

**ENHANCING L-ASPARAGINASE CATALYTIC
ACTIVITY FOR IMPROVED ANTILEUKEMIC
ACTIVITY: A COMPUTATIONAL STUDY ON
THERMOCOCCUS KODAKARENSIS L-
ASPARAGINASE MUTATIONS**

**A Thesis Submitted to
the Graduate School of Engineering and Sciences of
İzmir Institute of Technology
in Partial Fulfillment of the Requirements for the Degree of**

MASTER OF SCIENCE

in Bioengineering

**by
Pelinsu EKMEKÇİ**

**December 2023
İZMİR**

We approve the thesis of **Pelinsu EKMEKÇİ**

Examining Committee Members:

Assoc. Prof. Nur Başak SÜRMEİ ERALTUG

Department of Bioengineering, Izmir Institute of Technology

Assist. Prof. Dr. Arzu UYAR

Department of Bioengineering, Izmir Institute of Technology

Assist. Prof. Dr. Hümevra TAŞKENT SEZGİN

Department of Bioengineering, Izmir Institute of Technology

Assist. Prof. Dr. Mehmet Emin USLU

Department of Bioengineering, Manisa Celal Bayar University

Assist. Prof. Dr. Deniz KARATAŞ

Department of Bioengineering, Manisa Celal Bayar University

7 December 2023

Assoc. Prof. Nur Başak SÜRMEİ ERALTUG

Supervisor, Department of Bioengineering

Assist. Prof. Dr. Deniz KARATAŞ

Co-Advisor, Department of
Bioengineering, Manisa Celal Bayar
University

Assist. Prof. Dr. Ceyda ÖKSEL KARAKUŞ

Head of the Department of Bioengineering

Prof. Dr. Mehtap EANES

Dean of the Graduate School of
Engineering and Science

ACKNOWLEDGMENTS

I would like to thank to my advisor and co-advisor Assoc. Dr. Nur Bařak SÜRMEĻİ ERALTUĖ and Assist. Prof. Dr. Deniz KARATAŐ for their help, interest and support throughout this project. Getting to this point would not have been possible without their constant and masterful guidance.

I would like to express my endless gratitude to the members of Surmeli Laboratory; We thank Ekin KESTEVUR, TuĖçe SAKALLI, and Alper ŐAHİN for their valuable friendship and generous assistance.

I am grateful to my dear friends, Kutlu Őafak ALKAN, Yusuf Sinan SEZER, and Būőra ARSLAN, who supported me both professionally and morally throughout this process.

Finally, I would like to thank my mother and sister; were both with me for moral support.

ABSTRACT

ENHANCING L-ASPARAGINASE CATALYTIC ACTIVITY FOR IMPROVED ANTILEUKEMIC ACTIVITY: A COMPUTATIONAL STUDY ON THERMOCOCCUS KODAKARENSIS L-ASPARAGINASE MUTATIONS

In this study, thermostable *Thermococcus kodakarensis* L-asparaginase (TkA) enzyme, which lacks glutaminase activity, was studied for its structural and dynamic properties. The structural and dynamic properties of TkA, was investigated in its apo state and with the L-asparagine ligand to understand how the active site and general structure of the TkA enzyme changes with ligand binding and what effect this interaction has on the general behavior of the enzyme.

T11E, T55E and D86S mutants of TkA were examined by molecular docking and molecular dynamics simulations.

Binding results for molecular docking indicate that the structure is largely conserved, with root mean square deviation (RMSD) scores of -5.2 to -5.7 nm for wild-type TkA and mutants. RMSD and root mean square fluctuations (RMSF) data obtained as a result of molecular dynamics studies showed that the mutants had a stability close to that of the WT TkA enzyme, between 0.15 and 0.16 nm. In general, solvent accessible surface area (SASA) and radius of gyration analysis results support this analysis, while the D86S mutant gave more effective results than other mutants with SASA value of 260.38 nm²/ns and radius of gyration values of 2.61 nm/ns.

The total interaction energy of the ligand and WT TkA was -337.98 kJ/mol, while the interaction energy for D86S mutant was larger, at -363.03 kJ/mol. In conclusion, the study showing how the structure and dynamics of the TkA enzyme are affected by the binding of L-asparagine ligand helps to understand the stability and functional behavior of the enzyme.

ÖZET

GELİŞMİŞ ANTİLÖSEMİK AKTIVITE İÇİN L-ASPARAGINAZ KATALİTİK AKTİVİTESİNİN GELİŞTİRİLMESİ: THERMOCOCCUS KODAKARENSIS L- ASPARAGINAZ MUTASYONLARI ÜZERİNDE HESAPLAMALI BİR ÇALIŞMA

Bu çalışmada glutaminaz aktivitesi olmayan termostabil *Thermococcus kodakarensis* L-asparaginaz (TkA) enziminin yapısal ve dinamik özellikleri incelenmiştir. TkA enziminin aktif bölgesinin ve genel yapısının ligand bağlanmasıyla nasıl değiştiğini ve bu etkileşimin TkA'nın genel davranışı üzerinde ne gibi bir etkiye sahip olduğunu anlamak için TkA'nın yapısal ve dinamik özellikleri, apo durumunda ve L-asparajin ligandı ile araştırıldı.

TkA'nın T11E, T55E ve D86S mutantları, moleküler kenetleme ve moleküler dinamik simülasyonları ile incelendi.

Moleküler kenetlenme için bağlanma sonuçları, vahşi tip TkA ve mutantlar için kök ortalama kare sapma (RMSD) skorlarının -5,2 ila -5,7 nm olmasıyla yapının büyük ölçüde korunduğunu göstermektedir. Moleküler dinamik çalışmaları sonucunda elde edilen RMSD ve kök ortalama kare dalgalanmaları (RMSF) verileri, mutantların 0,15 ile 0,16 nm arasında WT TkA enzimine yakın bir stabiliteye sahip olduğunu gösterdi. Genel olarak çözücüyle erişilebilir yüzey alanı (SASA) ve dönme yarıçapı analiz sonuçları bu analizi desteklerken, D86S mutantı SASA değeri 260,38 nm²/ns ve dönme yarıçapı değeri 2,61 nm/ns ile diğer mutantlara göre daha etkili sonuçlar verdi.

Ligandın ve WT TkA'nın toplam etkileşim enerjisi -337,98 kJ/mol iken D86S mutantının etkileşim enerjisi -363,03 kJ/mol ile daha büyüktü. Sonuç olarak, TkA enziminin yapısının ve dinamiğinin L-asparajin ligandının bağlanmasından nasıl etkilendiğini gösteren çalışma, enzimin stabilitesinin ve fonksiyonel davranışının anlaşılmasına yardımcı olmaktadır.

TABLE OF CONTENTS

LIST OF FIGURES	viii
LIST OF TABLES.....	xv
CHAPTER 1.INTRODUCTION	1
1.1. Overview of Acute Lymphoblastic Leukemia (ALL) and its treatment with L-Asparaginase	6
1.2. Molecular Docking Studies.....	8
1.3. Molecular Dynamics Simulations	11
1.3.1. Molecular Dynamics Simulations for Protein Systems	14
1.3.2. Periodic Boundary Conditions (PBC)	14
1.3.3. Ewald Summation Techniques	15
1.3.4. Particle Mesh Ewald Method.....	15
1.3.5. Thermostats.....	15
1.3.6. Solvent Models	15
1.3.7. Energy-Minimization Methods.....	16
CHAPTER 2.LITERATURE REVIEW	17
CHAPTER 3.MATERIAL AND METHODS	26
3.1 . Tka Structural Alignment	27
3.2. Hotspot Wizard	27
3.3. In Silico Mutagenesis.....	28
3.4. Autodock Tools and Autodock Vina for Molecular Docking Studies.....	29
3.5. Molecular Dynamics Simulation via GROMACS.....	30
3.6. MD Simulation Analysis and Evaluation.....	32
3.6.1. Cluster Analysis.....	32
3.6.2. RMSD (Root Mean Square Deviation).....	32
3.6.3. RMSF (Root Mean Square Fluctuations)	33
3.6.4. Radius of Gyration.....	33
3.6.5. Number of Hydrogen Bond Analysis	34
3.6.6. SASA (Solvent Accessible Surface Area).....	34
3.6.7. Protein Ligand Interaction Energy.....	35
3.6.8. RDF (Radial Distribution Functions)	35
3.6.9. MSD and Diffusion Coefficient (Mean Square Displacement).....	36
CHAPTER 4.RESULTS AND DISCUSSION.....	37
4.1. Molecular Docking Studies and Mutations Selection.....	37
4.2. Molecular Docking Studies with WT Tka and L-asparagine Ligand.....	39

4.3. Molecular Docking Studies with Mutated TkAs and L-asparagine Ligand and Mutation Selection	43
4.4. Molecular Dynamics Simulations	47
4.4.1. Apo State WT TkA in Water 50 ns Simulations	48
4.4.2. Molecular Dynamics Simulations for WT TkA and Asparagine Ligand Complex	53
4.4.3. Molecular Dynamics Simulations for T55E TkA and Asparagine Ligand Complex	67
4.4.4. Molecular Dynamics Simulations for D86S TkA and Asparagine Ligand Complex	77
4.4.5. Molecular Dynamics Simulations for T11E TkA and Asparagine Ligand Complex	88
4.4.6. Comparison of All Mutants and WT TkA	99
CHAPTER 5. CONCLUSION	102
REFERENCES	104
APPENDICES	112
APPENDIX A	112
APPENDIX B	119
APPENDIX C	120
APPENDIX D	122

LIST OF FIGURES

<u>Figure</u>	<u>Page</u>
Figure 1.1. Two different basic mechanisms are known for the L-ASNase enzyme. (A) Two step substitution mechanism (B) One step mechanism. (Lubkowski et al. 2020)	2
Figure 1.2. Representations of the common structure model for L-ASNase from Escherichia coli, A) Quaternary structure is obtained from Escherichia coli, B) 2D secondary structures from Escherichia coli, C) Amino acid sequences that is obtained from Escherichia coli. (Lubkowski et al. 2020)	3
Figure 1.3. Most common structure of tetramer L-ASNase in A and their active sites in B that are obtained from Escherichia coli. (Lubkowski et al. 2020).....	4
Figure 1.4. Mechanism of L-asparaginase enzyme into the blood stream and effect on the tumor cells.	8
Figure 1.5. Protein dynamics: a hierarchy of major motions. Structural conformations and protein-ligand interactions obtained from molecular dynamics simulations have been visualised. Processes have been visualised from femtoseconds to milliseconds. For each scale, typical structural changes and interactions are shown (Surpeta, Sequeiros-Borja, and Brezovsky 2020).	12
Figure 1.6. General box environment during MD simulation. Protein molecule simulation in water (Lemkul 2019).	13
Figure 2.1. E. coli L-ASNase (PDBID:1NNS) and TkA 3D structure comparison via structural superposition (PDBID:5OT0). TkA chain A: red, TkA chain B: yellow and E. coli is indicated by blue color. Arrows indicate segments that adapt for different conformations while substrate.	18
Figure 2.2. Schematic representation reactions of a) L-asparaginase, b) L-glutaminase (Orabi et al. 2019).	19
Figure 2.3. 3D structure of TkA enzyme with B-hairpin and putative allosteric site representations. Yellow circle indicates active region. (Guo et al. 2017).....	20
Figure 2.4. A) Active site residues for TkA with PO_4^{3-} . B) E. coli L-ASNase with L-asparagine ligand.	21
Figure 3.1. Flow chart for overall methodology.	26

