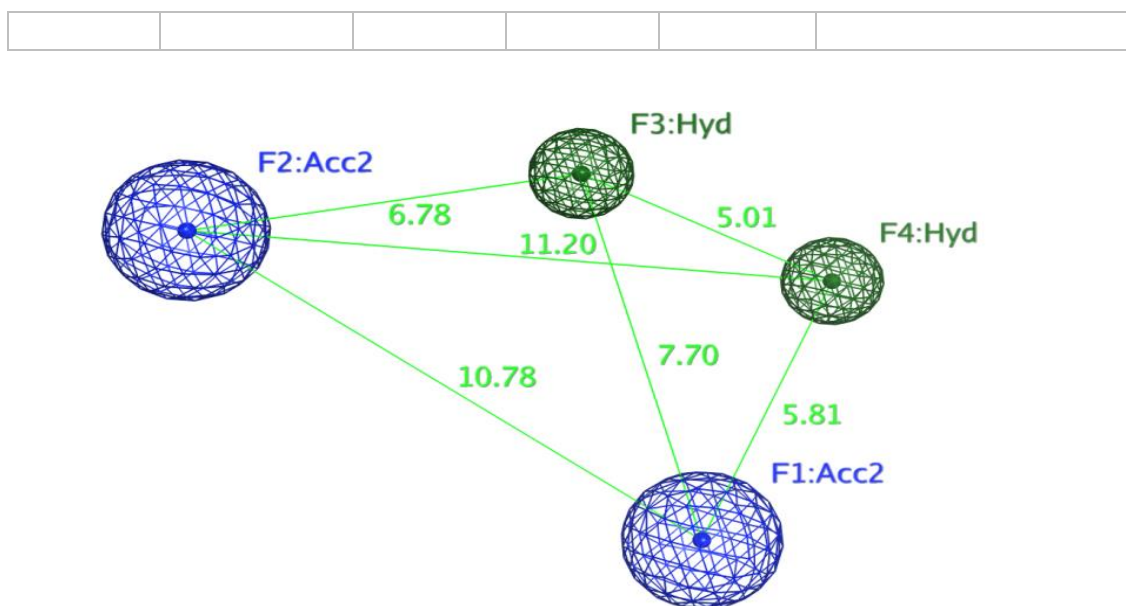


Figure 4.13: In the pharm_LIG_1/CREB pharmacophore model, the blue-colored sphere represents the Hydrogen bond acceptor while green colored sphere represents the hydrophobic features.

The pharmacophore model pharm_LIG_1/CREB, was then screened from the rest of the docked dataset for internal test validation. The model was able to classify the ligands as TP = 4, TN = 7, FP = 0 and FN = 1. The statistics of the classification were then evaluated via a confusion matrix. The resultant of the matrix showed an accuracy of 92%, asserting that the pharmacophore model was able to predict and classify the ligands as precisely (actives as actives, inactive as inactive). However, the value of precision and specificity or True Negative Rate (TNR), i.e., 1.000, declared that the model is specified. Similarly, the model's sensitivity or True Positive Rate (TPR) (0.8) indicated that the model is sensitive for all features associated with actives compounds (TP). The specificity (TNR) of 1.000 declared that the given feature predicted inactive more precisely than actives (table 4.4).

Table 4.4: Statistical evaluation of the pharmacophore model w.r.t internal dataset

Statistical approach	Implementation of results	Model Evaluation
Accuracy	$= 4+7/4+7+0+1$	$0.92 = 92\%$
Sensitivity	$= 4/4+1$	0.800
Specificity	$= 7/7+0$	1.000

Precision rate	= 4/4+0	1.000
----------------	---------	-------

Another statistical parameter was used to measure the predictive ability of the model. The Matthews Correlation Coefficient (MCC) measures the quality of agreement between predicted and actual values. It is a correlation coefficient that takes all variables (TP, TN, FP, and FN) and gives scores based on model classification. The value near to or closer to 1 indicates the best agreement between predicted and actual responses. The MCC value (calculated via eq 4) [102] for the pharmacophore model was 0.85, which indicated that the selected template and all true positive compounds with respective features have the ability to induce CREB response. Therefore, the model was further processed for virtual screening for the validation of external datasets.

$$MCC = \frac{TP.TN - FP.FN}{\sqrt{(TP+FP).(TP+FN).(TN+FP).(TN+FN)}} \dots\dots\dots Eq4$$

4.2.4 Virtual Screening

The evaluated model was then screened against the three external datasets. For this step, datasets from publicly accessible databases were retrieved and preprocessed. The first and second datasets consisted of chemical data from the Drug Bank Database (10631), ZINC Database (885) and while the third one constituted of FDA-approved drugs (48) of Epilepsy and Spinal muscular atrophy (EPI/SMA Database) (Appendix 3). To evaluate the metabolic profiles of all chemical entities, the substrate from all databases was passed through CYP Model. The selected entries from the CYPs, i.e., 201 from Drug Bank Database, 303 from the ZINC database, and 9 from EPI/SMA Database, were identified as non-inhibitors of the CYPs. These compounds were then screened against the in-house hERG pharmacophore model to avoid cardiotoxicity. At the end of the query, 185 (Drug Bank database), 177 (ZINC database), and 6 (EPI/SMA database) compounds were identified as non-inhibitors of hERG.

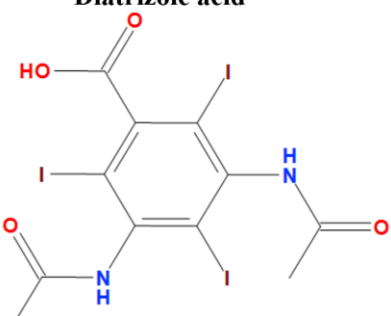
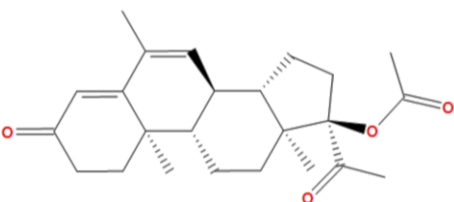
After ADMET toxicity analysis, the selected compounds were screened against the **pharm_LIG_1/CREB** pharmacophore model. The resultant hits demonstrated that 106 Drug Bank database entities, 113 ZINC database compounds, and 2

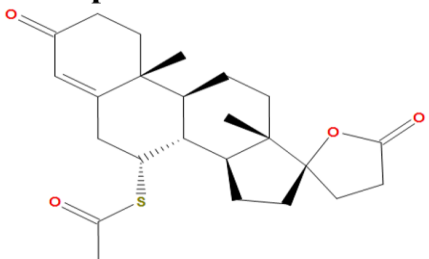
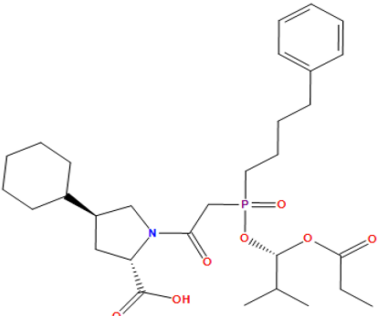
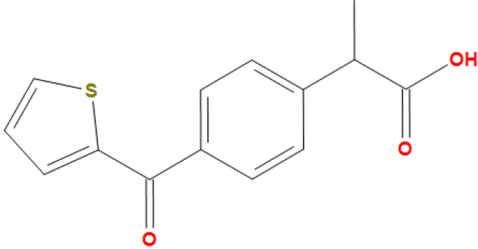
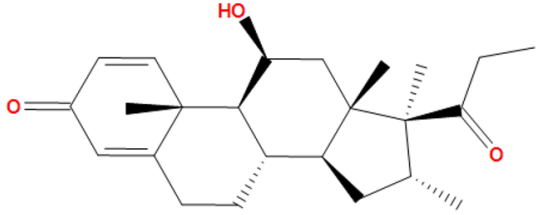
EPI/SMA database entries are potential entities. These identified compounds could be effective and safer against SMA-PME and, therefore, can be utilized further for lead optimization protocol.