<u>Figure</u>	<u>Page</u>
Figure 3.2. Illustrated demonstration of the Pymol interface and some instructions. The arrows indicate the parts that should be used for mutagenesis protocols.....	28
Figure 3.3. Hydrogen bond geometry in water. (1) shows H-bond between two residues. (2) indicates H-bonding bridge via water molecules.....	34
Figure 4.1. Multiple Sequence Alignment between TkA, E. coli, P horiskoshii, E chrysantemi and H.pylori. The parts indicated with black frames in the figure show the active site residues of TkA. It also shows similarities with other organisms....	38
Figure 4.2. A comprehensive docking study and docking scores by Baral et al. for the l-asparaginase family that categorized as gene and type of l-asparaginase. ^B indicate ansb gene and 1 or 2 indicates type of l-asparaginase.....	41
Figure 4.3. Three-dimensional structure of TkA and its ligand asparagine after docking. The green color indicates the L-asparagine ligand, and the orange parts represent active site residues. The image on the left was created with Ligplot+ and shows hydrogen bond formation.	42
Figure 4.4. Active site residues and asparagine ligand of E. coli L-asparaginase. Showing Thr11 and Thr85 residues that are involved in the nucleophilic attack on the substrate.....	43
Figure 4.5. Amino acid substitution matrix of the protein. The table shows the possible percentages of amino acid substitutions that can occur at certain positions in the protein. Each row represents a different amino acid combination (shown left), and each column represents a possible substituting amino acid (shown above). The color coding visualizes the level of the change score: Green indicates a higher match; red indicates a lower match. The 'X' sign is an indication that a change at that position is unlikely.....	45
Figure 4.6. A) Structure of T55E TkA highlighted in orange indicate active site residues and green indicates the amino acid ligand. B) 3D TkA structure of the Asp86Ser mutation, structures highlighted in orange indicate active site residues and green indicates the amino acid ligand. C) 3D TkA structure of Thr11Glu mutation, structures highlighted in orange indicate active site residues and green indicates the amino acid ligand. Additionally, regions with mutations are shown in yellow.	46
Figure 4.7. Shows RMSD graphs of WT Apo TkA. A) Shows the RMSD graphs of two separate repetitions, while B) shows to the average RMSD graph of two repetitions.	49

<u>Figure</u>	<u>Page</u>
Figure 4.8. RMSF plot and Chimera images of Apo TkA. A) RMSF graphs taken from two replicate studies of Apo TkA, B) shows the average RMSF graph obtained from RMSF graphs. C) As a result of cluster analysis, comparison of the most frequent encountered conformation (pink) during the simulation and the apo TkA conformations at the starting point (blue).	50
Figure 4.9. Radius of gyration plot to evaluate protein compactness. A) shows two repetitions for apo TkA, B) shows an average plot obtained from two repetitions of Apo TkA plot.	51
Figure 4.10. SASA analysis plot for Apo TkA. A) shows two repetitions studies for apo TkA and B) shows average SASA plot from two repetitions studies of apo TkA..	52
Figure 4.11. A) Backbone-to-backbone RMSD plot for WT TkA and asparagine ligand complex with 3 repetitions. B) Average RMSD plot from 3 repetitions for WT TkA with ligand. C) Comparing RMSD data for WT apo TkA and WT TkA with ligand.	54
Figure 4.12. A) RMSF graphs of three replicates for WT TkA with l-asparagine ligand. B) Average RMSF graph for WT TkA with l-asparagine ligand. C) Comparing RMSF graphs between WT TkA and its l-asparagine ligand and WT apo TkA.....	57
Figure 4.13. Initial position image of WT TkA (light pink) and l-asparagine ligand (red) and image of WT TkA (light blue) and l-asparagine ligand (cyan) at 10ns. The part shown in yellow here represents the hairpin: Residues: β 2(14-16) - β 3(20-22). Blue regions indicate the allosteric site: Residues: α 8(293-306) - β 15(258-262). Green regions α 1(27-33)– α 4(122-133) are regions that tend to move towards the active site during ligand and enzyme interaction.	58
Figure 4.14. A) Radius of gyration graphs of WT TkA with l-asparagine ligand from three different replicate runs. B) Average radius of gyration from three replicate runs. C) Comparative radius of gyration graph of WT apo TkA and WT TkA l-asparagine ligand complex.	59
Figure 4.15. The number of hydrogen bonds between atom pairs within 0.35 nm distance.	60
Figure 4.16. The images show hydrogen bond formation between l-asparagine ligand and some residues that are in active site of WT TkA.....	61

<u>Figure</u>	<u>Page</u>
Figure 4.17. A) SASA graphs of WT TkA with l-asparagine ligand from three different replicate runs. B) Average SASA from three replicate runs. C) Comparative SASA graph of WT apo TkA and WT TkA l-asparagine ligand complex.....	63
Figure 4.18. WT TkA and l-asparagine ligand interaction energies plots with three repeated studies.	64
Figure 4.19. A) RDF plots are a comparative plot of 3 replicates. B) is the RDF graph of 5 different residues that may be important.	66
Figure 4.20. A) RMSD graphs of T55E TkA two replicates. B) T55E TkA average RMSD graph obtained from two repetitions. C) RMSD analysis; Comparison with WT TkA and T55E mutant TkA.	68
Figure 4.21. A) RMSF graphs of T55E TkA two replicates. B) T55E TkA average RMSF graph obtained from two repetitions. C) RMSF analysis; Comparison with WT TkA and T55E mutant TkA.	69
Figure 4.22. Initial position image of T55E TkA (light pink) and l-asparagine ligand (red) and image of T55E TkA (light blue) and l-asparagine ligand (cyan) at 10ns. The part shown in yellow here represents the hairpin: Residues: β 2(14-16) - β 3(20-22). Blue regions indicate the allosteric site: Residues: α 8(293-306) - β 15(258-262). Green regions α 1(27-33)– α 4(122-133) are regions that tend to move towards the active site during ligand and enzyme interaction.	70
Figure 4.23. A) Radius of gyration plot for T55E TkA with two repeats. B) Average plot of radius of gyration for T55E TkA and its ligand. C) Comparative radius of gyration for T55E TkA and WT TkA.	71
Figure 4.24. The number of hydrogen bonds between atom pairs within 0.35 nm distance.	72
Figure 4.25. The images show hydrogen bond formation between l-asparagine ligand and some residues that are in active site of T55E TkA.	73
Figure 4.26. Images A show SASA plots with two repeated studies and average SASA plot respectively. Image B shows T55E TkA and WT TkA are compared on the same plot.	74
Figure 4.27. A) shows WT TkA an l-asparagine ligand interaction energy plot. B) shows T55E TkA interaction energy graph. Red regions represent LJ-SR and black regions represent Coul-SR energy.	75

<u>Figure</u>	<u>Page</u>
Figure 4.28. A) shows comparison of WT TkA RDF values and T55E TkA RDF values. B) RDF analysis is given specifically for residues for WT TkA. C) RDF analysis is given specifically for residues for T55E TkA.	76
Figure 4.29. A) RMSD graphs of D86S TkA two replicates. B) D86S TkA average RMSD graph obtained from two repetitions. C) RMSD analysis; Comparison with WT TkA and D86S TkA with l-asparagine ligand.	79
Figure 4.30. A) RMSF graphs of D86S TkA two replicates. B) D86S TkA average RMSF graph obtained from two repetitions. C) RMSF analysis; Comparison with WT TkA and D86S TkA.	80
Figure 4.31. Initial position image of D86S TkA (light pink) and l-asparagine ligand (red) and image of D86S TkA (light blue) and l-asparagine ligand (cyan) at 10ns. The part shown in yellow here represents the hairpin: Residues: β 2(14-16) - β 3(20-22). Blue regions indicate the allosteric site: Residues: α 8(293-306) - β 15(258-262). Green regions α 1(27-33)– α 4(122-133) are regions that tend to move towards the active site during ligand and enzyme interaction.	81
Figure 4.32. A) Radius of gyration plot for D86S TkA and its ligand l-asparagine with two repeats. A) Average radius of gyration plot for D86S TkA with l-asparagine ligand. C) Comparative radius of gyration for D86S TkA and WT TkA.	82
Figure 4.33. The number of hydrogen bonds between atom pairs within 0.35 nm distance.	83
Figure 4.34. The images show hydrogen bond formation between l-asparagine ligand and some residues that are in active site of D85S TkA.	84
Figure 4.35. A) SASA plot for D86S TkA and asparagine ligand with two repetitions. B) Average SASA plot for D86S TkA and l-asparagine ligand. C) SASA plot for comparison of D86S TkA and WT TkA.	85
Figure 4.36. A) WT TkA and l-aspragine ligand interaction energy plot. B) D86S TkA and l-asparagine ligand interaction energy plot. Red regions represent LJ-SR and black regions represent Coul-SR energy.	86
Figure 4.37. A) shows comparison of WT TkA RDF values and D86S TkA RDF values. B) RDF analysis is given specifically for residues for WT TkA. C) RDF analysis is given specifically for residues for D86S TkA.	87

<u>Figure</u>	<u>Page</u>
Figure 4.38. A) RMSD graphs of T11E TkA two replicates. B) T11E TkA average RMSD graph obtained from two repetitions. C) RMSD analysis; Comparison with WT TkA and T11E TkA with l-asparagine ligand.	89
Figure 4.39. A) RMSF graphs of T11E TkA two replicates. B) T11E TkA average RMSF graph obtained from two repetitions. C) RMSF analysis; Comparison with WT TkA and T11E TkA.	90
Figure 4.40. Initial position image of T11E TkA (light pink) and l-asparagine ligand (red) and image of T11E TkA (light blue) and l-asparagine ligand (cyan) at 10ns. The part shown in yellow here represents the hairpin: Residues: β 2(14-16) - β 3(20-22). Blue regions indicate the allosteric site: Residues: α 8(293-306) - β 15(258-262). Green regions α 1(27-33)– α 4(122-133) are regions that tend to move towards the active site during ligand and enzyme interaction.	91
Figure 4.41. A) Radius of gyration plots for T11E TkA and ligand l-asparagine with two repeat studies. B) Average RMSF plot for T11E TkA and l-asparagine ligand. B) Comparison of radius of gyration WT TkA and T11E TkA.	93
Figure 4.42. The number of hydrogen bonds between atom pairs within 0.35 nm distance.	94
Figure 4.43. The images show hydrogen bond formation between l-asparagine ligand and some residues that are in active site of T11E TkA.	95
Figure 4.44. A) SASA plots of T11E TkA and its ligand l-asparagine is shown with two repeats. B) Average SASA plots of T11E TkA and ligand l-asparagine. C) WT TkA with l-asparagine and T11E TkA with l-asparagine ligand comparison via SASA plots.	96
Figure 4.45. A) WT TkA and l-asparagine ligand interaction energy plot. B) T11E TkA and l-asparagine ligand interaction energy plot. Red regions represent LJ-SR and black regions represent Coul-SR energy.	97
Figure 4.46. A) RDF analysis between WT TkA's ligand and its active site as well as T11E TkA's ligand and its active site. B) shows RDF analysis between some important residue on the active site and WT TkA' ligand. C) shows RDF analysis between some important residue on the active site and T11E TkA' ligand.	98
Figure A.1. Parameter file context of energy minimization step "em.mdp".....	110
Figure A.2. Parameter file contents of the NVT ensemble step "nvt.mdp".....	111
Figure A.3. Parameter file context of NPT ensemble step "npt.mdp"	112

<u>Figure</u>	<u>Page</u>
Figure A.4. Parameter file context of NPT ensemble step "md.mdp"	113
Figure A.5. Parameter file context of "ie.mdp" that	114
Figure A.6. An example topology file content of protein in water simulation.....	115
Figure A.7. Topology file content of protein and ligand complex MD simulations.....	116
Figure D.1. Apo WT TkA in water MD simulation with initial parameter plots.....	120
Figure D.2. WT TkA with ligand, MD simulation with initial parameter plots.....	121
Figure D.3. T55E TkA and L-asparagine ligand initial parameter plots.....	122
Figure D.4. D86S TkA and L-asparagine ligand initial parameter plots.....	123
Figure D.5. T11E TkA and L-asparagine ligand initial parameter plots.....	124

LIST OF TABLES

<u>Table</u>	<u>Page</u>
Table 2.1. Active site residues for TkA and EcA. Prime indicates neighbor subunit that is B chain in this case.....	20
Table 3.1. GROMACS file formats (extensions) and their tasks.....	31
Table 4.1. L-asparaginase active site residues from different strains. The apostrophe at Tyr233 and Glu275 indicates active site residue from the B chain.....	37
Table 4.2. Best docking scores for rigid docking studies on the TkA. Studies were performed as 10 runs, 20 runs, 50 runs and 100 runs.....	40
Table 4.3. Flex docking results according to the selection of active site residues in different combinations. -5.87 kcal/mol is chosen as best score.....	41
Table 4.4. Docking scores of WT TkA, WT EcA and mutant TkAs. T11E, T55E and D86S are used for MD simulation candidates.....	44
Table 4.5. RMSF values of active site residues before ligand binding (Apo State) and after ligand binding.....	56
Table 4.6. Overall Results from MD studies and comparisons. According to the RDF value, the change in the mutation region is indicated as increase or decrease. The NA value corresponds to the values for which the data in the table is not available.....	100

CHAPTER 1

INTRODUCTION

L-asparaginase (L-ASNase) is an enzyme that catalyses the hydrolysis of L-asparagine (L-asn) into aspartic acid (L-asp) and ammonia (NH_3) in figure 1.1.A. There are two common mechanisms for this reaction. First mechanism is called two step substitution. the catalysis of hydrolysis of L-asparagine starts with a nucleophilic attack to carboxamide carbon of L-Asn and forms a tetrahedral intermediate (TI) in figure 1.1.B. While this step, an ammonia (NH_3) molecule is released. Because of release of NH_3 , this step is irreversible and ends with the formation of acyl-enzyme intermediate (AEI). In second step, a water molecule substitutes and with the formation of another TI in figure 1.1.C, AEI dissolves and the second step ends with L-ASNase and L-asp. The second mechanism is one-step and nucleophilic attack of H_2O and escape of NH_3 directly happens at the same time. Both mechanisms are represented figure 1.1 below (Lubkowski et al. 2020).

Generally, L-ASNase has tetrameric structure which is formed by dimer of two intimate dimers. The whole tetramer structure has a complete active site formed by two unfunctional active sites from intimate dimers. Also, the flexible loops in the active site have important roles in the function of the enzyme. When active site flexible loops are changed, active site also becomes unfunctional (Lubkowski et al. 2020).

The general quaternary structure, 2D secondary structures and amino acid sequences of *Escherichia coli* L-asparaginase (*Eca*) are shown in figure 1.2 and 3D tetramer structure and active pocket structure are shown in figure 1.3 below.

L-ASNase was first found in bovine tissue and its enzymatic activity was observed on ammonium and L-aspartate by Lang in 1904. Following him, Furth and Friedmann observed L-ASNase enzymatic activity in horse and pig tissues and organs too in 1910. They concluded the amount of the enzyme in these mammals are similar. Later, Clementi showed that the enzyme amount in different mammals is changing in 1922. Clementi's research also drew attention for the first time to the anticarcinogenic effect of guinea pig serum. The both enzyme amount difference and anticarcinogen

effects of guinea pig serum threw light on the research conducted to discovery of antilymphoma effects of L-asparaginase (Egler, Ahuja, and Matloub 2016).

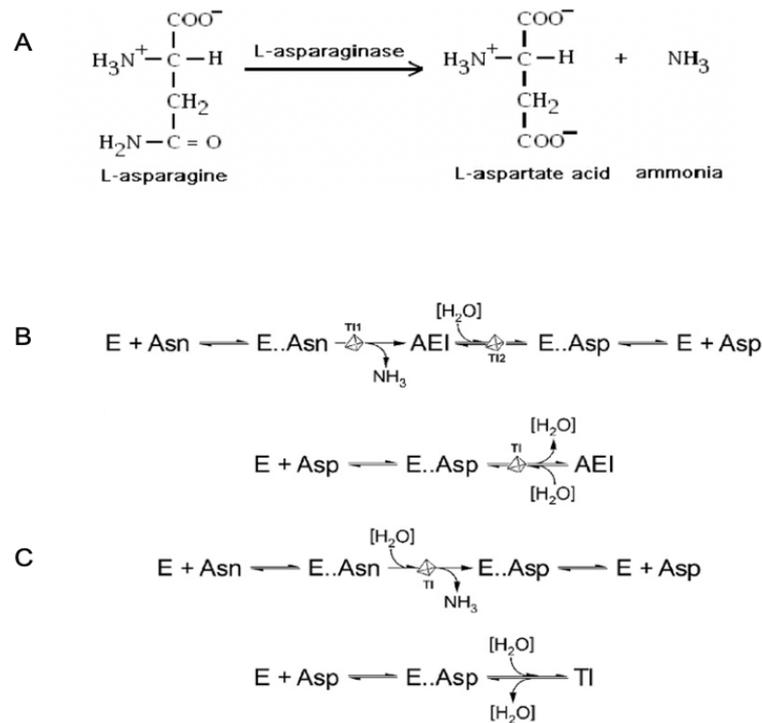


Figure 0.1. Two different basic mechanisms are known for the L-ASNase enzyme. (A) Two step substitution mechanism (B) One step mechanism (Lubkowski et al. 2020).

The discovery of L-ASNase antilymphoma effects dates to 1950s starting with the trials of several sera samples to test their effectiveness on animals with transplanted lymphomas. Guinea pig serum, rabbit serum, and horse serum were injected to the subcutaneous lymphoma transplanted mouse and rats to observe the effect of sera on progressing lymphoma. Guinea pig serum injected mouse and rats were seen healthy, and not show illness signs. But untreated control group and other groups treated with rabbit serum and horse serum were not seen as healthy, and the subcutaneous lymphoma were enlarged (Egler, Ahuja, and Matloub 2016). Kidd's following study suggested that the active agent is a protein based on the experiments (Egler, Ahuja, and Matloub 2016).

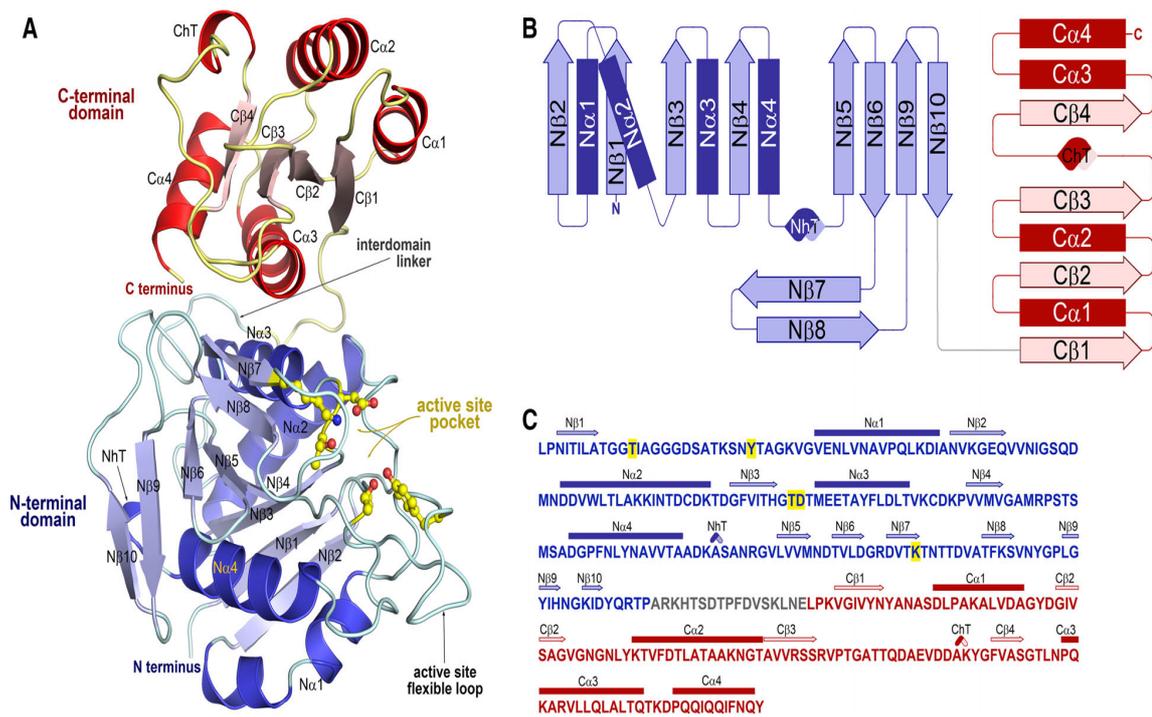


Figure 0.2. Representations of the common structure model for L-ASNase from *Escherichia coli*, A) Quaternary, structure is obtained from *Escherichia coli*, B) 2D secondary structures from *Escherichia coli*, C) Amino acid sequences that is obtained from *Escherichia coli* (Lubkowski et al. 2020).

In 1960s, following research by Broome focused on what exactly the active antilymphoma agent in guinea pig serum is. To uncover this agent, the researcher used the information from Clementi's study and hypothesised L-ASNase might be the antilymphoma agent. Broome identified properties of L-ASNase and designed experiments to extract L-ASNase from guinea pig serum. Later he used the serum containing L-ASNase, the serum without L-ASNase and extracted L-ASNase. Broome's results were showed that L-ASNase is the active antilymphoma agent. (Egler, Ahuja, and Matloub 2016) From that time, L-ASNase as a drug against lymphoma forms and especially acute lymphoblastic leukaemia (ALL) have been developing. Currently for three types of ALL therapeutic L-ASNases are effectively used. The first one was found and isolated from EcA very rapidly after Broome's discovery. Before finding L-ASNase in EcA, researchers isolated L-ASNase from guinea pig serum and then they applied their

methodology to isolate L-ASNase from *EcA* and *Bacillus coagulans* (Salzer et al. 2013; Egler, Ahuja, and Matloub 2016).

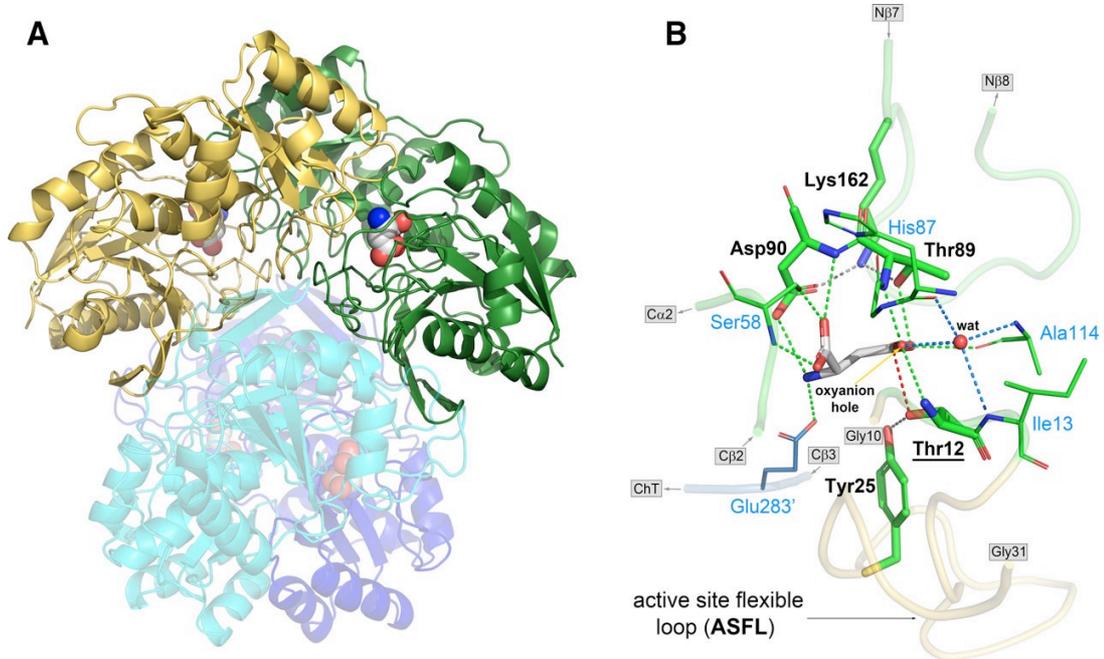


Figure 0.3. Most common structure of tetramer L-ASNase in A and their active sites in B that are obtained from *Escherichia coli* (Lubkowski et al. 2020).