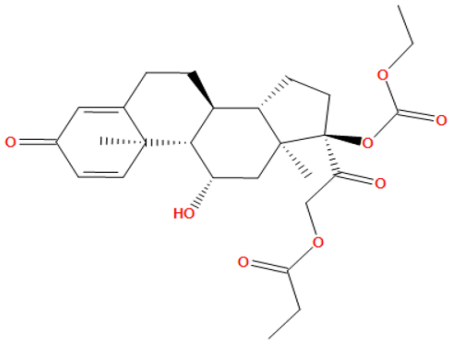
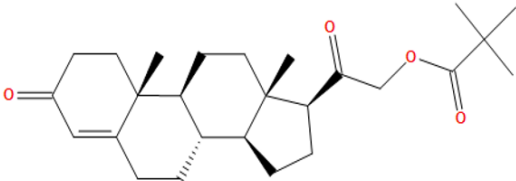
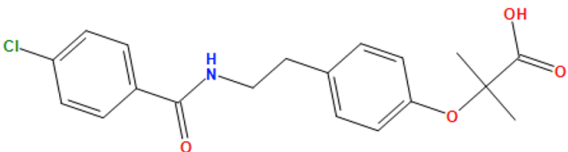
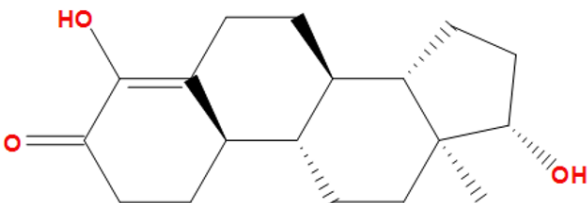
4.2.5 Biological Evaluation

To validate results from virtual screening, the biological assessment of the hits was made. For this purpose, compounds selected from the Drug Bank database were investigated in the literature. The possibility of these entities as FDA-approved drugs for the treatment of other diseases was explored. About 11 hits were used as medicinal products in many treatment procedures (table 4.5). Similarly, selected compounds with a unique ZINC id were then explored for their natural origin. The possibility of secondary metabolites and their derivatives was studied. Total 7 compounds possessed botanical, pharmacological, and chemical information (table 4.6). In the end, information for epilepsy or SMA based drugs (2) were explored (table 4.7). The compounds that remain unidentified are in the clinical phase, and no biological information was available at the moment.

Table 4.5. 2D structures of hits obtained after virtual screening by Drug Bank database for repurposing against CREB protein.

Chemical Compound	Properties
<p style="text-align: center;">Diatrizoic acid</p>  <p style="text-align: center;">Molecular Weight = 613.91 g/mol</p>	<ul style="list-style-type: none"> * An organic, iodinated compound. * Shows radiopaque based X-ray contrast activity.
<p style="text-align: center;">Megestrol acetate</p>  <p style="text-align: center;">Molecular Weight = 384.5 g/mol</p>	<ul style="list-style-type: none"> * Acetate salt of megestrol * Derivative of progesterone, causing inhibition of LH hormone. * Use for the treatment of anorexia and cachexia. * Side effects include weight gain, trouble in sleeping and stomach upset.

<p style="text-align: center;">Spironolactone</p>  <p style="text-align: center;">Molecular Weight = 416.6 g/mol</p>	<ul style="list-style-type: none"> * It is a synthetic 17-spironolactone corticosteroid. * Perform antihypertensive, diuretic and antiandrogen activities. * Side effects include vomiting, dizziness and stomach upset.
<p style="text-align: center;">Fosinopril</p>  <p style="text-align: center;">Molecular Weight = 563.7 g/mol</p>	<ul style="list-style-type: none"> * Act as ACE inhibitor. * Diminish vasopressin activity. * Side effects include joint pain, rashes, itching, diarrhea, vomiting, and elevated heartbeat.
<p style="text-align: center;">Suprofen</p>  <p style="text-align: center;">Molecular Weight = 260.3 g/mol</p>	<ul style="list-style-type: none"> * Act as non-steroidal anti-inflammatory drug, antirheumatic drug and a non-narcotic analgesic. * Side effects include vomiting, sensitivity to light, eye burning, and itching.
<p style="text-align: center;">Rimexolone</p>  <p style="text-align: center;">Molecular Weight = 370.5 g/mol</p>	<ul style="list-style-type: none"> * Act as a synthetic glucocorticoid. * Possesses immunosuppressive and anti-inflammatory activity. * Side effects include eye sensitivity and burning.

<p style="text-align: center;">Prednicarbate</p>  <p style="text-align: center;">Molecular Weight = 488.6 g/mol</p>	<ul style="list-style-type: none"> * Synthetic non-halogenated double-ester derivative. * Possesses antipruritic, anti-inflammatory, and vasoconstrictive properties. * Side effects include joint pain, rashes, vomiting, and elevated heartbeat.
<p style="text-align: center;">Cianidanol</p>  <p style="text-align: center;">Molecular Weight = 290.27 g/mol</p>	<ul style="list-style-type: none"> * A plant derivative consists of flavonoids. * Use to deal with chronic osteoarthritis. * Side effects include elevated heartbeat, tremor, anxiety, and shortness of breath.
<p style="text-align: center;">Bezafibrate</p>  <p style="text-align: center;">Molecular Weight = 361.8 g/mol</p>	<ul style="list-style-type: none"> * Agonist of PPAR-alpha. * Role in antilipidemic activity. * Side effects include joint pain, vomiting, and stomach upset.
<p style="text-align: center;">Oxabolone</p>  <p style="text-align: center;">Molecular Weight = 290.4 g/mol</p>	<ul style="list-style-type: none"> * DEA Schedule III controlled substance. * Side effects include joint pain, rashes, and elevated heartbeat.

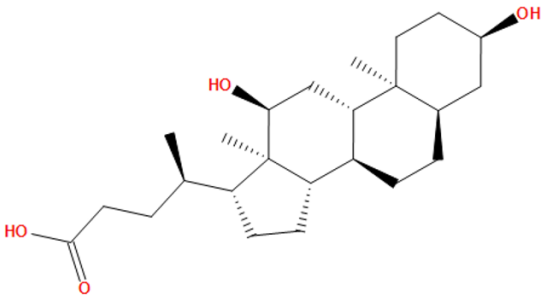
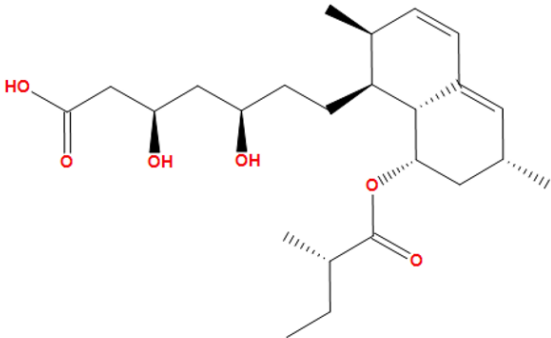
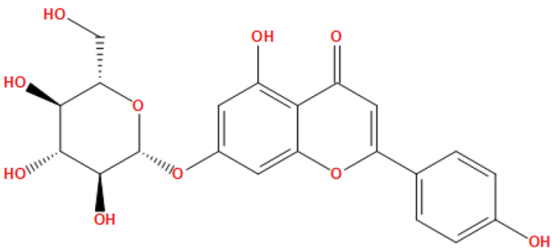
<p style="text-align: center;">Deoxycholic acid</p>  <p style="text-align: center;">Molecular Weight = 392.6 g/mol</p>	<ul style="list-style-type: none"> * Bile acid; act as a human blood serum metabolite. * Conjugate of deoxycholate acid. * Side effects include joint pain, urination problem and drowsiness.
<p style="text-align: center;">Lovastatin acid</p>  <p style="text-align: center;">Molecular Weight = 422 g/mol</p>	<ul style="list-style-type: none"> * Act as Aspergillus metabolite, Hydroxymethylglutaryl-CoA reductase inhibitor. * Side effects include elevated heartbeat, dry mouth, vomiting and constipation.

Table 4.6. 2D structures of hits after =virtual screening by ZINC database with their natural origin.

Zinc ID	Chemical Compound	Properties
ZINC03871576	<p style="text-align: center;">Apigenin7-beta-L-glucoside</p>  <p style="text-align: center;">Molecular Weight = 432.4 g/mol</p>	<ul style="list-style-type: none"> * Derivative of Apigenin. * Act as anti-inflammatory and antioxidant. * Found in onions, parsley, tea, oranges, wheat sprouts, and chamomile.