They also tested how both isolates are effective against lymphatic leukaemia and they found *E. coli* L-ASNase is effective but *B. coagulans* is not effective against lymphatic leukaemia. In the following two years, first clinical trial of L-ASNase on human was done and concluded as even though it is effective on human ALL, it requires further research because of a side-effect, haemolysis. Meanwhile, like Mashburn and others, more researchers were looking for the alternative sources of therapeutic L-ASNase and the one isolated from *Erwinia chrysanthemi* (*E. chrysanthemi*) is found to have that having more enzymatic activity than the L-ASNase from *E. coli* (Salzer et al. 2013; Egler, Ahuja, and Matloub 2016).

Trial studies showed that, hypersensitivity was raising in the patients especially for *E. coli* L-ASNase, but the same study showed that there was no cross-sensitivity for *E. chrysanthemi* which was pointing out that, there should be more studies about the

enzyme and its source, or new developments to prevent hypersensitivity (Zalewska-Szewczyk et al. 2009). Hypersensitivity derived from antibodies and the possible preparation step derived antibodies are still a question for drug development, and evaluation studies. Following it, FDA approved *E. coli* L-ASNase as an ALL-therapeutic agent in 1978. But hyperactivity and immunogenicity remained as a problem and a research topic for researchers. In 1981, researchers modified *E. coli* L-ASNase with monomethoxy polyethylene glycol to make an initial pegylated *E. coli* L-ASNase (PEG ASNase). Their research outcomes showed that PEG L-ASNase has very low immunogenicity against L-ASNase anti-bodies. Besides these profits, PEG L-ASNase persisted more than three weeks after a single injection (Kamisaki et al. 1981).

In 1975, 1978, 1983, 1999, and 2005 clinical trials were conducted among USA and Europe and results enlighten important information about dosage of native *E. coli* L-ASNase and *E. chrysanthemi* L-ASNase. Within this period, *E. chrysanthemi* is recognised as an ALL-therapeutic agent in United Kingdoms, PEG L-ASNase is recognised as an ALL-therapeutic agent by FDA in 1994 and it is recognised as first-line treatment for ALL by FDA in 2006, later *E. chrysanthemi* derived L-ASNase also recognised as ALL-therapeutic agent by FDA in 2011 (Salzer et al. 2013).

Besides direct usage of L-ASNase in cancer treatment, it is also found that it can be used as preventive agent against cancer in food industry. Above 120°C and in low humidity, asparagine reacts with reducing sugars and turns into acrylamide (AA). This is because carbonyl group of sugar reacts with amino group of the amino acid and produces glycosylamine and depending on the reactants, it leads further polymers.

This reaction is called Maillard reaction (Maillar et. al 1912). Also known browning because it gives the fried potato, grilled meat, and baked bread with a flavour. But acrylamide is a known carcinogen and both European Food Safety Authority (EFSA) and World Health Organization (WHO) points out that the daily uptake of acrylamide in food, is a health risk. WHO reported crisp breads may contain AA between 50 to 100 µg/kg and EFSA set the limit of daily uptake of AA to 170 µg/kg body weight in 2015.(Covino et al. 2023)To lower the risk, L-ASNase is recruited and applied in food production process to lower the asparagine amount in food and to prevent acrylamide formation during frying, baking, and cooking. For fried potatoes, first results were unsatisfactory because of the heat factor but, it is overcome by using thermostable L-ASNase. Initially, asparagine reduction rate was around approximately 22% and respectively 56%. Following usage of thermostable L-ASNase increased this reduction

rate to 80% compared with the controls. For bread applications and the research showed that the asparagine and acrylamide rate is reduced up to level of 78% to 86% respectively. For coffee this rate reached to 80.7% respectively. Later, because of the nature of enzymatic activity, immobilized L-ASNase was also introduced to food industry. Immobilized L-ASNase showed a rate of 90% reusability which is very beneficial for industry, but asparagine and acrylamide reduction rate was only up to 60% (Jia et al. 2021). The research on eliminating this possible risk of cancer in food industry is requiring more development but the contribution of L-ASNase is indisputable.

In recent years, extensive research on L-ASNase has led to increased knowledge about various strains. Among these, the *Pyrococcus horikoshii* L-ASNase (PDBID:1WNF), *Helicobacter pylori* L-ASNase (PDBID:2WT4), and *Thermococcus kodakarensis* L-ASNase (PDBID:5OT0) have been identified. The primary goal of using L-ASNase from different strains is to reduce allergic reactions and decrease sensitivity in the body.

1.1. Overview of Acute Lymphoblastic Leukemia (ALL) and its treatment with L-Asparaginase

As time progresses, the global impact of cancer becomes more pronounced. Acute Lymphoblastic Leukemia (ALL), which is one of the various types of cancer. It arises from the uncontrolled excessive production and accumulation of lymphoid progenitor cells. It is a malignancy of white blood cells (Baral et al., 2021). ALL has a significant effect on young children, particularly those aged between 1 and 4, as well as on adolescents and young adults (González-Torres et al. 2020).

According to the World Health Organization (WHO) data, leukemia and its subtypes have been studied as a group. Considering the most recent WHO data, leukemia, including all age groups and both genders, ranks 13th among the most frequently observed cancer types worldwide, with a new case count of 474,519 in the year 2020. Similarly, based on 2020 data, encompassing both genders and all age groups, leukemia ranks 10th among all cancer types globally, resulting in the death of 311,594 patients. When assessing data for both genders combined, leukemia's incidence and mortality rates appear to be higher in male children and within the 0-5 age range for both genders.

Furthermore, as per the 2020 WHO data, the country with the highest prevalence of the disease is Northern America (Sung et al., 2021).

The 2023 studies by the American Cancer Society, focusing on acute lymphoblastic leukemia in children and adolescent age groups in the United States, out of 6,540 new cases, 3,660 are male and 2,880 are female. The mortality rates are 700 for male children and 690 for female children. Assuming the disease continues to affect every 5,000 individuals, it is projected that the incidence will increase from 80.5 thousand people in 2020 to 81.8 thousand in 2025, and 82.5 thousand in 2030 (Siegel et al. 2023).

Based on all these statistical data, numerous children requiring treatment for ALL will continue to exist not only in the present day but also in the future. To administer the correct treatment, understanding the underlying biochemical mechanisms is highly crucial. While healthy cells can synthesize the amino acid L-asparagine with the assistance of the L-asparagine synthetase enzyme, Acute Lymphoblastic Leukemia (ALL) cells are deficient in this enzyme or can synthesize it in small amounts. Due to this deficiency, L-asparagine becomes an essential amino acid for ALL cells and must be obtained from the extracellular environment. In the absence of this amino acid, cancer cells will lose a critical growth factor and will not be able to sustain their survival. Based on all this information, the enzyme l-asparaginase is used to deprive ALL cells of the amino acid l-asparagine. Mechanism of L-ASNase enzyme is seen in figure 1.4.

By inducing depletion of asparagine in the bloodstream with the help of l-asparaginase, their proliferation is significantly hindered (Derst et al., 2000).

The exploration for new production methods of L-asparaginase underscores its critical role in clinical treatments.

In silico exploration, particularly in the realm of bioinformatics, aims to dissect and comprehend the intricate structure, functionality, and interactive dynamics of enzymes like L-asparaginase. These computational explorations shed light on the enzyme's specific interactions with its substrates and its catalytic efficiencies. Techniques like molecular docking, dynamic simulations, and structural predictions are instrumental in demystifying the interactions of L-asparaginase with other molecules, paving the way for the innovative engineering of enzyme variants.

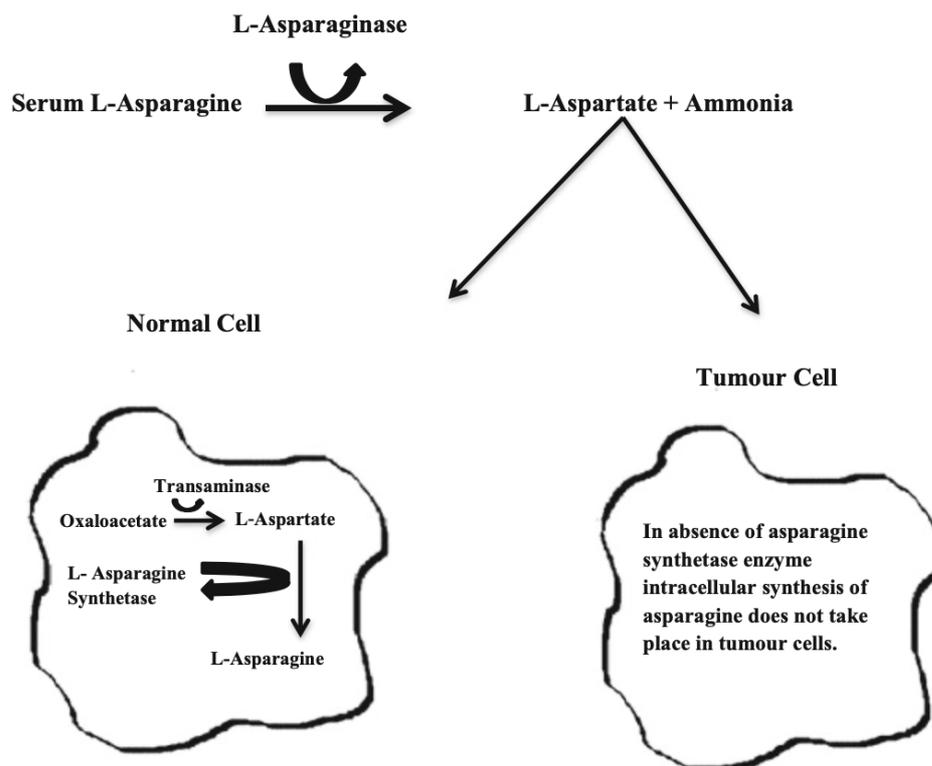


Figure 0.4. Mechanism of L-asparaginase enzyme into the blood stream and effect on the tumor cells (Batool et. al., 2016).

1.2. Molecular Docking Studies

The rapidly growing field of bioinformatics, along with its integration into wet lab experiments, accelerates experimental processes and saves time in research. Molecular modeling methods, commonly employed in drug development studies, enable a higher number of experiments in shorter timeframes.

X-ray crystallography, nuclear magnetic resonance (NMR) spectroscopy, electron microscopy, and other experimental techniques are commonly employed to determine the 3D structures of biomolecules. For instance, in X-ray crystallography, a crystal of the biomolecule is exposed to X-ray radiation, and the resulting diffraction pattern is used to reveal the distribution of electron density. This information is then used to construct the atomic model of the molecule. All these three-dimensional structures obtained from NMR and X-ray crystallography are combined by repeatedly visualizing them from various

angles using different imaging methods. They are then uploaded to the Protein Data Bank data system. Files can be accessed in various file formats such as ".pdb," and computer-based editing and rational designs are carried out. These methods can be used to explore information about the distances between atomic nuclei that could be used to reconstruct the structure. However, molecular docking methods are used to identify the binding energies and locations between molecules (Tao et al. 2020; Vranken 2007).

Molecular docking studies are frequently utilized, particularly in the field of computer-aided drug design. Through this method, even the interactions of small ligands with proteins are examined in three dimensions, and binding energies are calculated. In recent years, molecular binding investigations have gained even more significance as they explore the flexibility of protein-ligand interactions and the various conformations occurring during binding. Throughout the study, the flexibility and relaxation of the receptor enhance more accurate pose predictions and enable a more realistic examination of bindings (Jakhar et al., 2020).

The straightforward mechanism of molecular docking relies on providing a score known as the binding energy score, which expresses the intensity of the interaction between the ligand and the protein. Furthermore, it enables the prediction of the binding site's location. Molecular docking studies shed light on van der Waals bonds, hydrogen bonds, hydrophobic interactions, and electrostatic attraction interactions (Tao et al. 2020).

Hydrogen bonding, a weak electrostatic attraction formed between a hydrogen atom and an electronegative atom, plays a pivotal role in various chemical and biological systems, contributing to molecular shape changes, chemical reactions, and biological recognition processes. Van der Waals bonds, on the other hand, represent weak attractive forces between molecules, arising from transient positive and negative charge distributions due to electron movements. These bonds contribute to holding molecules together and stabilizing chemical compounds. Hydrophobic interactions occur between molecules with hydrophobic (water-repelling) regions that do not interact favorably with water. These interactions result from the tendency of hydrophobic regions to minimize contact with water. Electrostatic attraction interactions rely on the attraction between molecules due to their opposite electric charges. Positive-charged atoms or molecules are attracted to negative-charged atoms or molecules. Both hydrophobic interactions and electrostatic attraction interactions assist in arranging molecules into specific conformations and enabling proteins to perform their biological functions (Tam et al. 2022).

The strength of all these forces is crucial in evaluating the affinity between the receptor and the ligand. These forces occur in various conformations during molecular docking. The binding site continuously changes until the most stable binding is achieved. Both spatial and energetic compatibility between the protein and the ligand is essential. Therefore, determining the optimum number of runs is crucial in molecular docking studies (Tao et al. 2020; Meng et al. 2012).

The flexibility of the receptor is crucial in molecular docking, and accordingly, molecular docking studies are divided into three categories. The first of these is "rigid docking." It is the simplest method and does not require multiple calculations. The docking process proceeds with the protein and ligand structures remaining unchanged. The second method is called "semi-flexible docking." In this docking approach, the protein structure remains fixed while flexibility is introduced to the ligand structure. The docking process considers critical regions of the ligand, providing a more comprehensive exploration. In some cases, the opposite situation occurs where critical regions of the protein are considered, and the protein's flexibility is emphasized. This leads to the third and final type of docking, known as "flexible docking." In this system, flexibility is recognized for both the protein and the ligand, and calculations are performed. Therefore, it represents the most complex calculations in docking studies. Additionally, docking work can be carried out in aqueous and non-aqueous environments. While aqueous environment makes biological systems more realistic due to solvent modeling and the ions contained in the solvent, docking studies carried out in an aqueous environment take much longer (Raval & Ganatra, 2022; Tao et al., 2020).

There has been significant development in the field of molecular docking, resulting in the creation of over 60 different docking tools and programs. Some of the notable ones include DOCK, AutoDock Tools, FlexX, Surflex, GOLD, ICM, Glide, Cdocker, LigandFit, MCDock, FRED, MOE-Dock, LeDock, AutoDock Vina, rDock, UCSF Dock, and several others. In these docking software programs, the strategies for ligand placement vary algorithmically from one another. However, except for the GOLD program, almost all programs allow for flexible ligand placement while keeping the receptor rigid. Each of them strives to achieve the highest binding score for the most accurate receptor-ligand conformation. In general, these docking programs are known to be able to predict experimental poses with mean square root mean deviations (RMSDs) of 1.5 to 2 Å (Pagadala, Syed, and Tuszynski 2017; Bissantz, Folkers, and Rognan 2000).

1.3. Molecular Dynamics Simulations

Molecular dynamics (MD) simulations, like molecular docking studies, are becoming increasingly important, especially in the fields of drug discovery and structural biology. Through molecular simulations, the atomic-level behaviors of proteins and other biomolecules can be captured with high precision, over specific time intervals, and at high resolution (Hollingsworth and Dror 2018; Hospital et al. 2015).

MD simulations were first experimented with simple gases in the late 1950s. By the late 1970s, MD simulation studies had expanded to include proteins. Ultimately, in 2013, they gained further recognition with the Nobel Prize in Chemistry, becoming even more popular. In general, MD simulations predict how atoms in a protein, or any molecular system will move over time. Additionally, MD simulations shed light on various biological processes, such as protein folding, ligand-receptor binding, biomolecule mutations, protonation, phosphorylation, or conformational changes, revealing how a molecule will react. Simulations can be examined with femtosecond temporal resolution (Hollingsworth and Dror 2018).

Investigating protein dynamics relies on observing different mechanisms of action. It is highly dependent on systems and environments on different time scales resulting from the movements of proteins. Examination of covalent bond vibrations at subangstrom levels shows the fastest of movements. It is very important to understand the mechanism of action of rotamers of side chains positioned in the protein backbone at the Angstrom level. During protein ligand binding events, ligands deeply embedded in the protein structure can travel angstroms at high rates. Therefore, studies performed at nanosecond levels are ideal for examining protein movements and behaviors. Therefore, studies at nanosecond levels are important for in-depth observation. Even microsecond studies can be performed. But microsecond and macrosecond levels are quite time consuming and costly. There is a visual to understand MD simulations in terms of time scale in figure 1.4 (Surpeta, Sequeiros-Borja, and Brezovsky 2020).

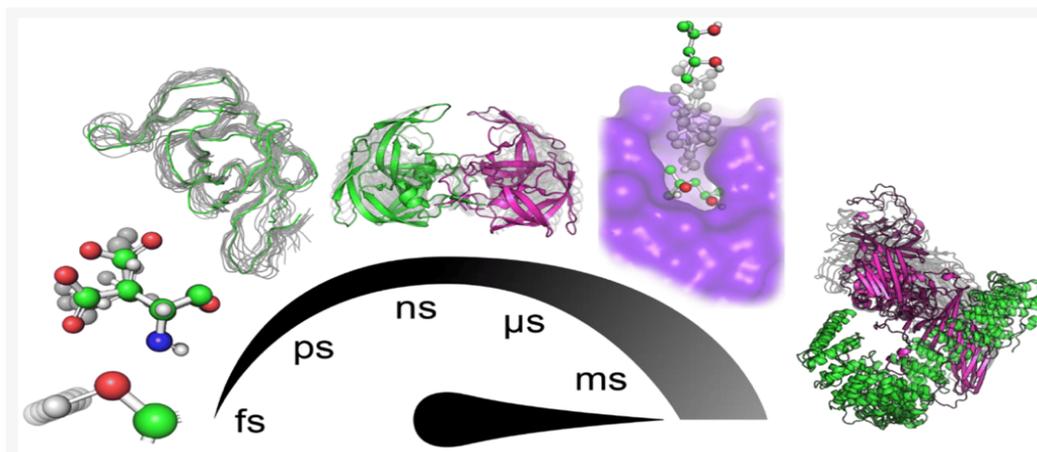


Figure 0.5. Protein dynamics: a hierarchy of major motions. Structural conformations and protein-ligand interactions obtained from molecular dynamics simulations have been visualised. Processes have been visualised from femtoseconds to milliseconds. For each scale, typical structural changes and interactions are shown (Surpeta, Sequeiros-Borja, and Brezovsky 2020).

From another perspective, the increasing popularity of MD simulations can be attributed to two main factors. First, they significantly bolster experimental studies with robust results. In fact, due to their reliance on physical modeling, MD simulations have started to provide even more accurate results compared to experimental data, despite offering approximate predictions. Second, the recent hardware advancements, particularly the availability of well-equipped and accessible Graphics Processing Units (GPUs), have contributed to the growing interest in MD simulations (Hollingsworth and Dror 2018).

GROMACS, AMBER, Nanoscale MD, OpenMM, CHARMM-GUI are widely used popular MD simulation tools. They mostly work with command prompt.

In simple terms, the logic behind MD simulations involves calculating the forces that biomolecule atoms apply to each other in a watery or lipid bilayer environment, considering their positions. These forces are then used to update the positions and velocities of each atom. All the rules are based on Newton's laws of motion. As a result, an orbit is generated, and throughout this orbit, the configuration of the biomolecule is defined at the atomic level at each point, yielding a three-dimensional movie. Parameters used throughout the process, such as the initial molecule or ligand, are known, which

ensures clearer and more reliable results (Hollingsworth and Dror 2018; Hospital et al. 2015).

MD simulations provide answers to questions that help us understand proteins. These can be summarized as follows:

1. Reflecting the mobility in various regions of biomolecules while assessing their flexibility. Experimental studies like X-ray crystallography are crucial for structure determination, but they yield an average structure. Examining the simulation of these structures measures how the biomolecule moves in equilibrium and what kinds of structural fluctuations it undergoes. Simulations also frequently allow the modeled structure to relax to a more suitable conformation, if available.
2. Furthermore, MD simulations assist in explaining the behavior of salt ions and water molecules during protein function or ligand binding. An example image is seen figure 1.6 (Hollingsworth and Dror 2018; Meng et al. 2012.; Hospital et al. 2015; Lemkul 2019).

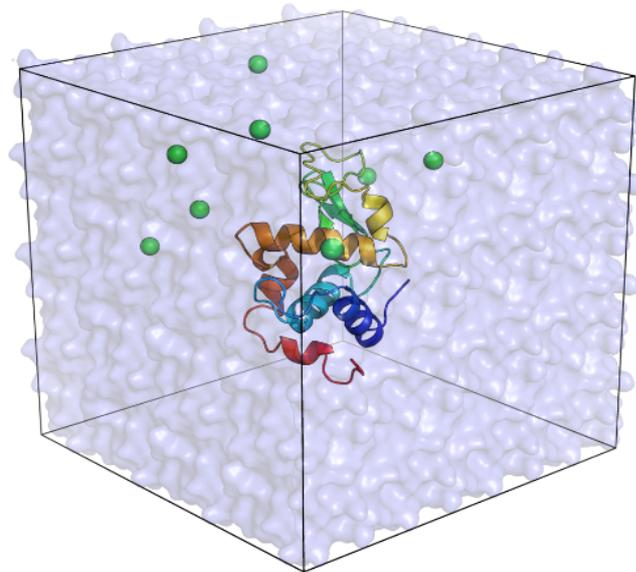


Figure 0.6. General box environment during MD simulation. Protein molecule simulation in water (Lemkul 2019).

1.3.1. Molecular Dynamics Simulations for Protein Systems

Various force fields are available for MD simulations, such as GROMACS AMBER, and CHARMM. The choice of each force field depends on the field of study and the expected outcomes of the simulation. These force fields provide insights into protein-ligand complexes, protein motion, and flexibility, contributing to the understanding of their interactive dynamics.

After determining the protein force field and solvent parameters, intermediate steps such as periodic boundary conditions (PBC), Ewald summation techniques, particle mesh Ewald method, thermostats parameters, solvent models, and energy-minimization methods are used to manipulate the system, adjusting the desired parameters and properties for the protein-ligand complex system (Singh 2020).

These parameters are explained as follows.