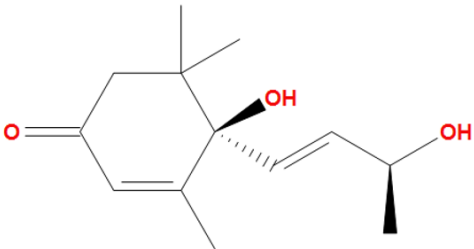
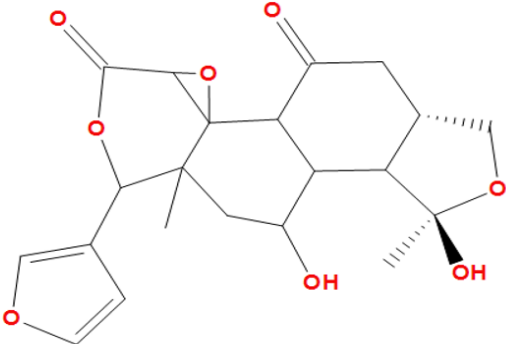
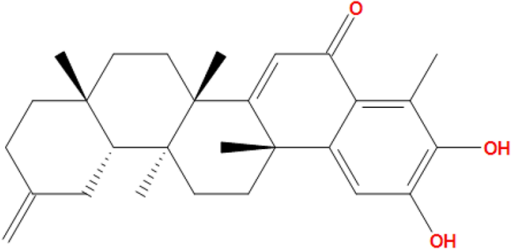
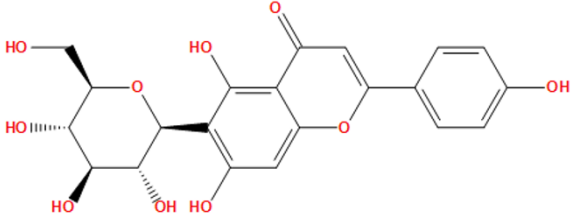
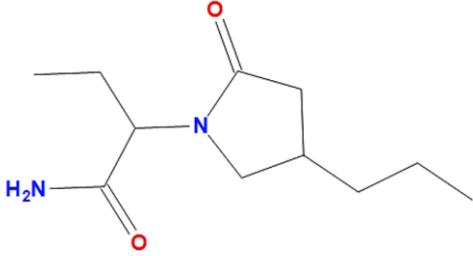
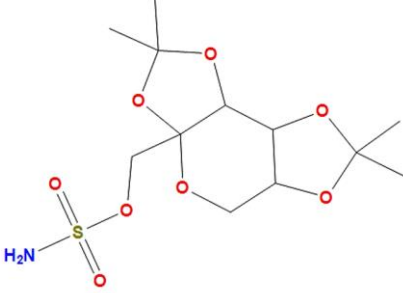
ZINC04096339	<p style="text-align: center;">Vomifoliol</p>  <p style="text-align: center;">Molecular Weight = 224.3 g/mol</p>	<ul style="list-style-type: none"> * Derivative of Sesquiterpenoids. * Found in grapes.
ZINC95486237	<p style="text-align: center;">Clausenolide</p>  <p style="text-align: center;">Molecular Weight = 404 g/mol</p>	<ul style="list-style-type: none"> * Natural racemic pyrrolidone compounds. * Found in leaves of Clausena lansium (Lour).
ZINC95486223	<p style="text-align: center;">6-Oxoisoiguesterin</p>  <p style="text-align: center;">Molecular Weight = 420 g/mol</p>	<ul style="list-style-type: none"> * Natural terpenoids. * Extracted from leaves of Quinone.
ZINC95486143	<p style="text-align: center;">Isovitexin</p>  <p style="text-align: center;">Molecular Weight = 432.4 g/mol</p>	<ul style="list-style-type: none"> * Derivative of apigenin. * Extracted from onions, grapes, and wheat. * Act as alpha-glucosidase inhibitor

Table 4.7. 2D structures of hits after virtual screening by FDA-approved EPI/SMA drug database for repurposing against CREB protein.

Molecular ID	Compound	Properties
EPI_1	<p style="text-align: center;">Brivaracetam</p>  <p style="text-align: center;">Molecular Weight = 212.29 g/mol</p>	<ul style="list-style-type: none"> * Orally available drug. * Derivative of Levetiracetam. * Binds to synaptic vesicle protein 2A in Central nervous system. * Role in inducing action potential in brain. * Use as anti-epileptic and anti-convulsant. * Side effects include diarrhea, stomach upset and vomiting.
EPI_36	<p style="text-align: center;">Topiramate</p>  <p style="text-align: center;">Molecular Weight = 339.36 g/mol</p>	<ul style="list-style-type: none"> * Promote post-synaptic response in neurons. * Attenuate GABA inhibition. * Use as anti-convulsant. * Side effects include constipation, nausea, shortness of breath and anxiety.

CHAPTER 5

DISCUSSION

In SMA-PME, acid ceramidase (aCDase) deficiency is responsible for progressive neuromuscular dysfunctionality [103]. Recent research highlighted recombinant replacement therapy as a potent option to overcome the deficit of aCDase [104]. Moreover, some cancer studies have also observed the induction of ASAHI gene expression (aCDase production) via CREB [105, 106].

Herein, sequences of recombinant aCDase (r-aCDase) and their predictive 3D structural models have been assembled and constructed, respectively. As the crystal structure of aCDase (5u7z) was published, homology modeling of the tag was performed. During the template identification step, it was observed that the structure of some tags like the SUMO tag and HIS tag was unavailable. However, the template for the GST tag was retrieved, and sequence alignment was executed. The model was built on this template with the highest accuracy of the residual match [107]. Molecular dynamic simulations further stabilized the validated molecular 3D structure after 50nsec. The consequent 3D structure possessed 149 residues in the allowed region (i.e., no steric clashes) and 3 in the disallowed region due to hindrance of backbone chain and residual atoms [108]. The GST tag model was then docked with aCDase via Cluspro. The resultant structure lost its ability to perform function due to a large loop, as described by Anthena D [109]. This limitation was overcome by de novo modeling of r-aCDase via trRosetta. All the sequences were submitted, and based on residual prediction; their respective 3D models were constructed. The TM value of the built model showed the confidence score of the prediction. The confidence score ranged from zero (poor) to 1 (best) declared structural stability, and a higher value indicated the better prediction of the model [95]. These models were substantiated after energy minimization by ERRAT and Ramachandran plots. The statistics showed the stability of the model even with some residues in the outlier region [110]. The energy minimized structures were then subjected to MD simulations. The RMSD and RMSF plots evaluated the stability of the protein and its respected residues over time. Based on the least deviated value of these factors [111], seq7, seq3, and seq4 were stated as the stable r-aCDase. Here it can be proposed that these structures are energetically stable and could be effective when used in recombinant therapy for further trials.

For the second module, molecular docking of CBP/CREB and 13 inducers proposed a weak positive correlation between GOLD score and pAC₅₀ value, suggesting an unstable binding pattern [100]. This abnormal behavior can be

attributed to the presence of CBP during the docking protocol. This hypothesis was further strengthened by an observed change in the binding cavity during the procedure. The same pattern has been observed in previous findings where the translocation of a ligand occurred due to conformational entropic change [112, 113]. Therefore, highly active conformer was deployed to MD simulation to stabilize the interacting complex. Since previous studies advocated for the structure's reliability, stability, and accuracy after MD [114] thus, the ligand retrieved after MD projected towards feature extraction. It was achieved by mapping the inter feature distance and atomic conformation. Previously, Saxena et al. proposed the pharmacophore model for the estradiol. Interestingly the pharmacophore generated after MD overlapped with the reported model's feature [115]. The generated model presented two H-bond Acceptors with two hydrophobic features. Moreover, the pharmacophore model also probed the stable interacting site after MD (ARG131, SEP133, TYR134, and ILE137), thus confirming the authenticity of the featured model. These retrieved features were then virtually screened against FDA-approved drugs from different databases. Considering the pharmacokinetic properties, the compounds of these databases were assessed based on ADMET property [116]. The compounds from the filters like CYPs and hERG (anti-target candidates) were expected to be non-carcinogen and non-toxic, thus reliable as potential inducers against the targeted protein [117]. The screened entities were then subjected to virtual screening using built pharmacophore model. The resultant hits then biologically validated and showed potential role as anti-inflammatory, anti-seizures, and neuroprotective compounds. However, the general side effects include nausea, elevated heartbeat and dizziness. These hits after robust in-silico validation were proposed to be effective against CREB and could be utilized further in the lead optimization step.

CHAPTER 6
CONCLUSION

In conclusion, the present study was designed with the therapeutic interventions of restoring aCDase level in the body and upregulation of ASAH1 gene expression. Keeping this goal in mind, the project was divided into two modules. The first module comprised of retrieving of r-aCDase sequence data and construction of 3D models. The stable models achieved at the end of tag's MD simulation were proposed to be effective and safer when used in recombinant therapy. For the second part of the project, CREB was used to modulate the normal physiological function of ASAH1 gene. It was accomplished by performing molecular docking to elucidate the binding hypothesis between CREB with its inducers. Due to the presence of CBP during the protocol, unstable binding patterns were attained. Based on activity value, a highly active ligand was selected and subjected to MD simulation to accomplished stable interaction and conformer. The stable residues in the actual binding cavity demonstrating the energy contours in the virtual receptor site were ARG131, SEP133, TYR134, and ILE137, respectively. The conformer was then used as a template for the feature extraction and predictive model formation. The bioactive pharmacophore exhibits two H-bond acceptors and two hydrophobic features. With respective features, the statistical evaluation was made against the rest of the ligand dataset. The accuracy of 92% declared the significance of the model. The virtual screening of the validated model was accomplished via an external dataset. The datasets were filtered for toxicity profiling against hERG and CYPs. The filtered compounds were then screened against the built pharmacophore model, and hits were recognized as the potential target for lead optimization protocol. During biological evaluation step, these hits were recognized as neuroprotective, anti- seizures and anti-inflammatory compounds with minimal side effects (dizziness, nausea, elevated heartbeat and sweating). Overall, the presented project gives a new in-silico dimension in the therapeutics of SMA-PME and will provide insightful measures in rational drug designing in the future.