1.3.2. Periodic Boundary Conditions (PBC)

Periodic Boundary Conditions (PBC) in MD simulations aim to confine the movement of atoms within defined boundaries. This ensures that atomic mobility remains within specific boundaries throughout the simulation period. Otherwise, the simulation box would move infinitely in all directions, resulting in the formation of a lattice. Isotropic Periodic Sum (IPS) is utilized when dealing with investigations into long-range interactions. This method prevents the occurrence of unnecessary symmetry artifacts caused by PBC. It allows for the determination of the self-diffusion coefficient with a cutoff radius greater than 2.2 nm and can be applied in any functional form for the potential energy (Singh 2020).

1.3.3. Ewald Summation Techniques

MD simulation techniques make use of the Ewald summation method to compute long-range Coulomb interactions. These long-range interactions are estimated as sums that converge very gradually (Singh 2020).

1.3.4. Particle Mesh Ewald Method

In this method, the potential energy can be examined into two summation parts: One of them is Ewald's standard direct sum. Other one is reciprocal sums. It utilizes classical Gaussian charge distributions. Direct sums use direct cut-offs, while reciprocal sums are determined with Fast Fourier Transform (FFT) on a grid. The grid charges interpolate at points of grid. Finally, forces are computed using analytical differentiation (Singh 2020).

1.3.5. Thermostats

In the MD simulation system, energy is added to and subtracted from the boundaries, while constant volume, particle count, and specified temperature are maintained for the canonical ensemble. This is the section where the energy exchanges occurring in endothermic and exothermic processes are examined (Singh 2020).

1.3.6. Solvent Models

MD simulations require realistic solvent environments that best mimic real cell-like conditions. There are four main types of solvent models, with the majority being

aqueous environments: SPC/E, TIP3P, TIP4P, and TIP5P. Each model is optimized for the diffusivity, radial distribution, and density anomalies of water.

1.3.7. Energy-Minimization Methods

In the MD simulations, another crucial approach is energy minimization techniques. To begin, it's essential to identify low-energy regions, and this is achieved through the grid search method and the zero-degree method function for energy minimization. Regarding energy minimization, the steepest descent method and the conjugate gradient algorithm methods are the initial steps, both being derivative-based techniques. In the second derivative stage, the Newton-Raphson algorithm is employed. The steepest descent method is particularly useful for refining the geometry of proteins and improving contact points by implementing minor structural modifications. It guides the system down from the highest point on the potential energy surface. On the other hand, the conjugate gradient algorithm methods make use of both current and existing gradient information. Each of these methods carries its own set of advantages and is suited to specific scenarios within molecular dynamics simulations (Singh 2020).

CHAPTER 2

LITERATURE REVIEW

The L-ASNase enzyme, which is often the subject of various studies, is generally aimed to be purified and its immunogenic effects minimized to reduce allergic reactions and hypersensitivity, including anaphylactic shock, in patients. Prolonged use of this enzyme can lead to antigenic effects in the patient's body. Considering all these reasons and its beneficial effectiveness, the exploration and development of new l-asparaginase strains hold significant importance (Mohideen 2020).

However, since wet lab studies are costly and time-consuming, interest in *in silico* studies have increased over time, especially in redesign studies in the field of proteins.

In this thesis, which is based on a detailed investigation of the structure of *Thermococcus kodakarensis* L-asparaginase (TkA) enzyme and the examination of the behavior of this enzyme in relation to the L-asparagine ligand, various mutation experiments were also carried out in the active site residues of the TkA enzyme. Thanks to these mutations, the behavior of the enzyme and the ligand relative to each other was further examined under the influence of mutation. Especially in terms of *in silico* studies, it was aimed to close the knowledge gap in the literature with a detailed *in silico* examination of this enzyme (Guo et al. 2017).

According to literature sources, the three-dimensional structure of the TkA enzyme was revealed by *in silico* studies by Guo et al. However, the structural analysis of this thermostable enzyme was revealed by comparing it with the *E. coli* L-ASNase enzyme, which has been extensively researched from past to present. The rmsd values reflecting the similarity ratio of the A chains of *E. coli* L-ASNase I and II and the TkA enzyme are 1.11Å and 1.67 Å, respectively. Both structures have a hairpin structure close to the active site, which opens and closes during substrate binding, and segments that adapt to conformational changes in regions close to it. Figure 2.1 shows superimposition between TkA and *E. coli* l-asparaginases (Guo et al. 2017).

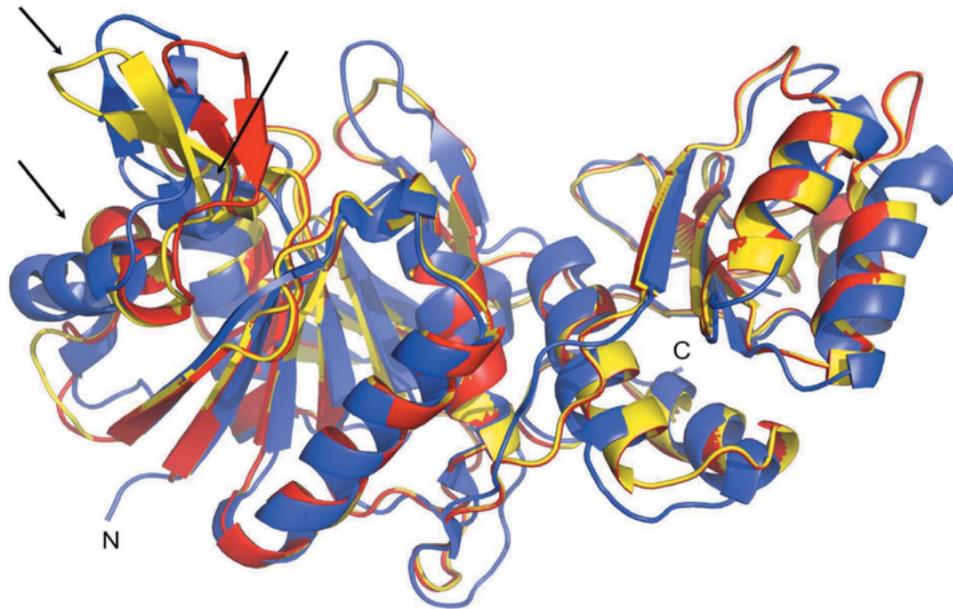


Figure 0.1. *E. coli* L-ASNase (PDBID:1NNS) and *TkA* 3D structure comparison via structural superposition (PDBID:5OT0). *TkA* chain A: red, *TkA* chain B: yellow and *E. coli* is indicated by blue color. Arrows indicate segments that adapt for different conformations while substrate (Guo et al. 2017).

EcA has two L-ASNase genes: *AnsA* and *AnsB*. *AnsA* gene-encoded L-ASNase I have a low affinity against l-asparagine while *AnsB* gene-encoded L-ASNase II has a high affinity against the l-asparagine ligand that's in figure 2.2. (Guo et al. 2017)

The most critical limitation is that glutaminase affinity of l-asparaginase. The enzyme can react with glutamine amino acid instead of asparagine and glutamic acid that is seen in figure 2.2 as chemical structure and reactions with l-asparaginase. These reactions reduce the drug's effectiveness and cause immune response as well as contribute to high cost. *E. coli* L-ASNase also has glutaminase activity. Thus, the scientist has shifted towards their research on thermostable l-asparaginases that have minimal or absent glutaminase activities (Orabi et al. 2019).

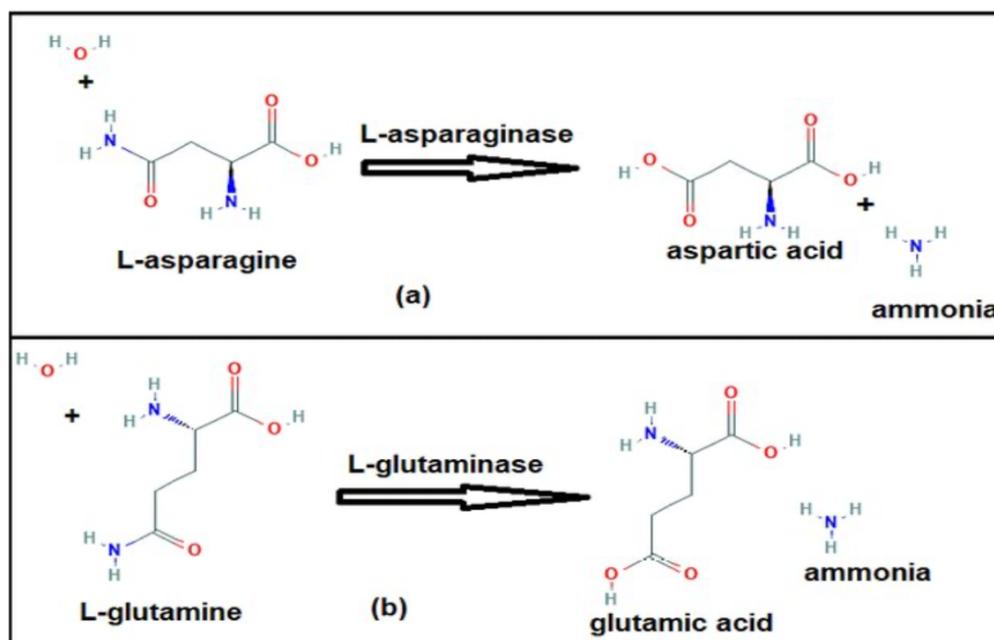


Figure 0.2. Schematic representation reactions of a) L-asparaginase, b) L-glutaminase (Orabi et al. 2019).

The ligand binding site, also known as the active site of TkA, is in a pocket formed by residues of two closely held dimer monomers. Beta-hairpin structure is composed of second and third beta-sheets. It is quite flexible, and they often lack electron density. The hairpin plays a role while substrate binding and catalysis as open and close states. Also $\alpha 1$ helix and $\alpha 4$ move towards the active site during enzyme activity. Furthermore, TkA has putative allosteric site like *E. coli* L-ASNase. The putative allosteric site is located between $\alpha 8$ helix and $\beta 15$ strand. The 3D structure of TkA enzyme is shown in figure 2.3 (Guo et al., 2017).

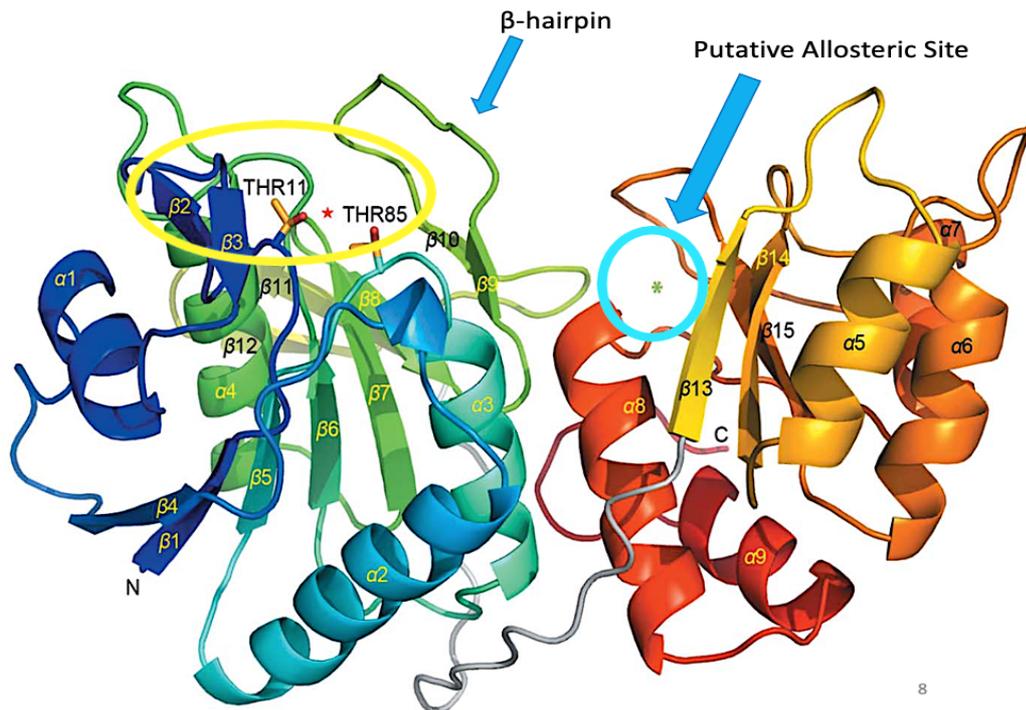


Figure 0.3. 3D structure of TkA enzyme with B-hairpin and putative allosteric site representations. Yellow circle indicates active region (Guo et al. 2017).

Table 2.1. Active site residues for TkA and EcA. Prime indicates neighbor subunit that is B chain in this case.

TkA	EcA
Thr11	Thr12
Tyr21	Tyr25
Ser54	Ser58
Thr55	Gln55
Thr85	Thr89
Asp86	Asp90
Lys156	Lys162
Tyr233'	Asn233'
Glu275'	Glu283'

Although the crystal structure of the *E. coli* enzyme has been defined in the literature according to the asparagine substrate, there is no study with the L-asparagine ligand in a similar region and among the residues for *TkA*. However, there are studies conducted with *TkA* and its ligand PO_4^{3-} . Therefore, the comparisons made for *E. coli* and *TkA* belong to these two regions. The asparaginase family, which has highly similar and conserved active sites, is also highly similar between *E. coli* L-ASNase and *TkA*. and Guo et. al. compared the two enzymes.

Thr12 and Thr89 are the residues that initiate the nucleophilic attack in this region in *E. coli*. During nucleophilic attack, Thr89 engages with Thr12 in a sequence of 'ping-pong' nucleophilic attacks. Tyr25 residue, one of the other active site residues, is also found in hairpin and plays a role as an intermediate element during the nucleophilic attack. The amino acids of the active site of *E. coli* and *TkA* enzymes are highly conserved, like other members of the asparaginase family. Active site residues of *TkA* and *E. coli* asparaginase are shown in table 2.1 and illustration is shown in figure 2.4.

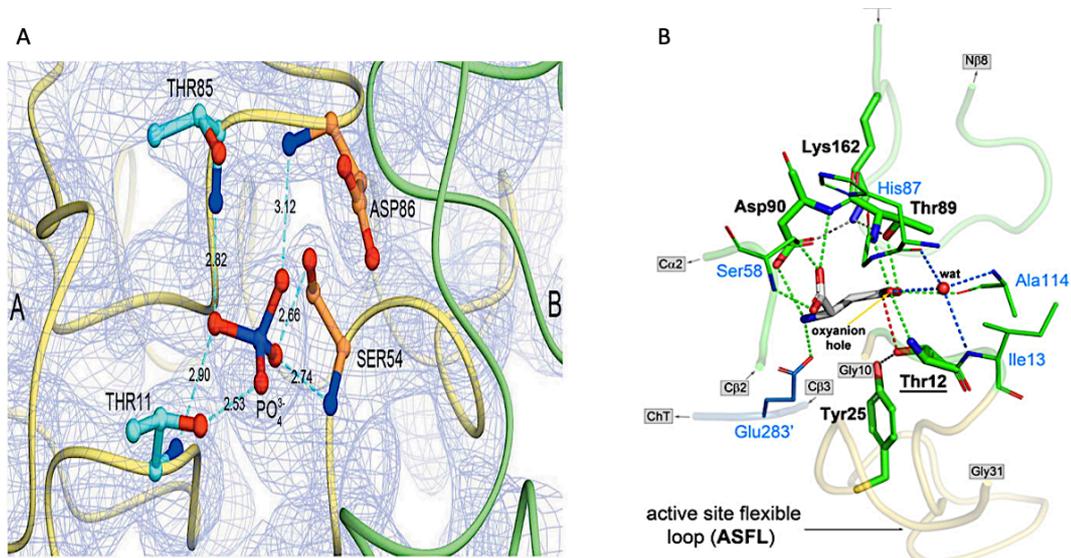


Figure 0.4. A) Active site residues for *TkA* with PO_4^{3-} . B) *E. coli* L-ASNase with L-asparagine ligand (Guo et al. 2017).

In the literature, studies on l-asparaginase obtained from various strains and their crystal structures have been investigated. However, there is still a lack of structural information, especially for TkA and its mutations. Therefore, the present study focuses on the molecular docking and molecular simulation methods to model TkA for obtaining structural insights. The aim is to investigate enzyme-ligand interactions and functional characterization of the enzyme. Examining their effects on the functional structure by trying various mutations and providing an overview of the asparagine ligand and its mechanism. This thesis is based on the investigation of the innovative l-asparaginase enzyme, and its mutations obtained from *TkA* strain through *in silico* molecular modeling and docking studies. The purpose of these analysis was to create reference ranges for *TkA*, which is relatively less studied in the literature, and to ensure the optimization and accuracy of *in silico* analyzes performed over the obtained reference ranges (Guo et al. 2017).

Throughout the studies, molecular docking experiments were conducted using Autodock Tools and Autodock Vina, along with UCSF Chimera for verification purposes via Vina's tools. The molecular docking efforts aimed both to understand the initial protein-ligand interactions and to use the resulting binding scores to compare with literature for selecting potential mutations. The HotSpot Wizard web server was instrumental in identifying 'hot spots' on TkA for potential mutations, crucial for tailoring enzymes' substrate specificity, activity, or enantioselectivity (Pavelka et al., 2009). Subsequently, MD simulations with GROMACS enabled a thorough analysis of mutant TkA, wild-type TkA, and TkA in water, shedding light on their dynamic behaviors and structural alterations. GROMACS is well-regarded for its high-performance molecular simulations and a broad spectrum of force fields, which makes it ideal for simulating complex biomolecular systems like TkA and its variants.

Simulations of mutant TkA, wild-type TkA, and TkA in water were performed to obtain graphs of RMSD (Root Mean Square Deviations), RMSF (Root Mean Square Fluctuations), gyration radius, Hydrogen bond, SASA (the Solvent Accessible Surface Area), protein-ligand interaction, RDF (Radial Distribution Function) and MSD (Mean Square Deviations).

RMSD is used to measure structural changes that a molecule or complex undergoes over time during molecular dynamics simulations. This measurement provides information about the overall stability and conformational changes of the system, allowing it to be seen how molecules and complexes develop structurally and which

binding modes are more stable. RMSD is based on superimposing the structures obtained from different frames of the studied structure and minimally translating and rotating the arithmetic mean of the corresponding atomic coordinates. This can be achieved when the superimposed structures reach the minimum root mean square deviation. Another approach here is to align the specified atom subsets (Wu and Wu 2010; Pitera 2014).

RMSF refers to the measure of the average positional fluctuations of atoms of molecules over a given period. With RMSF data, comments can be made about the flexibility of molecules. It can be used to understand the active sites and ligand binding sites of proteins. In RMSF analysis, it is possible to interpret how much each atom fluctuates positionally in a certain time interval and how much it deviates from the average position. Low RMSF values indicate that atoms or residues for proteins are more stable, that is, fluctuate less, in cases where they are considered. On the contrary, higher RMSF values indicate more fluctuations (Kumer et al., 2021; Ram Lamichhane & Ghimire).

The radius of gyration analysis evaluates the overall structural integrity and compactness of a protein or protein-ligand complex. It facilitates the understanding of changes in the protein's overall structure and the impact of ligand binding on the protein's structural integrity. Gyration radius generally describes how molecules expand or fold in space. A wider protein indicates that the protein is less compact, while a smaller gyration radius indicates a denser and more compact structure (Ambrose et al. 2022).

Hydrogen bonds (H-bonds) analysis is another method that is very important for understanding protein dynamics. Because hydrogen bonds are necessary for protein structure formation, molecular stabilization, protein catalytic mechanisms and physiological functions. It is a very fundamental research topic for the investigation of short-range interactions for molecular systems. Also, H-bond analysis is crucial for examining interactions between proteins and ligands. This analysis helps us understand the formation and durability of hydrogen bonds, which are fundamental for intermolecular recognition and binding strength. Regular monitoring of hydrogen bonds can indicate how stable a protein-ligand complex is and identify the key interactions that contribute to this stability (Jeyaram and Radha 2022; Fresch and Collini 2023).

The surface area of proteins and other macromolecules is measured using a technique called solvent accessible surface area, or SASA. The area where the molecule can make direct contact with solvent (which is frequently water) molecules is referred to as its surface area. Because SASA values reveal details on the solubility, interactions, and

functional status of proteins, they are significant in the structural biology and biochemistry of proteins. SASA, or Solvent Accessible Surface Area, is a measurement of the surface area of proteins or other macromolecules in contact with water. It provides important information about the hydrophobicity and hydrophilicity of proteins and molecules. Increasing SASA surface indicates that the molecule is less hydrophobic due to more contact with water molecules (Nakagawa and Tamada, 2021; Topham and Smith 2019; Durham et al. 2009).

Protein-ligand interaction analysis, including the effects of Coulombic and van der Waals forces, thoroughly examines how proteins bind with their ligands and the strength of these bonds. Such analyses assist in determining binding affinity and identifying the key amino acid residues involved in the binding process (Ambrose et al. 2022). Two separate features were discussed in the study: Lennard-Jones short-range (LJ-SR) and Coulombic short-range (Coul-SR) Potential terms. The LJ-SR term is a potential that models the repulsive and attractive van der Waals forces between atoms. The LJ potential includes two parts: one attractive (effective at long distances) and the other repulsive (effective at short distances). The repulsive part occurs when atoms get too close to each other, that is, at the point where the atoms' electron clouds begin to overlap. The attractive part arises from the weak attractive forces (van der Waals forces) between atoms or molecules at longer distances. In the text, the LJ-SR potential is often stated as a negative value, which may indicate a net attractive interaction. However, depending on the attractive and repulsive parts of the LJ potential, this may change. The term potential: Coulombic (Coul-SR) potential refers to the electrostatic interactions between charged particles and is defined by Coulomb's law. Whatever the sign of the charges, this potential can always be attractive or repulsive; Charges of the same sign repel each other (repulsive interaction), while charges of opposite signs attract each other (attractive interaction). In the text, the Coul-SR potential is expressed as the attractive interaction energy, meaning that the ligand has a strong attractive interaction with the protein (Londhe et al. 2019; Ambrose et al. 2022).

The RDF (Radial Distribution Function) serves as a measure in molecular systems to observe the spatial arrangement of particles relative to a reference point, and in protein-ligand systems, it helps understand the distribution of water molecules around specific amino acids or ligands, contributing to insights into the system's hydration and solvation. Essentially, RDF can also be described as the relative distribution of two molecules with respect to each other (Chen et al., 2008).

Monitoring the motion of a particle or molecule is another method of understanding and interpreting molecular dynamics systems. Thus, information is obtained about the diffusion of molecules or particles. In protein ligand systems, when this situation is examined for the ligand, information can be obtained about the diffusion of the ligand. However, dynamic systems are quite complex systems, and the movements of the ligand cannot follow a constant movement over time. Therefore, instead of investigating each step distance of the ligand, the square of the distance traveled by the ligand is calculated. Thus, during the diffusion calculation, instead of constantly obtaining negative and positive values, positive values are added and thus how the square distance increases can be interpreted (Samanta and Roccatano 2013).

With these comprehensive approaches, the structural and dynamic properties of mutant and wild-type TkAs have been extensively investigated, allowing the behavior of molecules to be examined in protein-ligand systems as well as in apo state protein simulations.