CHAPTER 7
REFERENCES

References:

1. Gan, J.J., et al., *Acid ceramidase deficiency associated with spinal muscular atrophy with progressive myoclonic epilepsy*. *Neuromuscular Disorders*, 2015. **25**(12): p. 959-963.
2. Topaloglu, H. and J. Melki, *Spinal muscular atrophy associated with progressive myoclonus epilepsy*. *Epileptic Disorders*, 2016. **18**(s2): p. S128-S134.
3. Dymont, D., et al., *Evidence for clinical, genetic and biochemical variability in spinal muscular atrophy with progressive myoclonic epilepsy*. *Clinical genetics*, 2014. **86**(6): p. 558-563.
4. He, X., et al., *Enzyme replacement therapy for Farber disease: Proof-of-concept studies in cells and mice*. *BBA clinical*, 2017. **7**: p. 85-96.
5. Jankovic, J. and V. Rivera, *Hereditary myoclonus and progressive muscular atrophy: a new syndrome*. *Transactions of the American Neurological Association*, 1978. **103**: p. 116-118.
6. Radhakrishnan, D.M., et al., *Spinal Muscular Atrophy and Progressive Myoclonic Epilepsy: A Rare Association*. *Journal of Neurosciences in Rural Practice*, 2021. **12**(01): p. 210-212.
7. Elsea, S.H., et al., *ASAH1 pathogenic variants associated with acid ceramidase deficiency (ACD): Farber disease and spinal muscular atrophy with progressive myoclonic epilepsy (SMA-PME)*. *Human Mutation*, 2020.
8. Fabian, P., et al., *Acid ceramidase deficiency: Farber disease and SMA-PME*. *Orphanet journal of rare diseases*, 2018. **13**(1): p. 121.
9. Gill, R.W. and P. Sanseau, *Rapid in silico cloning of genes using expressed sequence tags (ESTs)*. 2000.
10. Frohbergh, M., X. He, and E.H. Schuchman, *The molecular medicine of acid ceramidase*. *Biological chemistry*, 2015. **396**(6-7): p. 759-765.
11. Leclerc, J., et al., *Lysosomal acid ceramidase ASAH1 controls the transition between invasive and proliferative phenotype in melanoma cells*. *Oncogene*, 2019. **38**(8): p. 1282-1295.
12. Park, J.-H. and E.H. Schuchman, *Acid ceramidase and human disease*. *Biochimica et Biophysica Acta (BBA)-Biomembranes*, 2006. **1758**(12): p. 2133-2138.
13. Shtraizent, N., et al., *Autoproteolytic cleavage and activation of human acid ceramidase*. *Journal of Biological Chemistry*, 2008. **283**(17): p. 11253-11259.
14. Gebai, A., et al., *Structural basis for the activation of acid ceramidase*. *Nature communications*, 2018. **9**(1): p. 1-11.
15. Mahmoud, I.G., et al., *ASAH1-related disorders: Description of 15 novel pediatric patients and expansion of the clinical phenotype*. *Clinical Genetics*, 2020. **98**(6): p. 598-605.
16. Kono, M., et al., *Neutral ceramidase encoded by the Asah2 gene is essential for the intestinal degradation of sphingolipids*. *Journal of Biological Chemistry*, 2006. **281**(11): p. 7324-7331.
17. Dementiev, A., et al., *Molecular mechanism of inhibition of acid ceramidase by carmofur*. *Journal of medicinal chemistry*, 2018. **62**(2): p. 987-992.
18. Falluel-Morel, A., et al., *Interactions of PACAP and ceramides in the control of granule cell apoptosis during cerebellar development*. *Journal of molecular neuroscience*, 2008. **36**(1): p. 8-15.
19. Dymont, D.A., et al., *ASAH1-related disorders*, in *GeneReviews®[Internet]*. 2018, University of Washington, Seattle.
20. Jankovic, J. and V.M. Rivera, *Hereditary myoclonus and progressive distal muscular atrophy*. *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society*, 1979. **6**(3): p. 227-231.

21. Moser, H., A. Moser, and W. Chen, *The metabolic basis of inherited disease*. Ceramidase deficiency, 1989.
22. Bernardo, K., et al., *Purification, characterization, and biosynthesis of human acid ceramidase*. Journal of Biological Chemistry, 1995. **270**(19): p. 11098-11102.
23. Koch, J., et al., *Molecular cloning and characterization of a full-length complementary DNA encoding human acid ceramidase: identification of the first molecular lesion causing Farber disease*. Journal of Biological Chemistry, 1996. **271**(51): p. 33110-33115.
24. Li, C.-M., et al., *The human acid ceramidase gene (ASAH): structure, chromosomal location, mutation analysis, and expression*. Genomics, 1999. **62**(2): p. 223-231.
25. He, X., et al., *Purification and characterization of recombinant, human acid ceramidase: catalytic reactions and interactions with acid sphingomyelinase*. Journal of Biological Chemistry, 2003. **278**(35): p. 32978-32986.
26. Zhou, J., et al., *Spinal muscular atrophy associated with progressive myoclonic epilepsy is caused by mutations in ASAH1*. The American Journal of Human Genetics, 2012. **91**(1): p. 5-14.
27. Filosto, M., et al., *ASAH1 variant causing a mild SMA phenotype with no myoclonic epilepsy: a clinical, biochemical and molecular study*. European Journal of Human Genetics, 2016. **24**(11): p. 1578-1583.
28. Fabian, P., et al., *Acid ceramidase deficiency: Farber disease and SMA-PME*. Orphanet journal of rare diseases, 2018. **13**(1): p. 1-19.
29. van der Beek, N.A., et al., *A new case of SMA phenotype without epilepsy due to biallelic variants in ASAH1*. European Journal of Human Genetics, 2019. **27**(3): p. 337-339.
30. Kyriakou, K., et al., *Acid Ceramidase Depletion Impairs Neuronal Survival and Induces Morphological Defects in Neurites Associated with Altered Gene Transcription and Sphingolipid Content*. International journal of molecular sciences, 2020. **21**(5): p. 1607.
31. Delghandi, M.P., M. Johannessen, and U. Moens, *The cAMP signalling pathway activates CREB through PKA, p38 and MSK1 in NIH 3T3 cells*. Cellular signalling, 2005. **17**(11): p. 1343-1351.
32. Lucki, N. and M.B. Sewer, *The cAMP-responsive element binding protein (CREB) regulates the expression of acid ceramidase (ASAH1) in H295R human adrenocortical cells*. Biochimica et Biophysica Acta (BBA)-Molecular and Cell Biology of Lipids, 2009. **1791**(8): p. 706-713.
33. Cho, S.M. and H.J. Kwon, *Acid ceramidase, an emerging target for anti-cancer and anti-angiogenesis*. Archives of pharmacal research, 2019. **42**(3): p. 232-243.
34. Muñoz, Ú., et al., *Hepatocyte growth factor enhances alternative splicing of the Krüppel-like factor 6 (KLF6) tumor suppressor to promote growth through SRSF1*. Molecular Cancer Research, 2012. **10**(9): p. 1216-1227.
35. Zhang, J., et al., *Pathogenic gene variants and clinical phenotype features of 26 children with progressive myoclonic epilepsy*. Zhonghua er ke za zhi= Chinese journal of pediatrics, 2019. **57**(6): p. 458-464.
36. He, X., et al., *Purification and Characterization of Recombinant, Human Acid Ceramidase CATALYTIC REACTIONS AND INTERACTIONS WITH ACID SPHINGOMYELINASE*. Journal of Biological Chemistry, 2003. **278**(35): p. 32978-32986.
37. Sikora, J., et al., *Acid ceramidase deficiency in mice results in a broad range of central nervous system abnormalities*. The American journal of pathology, 2017. **187**(4): p. 864-883.

38. Beck, M., H.W. Moser, and K. Sandhoff, *Acid ceramidase deficiency: Farber lipogranulomatosis, spinal muscular atrophy associated with progressive myoclonic epilepsy and peripheral osteolysis*, in *Rosenberg's Molecular and Genetic Basis of Neurological and Psychiatric Disease*. 2020, Elsevier. p. 547-557.
39. Badv, R.S., et al., *A novel case report of spinal muscular atrophy with progressive myoclonic epilepsy from Iran*. *International medical case reports journal*, 2019. **12**: p. 155.
40. Kernohan, K.D., et al., *Whole-transcriptome sequencing in blood provides a diagnosis of spinal muscular atrophy with progressive myoclonic epilepsy*. *Human mutation*, 2017. **38**(6): p. 611-614.
41. Rajput, V.B., *Comparative structural and functional studies of acid ceramidase and acid ceramidase-like proteins using biophysical and computational techniques*. 2019.
42. Bashyam, M.D., et al., *Molecular analyses of novel ASAH1 mutations causing Farber lipogranulomatosis: analyses of exonic splicing enhancer inactivating mutation*. *Clinical genetics*, 2014. **86**(6): p. 530-538.
43. Alayoubi, A.M., et al., *Systemic ceramide accumulation leads to severe and varied pathological consequences*. *EMBO molecular medicine*, 2013. **5**(6): p. 827-842.
44. Michelena, T.M., et al., *The impact on the activity of acetylcholinesterase of a polylysine-ApoE peptide carrier targeting the blood brain barrier*. *Fundamental Toxicological Sciences*, 2018. **5**(4): p. 123-126.
45. Mitchell, J., C. Etchart, and F. El Turk, *Expanding the phenotype: Acid ceramidase deficiency presenting with features of SMA-PME and Farber disease*. *Molecular Genetics and Metabolism*, 2020. **129**(2): p. S111.
46. Hui, L., et al., *ASAH1 gene mutation causing spinal muscular atrophy with progressive myoclonic epilepsy in a Chinese child and literature review*. *Chinese Journal of Applied Clinical Pediatrics*, 2018. **33**(6): p. 470-474.
47. Alves, M.Q., et al., *Molecular basis of acid ceramidase deficiency in a neonatal form of Farber disease: identification of the first large deletion in ASAH1 gene*. *Molecular genetics and metabolism*, 2013. **109**(3): p. 276-281.
48. Ladisch, M.R. and K.L. Kohlmann, *Recombinant human insulin*. *Biotechnology progress*, 1992. **8**(6): p. 469-478.
49. Nagai, N., et al., *Recombinant human microplasmin: production and potential therapeutic properties*. *Journal of Thrombosis and Haemostasis*, 2003. **1**(2): p. 307-313.
50. Seo, W.-Y., et al., *Production of recombinant human procollagen type I C-terminal propeptide and establishment of a sandwich ELISA for quantification*. *Scientific reports*, 2017. **7**(1): p. 1-13.
51. Schuchman, E.H. and I. Galanin, *Development of enzyme replacement therapy for Farber disease and other disorders with ceramide storage*. *Mol Genet Metab*, 2014. **111**(2): p. S94.
52. Alam, M.B., et al., *Attenuation of melanogenesis by *Nymphaea nouchali* (Burm. f) flower extract through the regulation of cAMP/CREB/MAPKs/MITF and proteasomal degradation of tyrosinase*. *Scientific reports*, 2018. **8**(1): p. 1-14.
53. Balmanno, K. and S. Cook, *Tumour cell survival signalling by the ERK1/2 pathway*. *Cell Death & Differentiation*, 2009. **16**(3): p. 368-377.
54. Ferlinz, K., et al., *Human acid ceramidase: processing, glycosylation, and lysosomal targeting*. *Journal of Biological Chemistry*, 2001. **276**(38): p. 35352-35360.
55. Suresh, C., et al., *Penicillin V acylase crystal structure reveals new Ntn-hydrolase family members*. *Nature structural biology*, 1999. **6**(5): p. 414-416.
56. Okino, N., et al., *The reverse activity of human acid ceramidase*. *Journal of Biological Chemistry*, 2003. **278**(32): p. 29948-29953.

57. Bokhove, M., et al., *Structures of an isopenicillin N converting Ntn-hydrolase reveal different catalytic roles for the active site residues of precursor and mature enzyme*. *Structure*, 2010. **18**(3): p. 301-308.
58. Airola, M.V., et al., *Structural basis for ceramide recognition and hydrolysis by human neutral ceramidase*. *Structure*, 2015. **23**(8): p. 1482-1491.
59. Parveen, F., et al., *Role of ceramidases in sphingolipid metabolism and human diseases*. *Cells*, 2019. **8**(12): p. 1573.
60. Steven, A., et al., *What turns CREB on? And off? And why does it matter?* *Cellular and Molecular Life Sciences*, 2020: p. 1-19.
61. Resnick-Silverman, L., et al., *Target structure-based discovery of small molecules that block human p53 and CREB binding protein association*. *Chemistry & biology*, 2006. **13**(1): p. 81-90.
62. Wu, D., et al., *Neuroprotective Function of a Novel Hexapeptide QMDDQ from Shrimp via Activation of the PKA/CREB/BNDF Signaling Pathway and Its Structure–Activity Relationship*. *Journal of Agricultural and Food Chemistry*, 2020. **68**(24): p. 6759-6769.
63. Radhakrishnan, I., et al., *Structural analyses of CREB-CBP transcriptional activator-coactivator complexes by NMR spectroscopy: implications for mapping the boundaries of structural domains*. *Journal of molecular biology*, 1999. **287**(5): p. 859-865.
64. Chatterjee, S., et al., *The CBP KIX domain regulates long-term memory and circadian activity*. *BMC biology*, 2020. **18**(1): p. 1-23.
65. Song, Y., et al., *Structural Insights into the CRTC2–CREB Complex Assembly on CRE*. *Journal of molecular biology*, 2018. **430**(13): p. 1926-1939.
66. Middei, S., et al., *CREB selectively controls learning-induced structural remodeling of neurons*. *Learning & memory*, 2012. **19**(8): p. 330-336.
67. Wang, H., et al., *cAMP response element-binding protein (CREB): a possible signaling molecule link in the pathophysiology of schizophrenia*. *Frontiers in molecular neuroscience*, 2018. **11**: p. 255.
68. Altarejos, J.Y. and M. Montminy, *CREB and the CRTC co-activators: sensors for hormonal and metabolic signals*. *Nature reviews Molecular cell biology*, 2011. **12**(3): p. 141-151.
69. Huang ST, L.Y., Gullen EA, Cheng YC., *Impacts of baicalein analogs with modification of the 6th position of A ring on the activity toward NF-kappaB-, AP-1-, or CREB-mediated transcription*. *Bioorg. Med. Chem. Lett.*, 2008.
70. Huang, S.-T., et al., *Impacts of baicalein analogs with modification of the 6th position of A ring on the activity toward NF-κB-, AP-1-, or CREB-mediated transcription*. *Bioorganic & medicinal chemistry letters*, 2008. **18**(18): p. 5046-5049.
71. Lazennec, G., J.A. Thomas, and B.S. Katzenellenbogen, *Involvement of cyclic AMP response element binding protein (CREB) and estrogen receptor phosphorylation in the synergistic activation of the estrogen receptor by estradiol and protein kinase activators*. *The Journal of steroid biochemistry and molecular biology*, 2001. **77**(4-5): p. 193-203.
72. Moeenrezakhanlou, A., et al., *1α, 25-Dihydroxycholecalciferol activates binding of CREB to a CRE site in the CD14 promoter and drives promoter activity in a phosphatidylinositol-3 kinase-dependent manner*. *Journal of leukocyte biology*, 2007. **81**(5): p. 1311-1321.
73. Resende, C., et al., *Interleukin-1B signalling leads to increased survival of gastric carcinoma cells through a CREB-C/EBPβ-associated mechanism*. *Gastric cancer*, 2016. **19**(1): p. 74-84.

74. Fernández-Araujo, A., et al., *Key role of phosphodiesterase 4A (PDE4A) in autophagy triggered by yessotoxin*. Toxicology, 2015. **329**: p. 60-72.
75. Jiang, M., et al., *Retinoic acid induces caspase-8 transcription via phospho-CREB and increases apoptotic responses to death stimuli in neuroblastoma cells*. Biochimica et Biophysica Acta (BBA)-Molecular Cell Research, 2008. **1783**(6): p. 1055-1067.
76. Tardito, D., et al., *Early induction of CREB activation and CREB-regulating signalling by antidepressants*. International Journal of Neuropsychopharmacology, 2009. **12**(10): p. 1367-1381.
77. Woestenenk, E.A., et al., *His tag effect on solubility of human proteins produced in Escherichia coli: a comparison between four expression vectors*. Journal of structural and functional genomics, 2004. **5**(3): p. 217-229.
78. Butt, T.R., et al., *SUMO fusion technology for difficult-to-express proteins*. Protein expression and purification, 2005. **43**(1): p. 1-9.
79. Zhao, X., G. Li, and S. Liang, *Several affinity tags commonly used in chromatographic purification*. Journal of analytical methods in chemistry, 2013. **2013**.
80. Gopal, G.J. and A. Kumar, *Strategies for the production of recombinant protein in Escherichia coli*. The protein journal, 2013. **32**(6): p. 419-425.
81. Webb, B. and A. Sali, *Comparative protein structure modeling using MODELLER*. Current protocols in bioinformatics, 2016. **54**(1): p. 5.6. 1-5.6. 37.
82. Singh, A., *Deep learning 3D structures*. Nature methods, 2020. **17**(3): p. 249-249.
83. Moore, W., *Schrodinger*. 2015: Cambridge University Press.
84. Gee, C.T., et al., *Dual Labeling of the CBP/p300 KIX domain for 19F NMR leads to identification of a new small molecule binding site*. Chembiochem: a European journal of chemical biology, 2018. **19**(9): p. 963.
85. Tomoshige, S., Y. Hashimoto, and M. Ishikawa, *Efficient protein knockdown of HaloTag-fused proteins using hybrid molecules consisting of IAP antagonist and HaloTag ligand*. Bioorganic & medicinal chemistry, 2016. **24**(14): p. 3144-3148.
86. Annamala, M.K., K.K. Inampudi, and L. Guruprasad, *Docking of phosphonate and trehalose analog inhibitors into M. tuberculosis mycolyltransferase Ag85C: Comparison of the two scoring fitness functions GoldScore and ChemScore, in the GOLD software*. Bioinformation, 2007. **1**(9): p. 339.
87. Hu, Y., et al., *The importance of protonation and tautomerization in relative binding affinity prediction: a comparison of AMBER TI and Schrödinger FEP*. Journal of computer-aided molecular design, 2016. **30**(7): p. 533-539.
88. Vilar, S., G. Cozza, and S. Moro, *Medicinal chemistry and the molecular operating environment (MOE): application of QSAR and molecular docking to drug discovery*. Current topics in medicinal chemistry, 2008. **8**(18): p. 1555-1572.
89. Shaw, D.E., *"Desmond User Manual," p. 178*. 2020.
90. Martínez, L., *Automatic identification of mobile and rigid substructures in molecular dynamics simulations and fractional structural fluctuation analysis*. PloS one, 2015. **10**(3): p. e0119264.
91. Kaserer, T., et al., *Pharmacophore models and pharmacophore-based virtual screening: concepts and applications exemplified on hydroxysteroid dehydrogenases*. Molecules, 2015. **20**(12): p. 22799-22832.
92. Wishart, D.S., et al., *DrugBank 5.0: a major update to the DrugBank database for 2018*. Nucleic acids research, 2018. **46**(D1): p. D1074-D1082.
93. Irwin, J.J. and B.K. Shoichet, *ZINC- a free database of commercially available compounds for virtual screening*. Journal of chemical information and modeling, 2005. **45**(1): p. 177-182.

94. Munawar, S., et al., *Experimentally validated pharmacoinformatics approach to predict hERG inhibition potential of new chemical entities*. *Frontiers in pharmacology*, 2018. **9**: p. 1035.
95. Yang, J., et al., *Improved protein structure prediction using predicted interresidue orientations*. *Proceedings of the National Academy of Sciences*, 2020. **117**(3): p. 1496-1503.
96. Colovos, C. and T.O. Yeates, *Verification of protein structures: patterns of nonbonded atomic interactions*. *Protein science*, 1993. **2**(9): p. 1511-1519.
97. Sheik, S., et al., *Ramachandran plot on the web*. *Bioinformatics*, 2002. **18**(11): p. 1548-1549.
98. Molkenhain, N., S. Hu, and A.J. Niemi, *Discrete nonlinear Schrödinger equation and polygonal solitons with applications to collapsed proteins*. *Physical Review Letters*, 2011. **106**(7): p. 078102.
99. Hu, B., et al., *Computational insights into the sorption mechanism of polycyclic aromatic hydrocarbons by carbon nanotube through density functional theory calculation and molecular dynamics simulation*. *Computational Materials Science*, 2020. **179**: p. 109677.
100. Verdonk, M.L., et al., *Improved protein–ligand docking using GOLD*. *Proteins: Structure, Function, and Bioinformatics*, 2003. **52**(4): p. 609-623.
101. Valdemoro, C., D. Alcoba, and L. Tel, *Recent developments in the contracted Schrödinger equation method: Controlling the N-representability of the second-order reduced density matrix*. *International journal of quantum chemistry*, 2003. **93**(3): p. 212-222.
102. Halimu, C., A. Kasem, and S.S. Newaz. *Empirical comparison of area under ROC curve (AUC) and Mathew correlation coefficient (MCC) for evaluating machine learning algorithms on imbalanced datasets for binary classification*. in *Proceedings of the 3rd international conference on machine learning and soft computing*. 2019.
103. Yildiz, E.P., et al., *Spinal muscular atrophy with progressive myoclonic epilepsy linked to mutations in ASAH1*. *Clinical neurology and neurosurgery*, 2018. **164**: p. 47-49.
104. Wasserstein, M.P., et al., *Olipudase alfa for treatment of acid sphingomyelinase deficiency (ASMD): safety and efficacy in adults treated for 30 months*. *Journal of inherited metabolic disease*, 2018. **41**(5): p. 829-838.
105. Parker, D., et al., *Phosphorylation of CREB at Ser-133 induces complex formation with CREB-binding protein via a direct mechanism*. *Molecular and cellular biology*, 1996. **16**(2): p. 694-703.
106. Abdelghany, L., et al., *Dipyridamole induces the phosphorylation of CREB to promote cancer cell proliferation*. *Oncology Letters*, 2021. **21**(4): p. 1-1.
107. Chang, J.-M., P. Di Tommaso, and C. Notredame, *TCS: a new multiple sequence alignment reliability measure to estimate alignment accuracy and improve phylogenetic tree reconstruction*. *Molecular biology and evolution*, 2014. **31**(6): p. 1625-1637.
108. Hayward, S. and E.J. Milner-White, *Determination of amino acids that favour the α L region using Ramachandran propensity plots. Implications for α -sheet as the possible amyloid intermediate*. *Journal of Structural Biology*, 2021. **213**(2): p. 107738.
109. Nagi, A.D. and L. Regan, *An inverse correlation between loop length and stability in a four-helix-bundle protein*. *Folding and Design*, 1997. **2**(1): p. 67-75.
110. Balaji, S., R. Kalpana, and P. Shapshak, *Paradigm development: comparative and predictive 3D modeling of HIV-1 Virion Infectivity Factor (Vif)*. *Bioinformation*, 2006. **1**(8): p. 290.
111. Aier, I., P.K. Varadwaj, and U. Raj, *Structural insights into conformational stability of both wild-type and mutant EZH2 receptor*. *Scientific reports*, 2016. **6**(1): p. 1-10.

112. Cheng, F., et al., *admetSAR: a comprehensive source and free tool for assessment of chemical ADMET properties*. 2012, ACS Publications.
113. Chia-en, A.C., W. Chen, and M.K. Gilson, *Ligand configurational entropy and protein binding*. Proceedings of the National Academy of Sciences, 2007. **104**(5): p. 1534-1539.
114. Aamir, M., et al., *In silico prediction, characterization, molecular docking, and dynamic studies on fungal SDRs as novel targets for searching potential fungicides against Fusarium wilt in tomato*. Frontiers in pharmacology, 2018. **9**: p. 1038.
115. Saxena, A., et al., *Modelling the binding affinity of steroids to zebrafish sex hormone-binding globulin*. SAR and QSAR in Environmental Research, 2014. **25**(5): p. 407-421.
116. Rehman, Z., et al., *Molecular docking and pharmacophore models to probe binding hypothesis of inhibitors of hypoxia inducible factor-1*. Journal of Biomolecular Structure and Dynamics, 2021: p. 1-12.
117. Abdul-Hammed, M., et al., *Virtual screening, ADMET profiling, PASS prediction, and bioactivity studies of potential inhibitory roles of alkaloids, phytosterols, and flavonoids against COVID-19 main protease (Mpro)*. Natural product research, 2021: p. 1-7.

CHAPTER 8

APPENDIX

Appendix 1: Sequences of Recombinant aCDase (r-aCDase):

SEQ1 >seq1
 MGSSHHHHHSSGLVPRGSHMASMTGGQQMGRGSEFCTSIVAEDKKG
 HLIHGRNMDFGVFLGWNINNDTWVITEQLKPLTVNLDFQRNNKTVFK
 ASSFAGYVGMLTGFKPGLFSLTLNERFSINGGYLGILEWILGKKDVM
 WIGFLTRTVLENSTSYEEAKNLLTKTKILAPAYFILGGNQSGEGCVI
 TRDRKESLDVYELDAKQGRWYVVQTNYDRWKHPFFLDDRRTPAKMCL
 NRTSQENISFETMYDVLSTKPVLNKLTVYTTLIDVTKGQFETYLRDC
 PDPCIGW

SEQ2 >seq2
 MRTADREARPGLPSLLLLLLAGAGLSAASPPAAPRFNVSLDSVPELR
 WLPVLRHYDLDLVRAAMAQVIGDRVPKWVHVLIKVVLELERFLPQP
 FTGEIRGMCDFMNLSLADCLLVNLAYESSVFCTSIVAQDSRGHIYHG
 RNLDPFGNILRKLTVDVQFLKNGQIAFTGTTFIGYVGLWTGQSPHK
 FTVSGDERDKGWWENAI AALFRRHIPVSWLIRATLSESENFEEAVG
 KLAKTPLIADVYYIVGGTSPREGVVITRNRDGPADIWPLDPLNGAWF
 RVETNYDHWKPAPKEDDRRTSAIKALNATGQANLSLEALFQILSVVP
 VYNNFTIYTTVMSAGSPDKYMTRIRNPSRKHHHHHHHHHH

SEQ3 >seq3
 MESPI LGYWKIKGLVQPTRLLEYLEEKYEEHLYERDEGDKWRNKKF
 ELGLEFPNLPYYIDGDVKL TQSM A I RY IADKH NMLGGCPKERA E I S
 MLEGAVLDIRYGVSR IAYS KDFETLKVDFLSKLP EMLKMFEDRLCHK
 TYLNGDHVTHPDFM L YDALDVVLYMDPMCLDAFPKLVCFKKRIEAI P
 QIDKYLKSSKYIAWPLQGWQATFGGGDHPPKSDLEVL FQGPLETSLY
 KKAGTMPGRSCVALVLLAAAVSCAVAQHAPPWTEDCRKSTYPPSGPT
 YRGAVPWYTINLDLPPYKRWHELM LDKAPMLKVI VNSLKNMINTFVP
 SGKVMQVVDEKLPGLLG NFP GP FEEEMKGIAAVTDIPLGEI I SFNIF
 YELFTICTSIVAEDKKGHLI HGRNMDFGVFLGWNINNDTWVITEQLK
 PLTVNLDFQRNNKTVFKASSFAGYVGMLTGFKPGLFSLTLNERFSIN
 GGYLGILEWILGKKDAMWIGFLTRTVLENSTSYEEAKNLLTKTKILA
 PAYFILGGNQSGEGCVITRDRKESLDVYELDAKQGRWYVVQTNYDRW
 KHPFFLDDRRTPAKMCLNRTSQENISFETMYDVLSTKPVLNKLTVYT

TLIDVTKGQFETYLRDCPDPCIGW

SEQ4 >seq4
 HHHHHHMSDSEVNQEAKPEVKPEVKPETHINLKVSDGSSEIFFKIKK
 TTPLRRLMEAFAKRQ GKEMDSLRF LYDGIRIQADQTPEDLDMEDNDI
 IEAHREQIGGQHAPPWTEDCRKSTYPPSGPTYRGAVPWYTINLDLPP
 YKRWHELMLDKAPVLKVI VNSLKNMINTFVPSGKIMQVVDEKLPGLL
 GNFPGPFEEMKGIAAVTDIPLGEIISFNIFYELFTICTSIVAEDKK
 GHLIHGRNMDFGVFLGWNINNDTWVITEQLKPLTVNLDFQRNNKTVF
 KASSFAGYVGMLTGFKPGLFSLTLNERFSINGGYLGILEWILGKKDV
 MWIGFLTRTVLENSTSYEEAKNLLTKTKILAPAYFILGGNQS GEGCV
 ITRDRKESLDVYELDAKQGRWYVVQTN YDRWKHPFFLDDRRTPAKMC
 LNRTSQENISFETMYDVLSTKPVLNKLTVYTTLIDVTKGQFETYLRD
 CPDPCIGW

SEQ5 >seq5
 HHHHHHHHHHQHAPPWTEDCRKSTYPPSGPTYRGAVPWYTINLDLPP
 YKRWHELMLDKAPVLKVI VNSLKNMINTFVPSGKIMQVVDEKLPGLL
 GNFPGPFEEMKGIAAVTDIPLGEIISFNIFYELFTICTSIVAEDKK
 GHLIHGRNMDFGVFLGWNINNDTWVITEQLKPLTVNLDFQRNNKTVF
 KASSFAGYVGMLTGFKPGLFSLTLNERFSINGGYLGILEWILGKKDV
 MWIGFLTRTVLENSTSYEEAKNLLTKTKILAPAYFILGGNQS GEGCV
 ITRDRKESLDVYELDAKQGRWYVVQTN YDRWKHPFFLDDRRTPAKMC
 LNRTSQENISFETMYDVLSTKPVLNKLTVYTTLIDVTKGQFETYLRD
 CPDPCIGWEQKLISEEDL

SEQ6 >seq6
 MAHHHHHMSDSEVNQEAKPEVKPEVKPETHINLKVSDGSSEIFFKI
 KKTTPLRRLMEAFAKRQ GKEMDSLRF LYDGIRIQADQTPEDLDMEDN
 DIIEAHREQIGGGSEFRTQHAPPWTEDCRKSTYPPSGPTYRGAVPWY
 TINLDLPPYKRWHELMLDKAPVLKVI VNSLKNMINTFVPSGKIMQVV
 DEKLPGLLGNFPGPFEEMKGIAAVTDIPLGEIISFNIFYELFTICT
 SIVAEDKKGHLIHGRNMDFGVFLGWNINNDTWVITEQLKPLTVNLDF

QRNNKTVFKASSFAGYVGMLTGFKPGLFSLTLNERFSINGGYLGILE
 WILGKKDVMWIGFLTRTVLENSTSYEEAKNLLTKTKILAPAYFILGG
 NQSGEGCVITRDRKESLDVYELDAKQGRWYVVQTNYDRWKHPFFLDD
 RRTPAKMCLNRTSQENISFETMYDVLSTKPVLNKLTVYTTLIDVTKG
 QFETYLRDCPDPCIGW

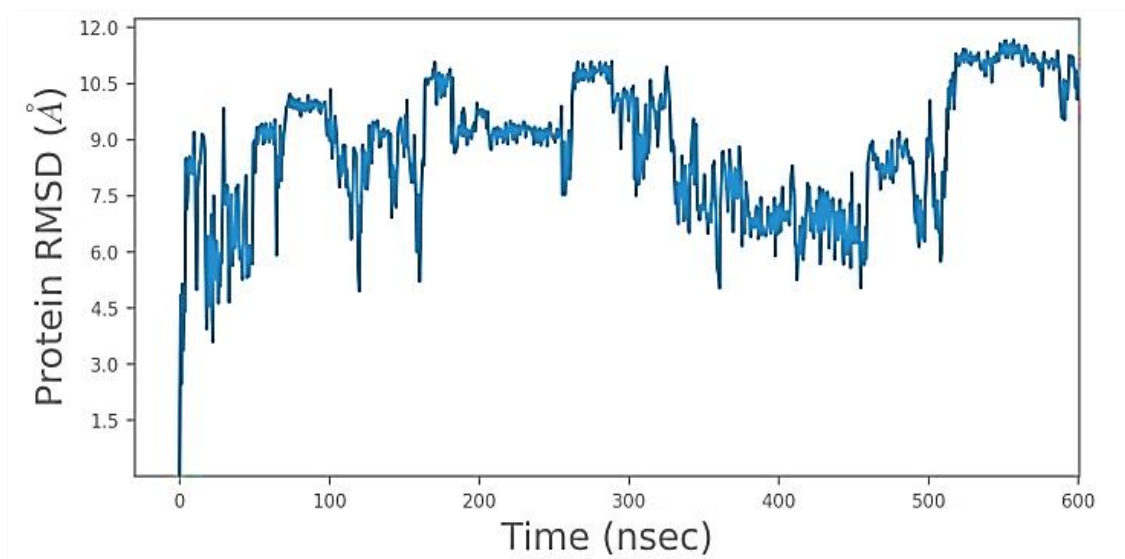
SEQ7

>seq7

MKFLVNVALVFMVVYISYIYAHHHHHHHHHQHAPPWTEDCRKSTYP
 PSGPTYRGAVPWYTINLDLPPYKRWHELMMLDKAPVLKVIIVNSLKNMI
 NTFVPSGKIMQVVDEKLPGLLGNFPGPFFEEEMKGIAAVTDIPLGEII
 SFNIFYELFTICTSIVAEDKKGHLIHGRNMDFGVFLGWNINNDTWVI
 TEQLKPLTVNLDFQRNNKTVFKASSFAGYVGMLTGFKPGLFSLTLNE
 RFSINGGYLGILEWILGKKDVMWIGFLTRTVLENSTSYEEAKNLLTK
 TKILAPAYFILGGNQSGEGCVITRDRKESLDVYELDAKQGRWYVVQT
 NYDRWKHPFFLDDRRTPAKMCLNRTSQENISFETMYDVLSTKPVLNK
 LTVYTTLIDVTKGQFETYLRDCPDPCIGW

Appendix 2: MD Simulation of CREB at 600nsec

First, CREB was utilized to perform the molecular docking studies due to its direct interaction with the collected inducers datasets. For this purpose, CREB was simulated for 600nsec, and then docking was performed (given below). However, when the docked complex was subjected to further MD for interactions stability, the structure of CREB got unstable again. That is why the complex of CBP/CREB was preferred to be docked with inducers in this project. The main purpose was to keep the CREB stable while exploring different conformation of the ligands.



Appendix 2: The RMSD vs time plot during the MD simulation of CREB protein.

Appendix 3: FDA-Approved drugs for Epilepsy (EPI) and Spinal Muscular Atrophy (SMA)-EPI/SMA Database:

Mol_ID	smiles	Drug Name	Function
EPI_1	<chem>O=C(N)C(CC)N1C(=O)CC(CCC)C1</chem>	Brivaracetam	Seizures Control
EPI_2	<chem>Oc1c(c(O)cc(CCCCC)c1)C1C(C(=C)C)CCC(C)=C1</chem>	Cannabidiol	Seizures Control
EPI_3	<chem>O=C(N)N1c2c(ccc2)C=Cc2c1cccc2</chem>	Carbamazepine	Anti Convulsant
EPI_4	<chem>Clc1cc2N(c3ccccc3)C(=O)CC(=O)N(C)c2cc1</chem>	Clobazam	Seizures Control
EPI_5	<chem>Clc1c(C2=NCC(=O)Nc3c2cc([N+](=O)[O-])cc3)cccc1</chem>	Clonazepam	Anti Convulsant
EPI_6	<chem>O=C(O)CN(C)C</chem>	Dimethylglycine	Seizures Control
EPI_7	<chem>O=C(O)C(CCC)CCC.O=C([O-</chem>	Divalproex	Anti

	<chem>]C(CCC)CCC</chem>		Convulsant
EPI_8	<chem>O=C1C(CC)(C)CC(=O)N1</chem>	Ethosuximide	Anti Convulsant
EPI_9	<chem>O=C1N(CC)C(=O)NC1c1cccc1</chem>	Ethotoin	Anti Convulsant
EPI_10	<chem>Fc1ccc(CNc2cc(N)c(NC(=O)OCC)cc2)cc1</chem>	Ezogabine	Seizures Control
EPI_11	<chem>O=C(OCC(COC(=O)N)c1cccc1)N</chem>	Felbamate	Anti Convulsant
EPI_12	<chem>Fc1ccc(C(N2CCN(CC=Cc3cccc3)C2)c2ccc(F)cc2)cc1</chem>	Flunarizine	Calcium Blocker (as adjuvant in Epilepsy therapy)
EPI_13	<chem>P(=O)(OCN1C(=O)C(c2cccc2)(c2ccc2)NC1=O)(O)O</chem>	Fosphenytoin	Anti Convulsant
EPI_14	<chem>O=C(NCc1cccc1)[C@H](NC(=O)C)COC</chem>	Lacosamide	Anti Convulsant
EPI_15	<chem>O=C(N)C(CC)N1C(=O)CCC1</chem>	Levetiracetam	Anti Convulsant
EPI_16	<chem>O=C1N(C)C(=O)CC1(C)c1cccc1</chem>	Methsuximide	Anti Convulsant
EPI_17	<chem>O=C(N)N1c2c(C(=O)Cc3c1cccc3)ccc2</chem>	Oxcarbazepine	Anti Convulsant
EPI_18	<chem>O=C1N(c2cccc2)C=C(c2nccc2)C=C1c1c(C#N)cccc1</chem>	Perampanel	Seizures Control
EPI_19	<chem>O=C1C(CC)(c2cccc2)C(=O)NC(=O)N1</chem>	Phenobarbitone	Seizures Control
EPI_20	<chem>O=C(NC(=O)Cc1cccc1)N</chem>	Phenacemide	Anti Convulsant

EPI_21	<chem>O=C1C(c2ccccc2)(c2ccccc2)NC(=O)N1</chem>	Phenytoin	Anti Convulsant
EPI_22	<chem>O=C(O)CC(CN)CC(C)C</chem>	Pregabalin	Anti Convulsant
EPI_23	<chem>O=C1C(CC)(c2ccccc2)C(=O)NCN1</chem>	Primidone	Anti Convulsant
EPI_24	<chem>Fc1c(Cn2nnc(C(=O)N)c2)c(F)ccc1</chem>	Rufinamide	Anti Convulsant
EPI_25	<chem>OC(C(C)(C)C)C=Cc1cc2OCOc2cc1</chem>	Stiripentol	Seizures Control
EPI_26	<chem>O=C1N(C)C(=O)OC1(C)C</chem>	Trimethadione	Seizures Control
EPI_27	<chem>O=C(O)C(CCC)CCC</chem>	Valproic	Anti Convulsant
EPI_28	<chem>O=C(O)CCC(N)C=C</chem>	Vigabatrin	Seizures Control
EPI_29	<chem>S(=O)(=O)(N)Cc1noc2c1cccc2</chem>	Zonisamide	Anti Convulsant
EPI_30	<chem>S(=O)(=O)(N)c1sc(NC(=O)C)nn1</chem>	Acetazolam	Anti Convulsant
EPI_31	<chem>Clc1cc2C(c3ccccc3)=NCC(=O)N(C)c2cc1</chem>	Diazepam	Seizures Control
EPI_32	<chem>O=C(O)CC1(CN)CCCC1</chem>	Gabapentin	Anti Convulsant
EPI_33	<chem>O=[N+](O-)]c1cc2C(c3ccccc3)=NCC(=O)Nc2cc1</chem>	Nitrazepam	Anti Convulsant
EPI_34	<chem>CC1OC(C)OC(C)O1</chem>	Paraldehyde	Anti Convulsant
EPI_35	<chem>O=C1C(CC)(c2ccccc2)C(=O)NC(=O)N1</chem>	Phenobarbital	Anti Convulsant
EPI_36	<chem>S(=O)(=O)(OCC12OC(C)(C)OC1C1OC(C)(C)OC1CO2)N</chem>	Topiramate	Anti Convulsant

EPI_37	<chem>O=C(O)CCC(N)C=C</chem>	Vigabatrin	Anti Convulsant
EPI_38	<chem>O=C(O)C1CN(CCC=C(c2c(C)ccs2)c2c(C)ccs2)CCC1</chem>	Gabitril	Seizures Control
EPI_39	<chem>Clc1c(C2=NC(O)C(=O)Nc3c2cc(Cl)c3)cccc1</chem>	lorazepam	Anti Convulsant
EPI_40	<chem>O=C(N)N1c2c(C(O)Cc3c1cccc3)cccc2</chem>	Eslicarbazepine	Anti Convulsant
EPI_41	<chem>FC(F)(F)c1cc(CC(NCC)C)ccc1</chem>	Fenfluramine	Seizures Control
EPI_42	<chem>Clc1c(Cl)cccc1-c1c(N)nc(N)nn1</chem>	Lamotrigine	Seizures Control
EPI_43	<chem>O=C(N)CN1C(=O)CCC1</chem>	Piracetam	Anti Convulsant
EPI_44	<chem>Brc1cc2C(c3ncccc3)=NC(CCC(=O)OC)c3n(c(C)cn3)-c2cc1</chem>	Remimazolam	Anti Convulsant
EPI_45	<chem>Clc1cc2C(c3c(F)cccc3)=NCc3n(c(C)nc3)-c2cc1</chem>	Midazolam	Seizures Control
SMA_1	<chem>S=P(OCC1C(OP(=S)(OCC2C(OP(=S)(OCC3C(OP(=S)(OCC4C(OP(=S)(OCC5C(OP(=S)(OCC6C(OP(=S)(OCC7C(O)C(OCCOC)C(n8c9N=C(N)NC(=O)c9nc8)O7)O)C(OCCOC)C(n7c8N=C(N)NC(=O)c8nc7)O6)O)C(OCCOC)C(N6C(=O)NC(=O)C(C)=C6)O5)O)C(OCCOC)C(N5C(=O)N=C(N)C(C)=C5)O4)O)C(OCCOC)C(n4c5N=C(N)NC(=O)c5nc4)O3)O)C(OCCOC)C(N3C(=O)NC(=O)C(C)=C3)O2)O)C(OCCOC)C(n2c3ncnc(N)c3nc2)O1)(OC1C(OCCOC)C(n2c3ncnc(N)c3nc2)OC1COP(=S)(OC1C(OCCOC)C(N2C(=O)N</chem>	Spinraza	Improve Voluntary Movement