CHAPTER 3

MATERIAL AND METHODS

Molecular docking and molecular dynamics simulation studies were performed for wild type *Thermococcus kodakarensis* l-asparaginase (WT TkA) and its ligand asparagine, as well as TkA mutations, starting with sequence alignment and literature review. A diagram in which the studies are expressed in general terms can be seen in figure 3.1.

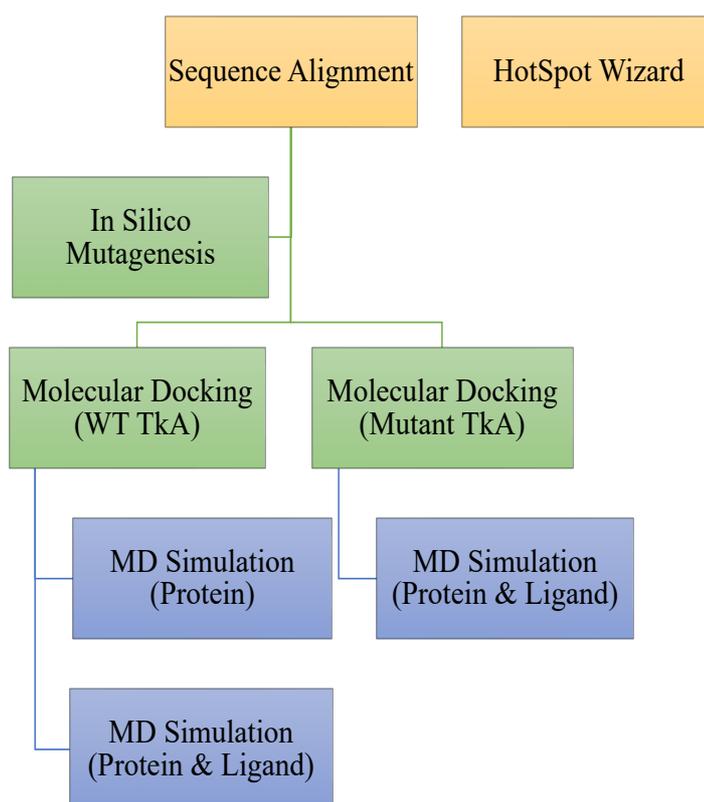


Figure 0.1. Flow chart for overall methodology.

3.1. TkA Structural Alignment

The behavior and conformational changes of *Thermococcus kodakarensis* type I thermostable l-asparaginase (TkA) and its l-asparagine ligand were examined. Additionally, different mutations were tried in the TkA active site residues. It was examined how TkA and l-asparagine ligand were affected by mutations while exhibiting conformational changes and behavior relative to each other.

TkA active site residues are known from the literature. Thr11, Tyr21, Ser54, Thr55, Thr85, D86, Lys156 residues from the A chain and Tyr233 and Glu275 residues from the B chain constitute the active site elements of TkA, which functions as a dimer.

In line with this information, sequence alignment was first performed, as seen in figure 3.1. Sequence alignment The TkA clustal Omega alignment method was used with L-ASNase and other enzymes in the l-asparaginase enzyme family. Other selected enzymes from the asparaginase family were *Erwinia chrysanthemi* (1HG1), *Escherichia coli* (3ECA), *Pyrococcus horikoshii* (1WNF) and *Helicobacter pylori* (2WT4).

3.2. Hotspot Wizard

Hotspot Wizard creates smart libraries for automatic identification of hotspots in proteins and their stability, catalytic activity, substrate specificity.

- It works by using advanced techniques such as homology modelling and quality assessment of models, and by performing comprehensive sequence analysis and alignment.
- Hotspot Wizard focus to predict how a protein looks and checks the model's quality. The tools compare protein sequences to design mutations.

For this reason, the hotspot wizard was used to determine the hotspot in the protein and to predict the regions where mutations could be created.

3.3. In Silico Mutagenesis

After collecting Hotspot Wizard and literature information, selected mutations were created with Pymol to examine Tka protein behavior.

Tka (5ot0) enzyme was obtained from RCSB as a “.pdb” file. After deleting HETATM and water molecules in Chimera, a relaxation structure was obtained using energy minimization and the pyrosetta web tool. Once the appropriate conditions were met, the startup file was loaded into Pymol. The protein window was selected by selecting the "Mutagenesis" option from the "Wizard" panel. The residue to be mutated and its location were selected from the window. With 'Apply' and 'Done' the process is completed and the mutant file is saved. The processes can be seen in figure 3.2.

Additionally, the new type of the mutated amino acid was confirmed by typing 'get_resi' on the command line.

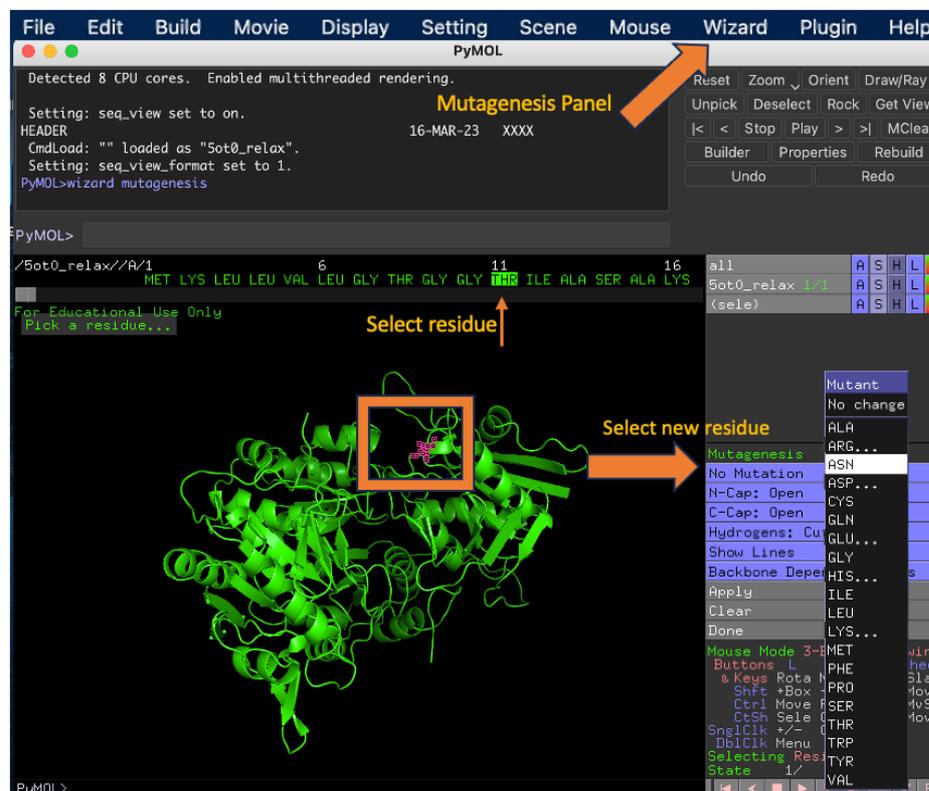


Figure 0.2. Illustrated demonstration of the Pymol interface and some instructions. The arrows indicate the parts that should be used for mutagenesis protocols.

3.4. Autodock Tools and Autodock Vina for Molecular Docking Studies

In order to statically model and energetically optimize the binding events between the TkA enzyme and its ligand asparagine, molecular docking studies were carried out in anhydrous environment. The studies were abandoned in Autodock Tools, Autodock Vina and Chimera to test the accuracy of the studies. Docking studies were carried out in two separate groups: rigid and flexible docking studies. Flexible docking was aimed to obtain more realistic results by considering the flexibility of certain amino acids in the active site of the protein.

Autodock Tools is used to find possible positions where the ligand can interact with the protein with the Lamarckian Genetic Algorithm, and based on scoring functions, Autodock Tools evaluates the interaction energy and stability of the ligand and the TkA enzyme. It shows a lower energy and a stronger binding potential. Autodock Tools was used in rigid docking studies for TkA and asparagine ligand. It enabled the evaluation of binding energies between TkA and asparagine ligand.

Autodock Vina uses a gradient optimization algorithm and finds energetically optimized docking positions. It also has an empirical scoring function. The function is used to predict binding affinity and stability of binding positions. Autodock Vina was mostly used in flex docking works. Docking work was evaluated using the docking scores provided by Autodock Tools.

TkA protein structure was obtained from PDB database (PDB ID: 5OT0). Asparagine as ligand was obtained from PubChem. Protein and ligand were prepared for molecular docking using AutoDock Tools (ADT) and AutoDock Vina. Water molecules, HETATM and water molecules have been cleared. Polar hydrogen atoms were added. Active site amino acids for flex docking were determined. First, Thr11 and Thr85, then all active site amino acids, namely thr11, tyr21, ser54, thr55, thr85, asp86 and lys156, tyr233 and glu275 from the B chain were selected as flexible. Grid box dimensions were determined by considering the coordinates containing the active region as (-29, 25, -26). Box dimensions set to (20, 20, 20). Exhaustiveness was chosen 8. Lamarckian Genetic Algorithm (LGA) was preferred for docking studies. Docking studies of WT TkA and mutant TkA enzymes for their ligand asparagine were performed for 10 runs, 20 runs, 50

runs and 100 runs. The results obtained were analyzed based on RMSD (Root Mean Square Deviation) values and energy scores.

Binding positions and energetic properties were compared for each working set for wild type and mutants. Based on the docking results obtained, the interaction mechanisms of WT TkA and mutants with asparagine were compared. The effects of flexibility and modifications of active site amino acids on ligand binding were evaluated. Relatively higher docking scores were taken into consideration to identify new mutants.

Additionally, TkA and ligand asparagine complex were compared with the results of *E. coli* l-asparaginase and ligand asparagine docking.

3.5. Molecular Dynamics Simulation via GROMACS

Molecular dynamics simulations were performed following two different GROMACS protocols. Among these, the "Protein in Water" protocol was used to examine the behavior of the WT TkA enzyme in aqueous media without ligands in Appendix B. The second one was the "Protein-Ligand Complex" protocol to understand dynamics of TkA enzymes with ligands in Appendix C. Also, the protocols are referred to as GROMACS. It can be accessed on the Tutorials page (Lemkul 2019). MD simulations were performed with complexes of mutant and WT TkA and asparagine ligands obtained from molecular docking studies. Since the apo state TkA enzyme does not contain a ligand, the PDB file prepared for docking studies was used. This study was conducted in two separate repetitions. While mutant experiments were examined in two separate repetitions, MD studies on wild type TkA and asparagine ligand were performed in three repetitions.

GROMACS uses various parameter files for the analysis of biological systems close to reality. GROMACS performs simulations using files in various formats. Definitions of these formats are given in Table 3.1.

During the simulation, the choice of force field and solvent used is important because the entire simulation takes place within certain limits of exposure to this force field and solvent. Specifically for intrinsically disordered peptides and proteins, the CHARMM36 force field was chosen because it is more successful in generating

polypeptide backbone conformations, while the TIP3P solvent model was chosen for its computational efficiency. The box edge in which the simulation will take place was studied as cubic and the size of the box edge was studied as 2.0 nm.

GROMACS, MD simulations basically consist of an energy minimization and two equilibration phases. Before the start of the simulation, in the first step, energy minimization, Particle Mesh Ewald (PME), which is used for long-range electrostatic interactions, was selected to minimize the energy to the lowest possible level. The parameter file is labelled "em.mdp", with other parameter files in Appendix A (figure A.1). Energy minimization phases were set to 100 ps.

Table 3.1. GROMACS file formats (extensions) and their tasks.

File Format Name	Descriptions
Coordinate file (.gro)	It contains the coordinates of the atoms. Initial configuration file or output file
Energy file (.edr)	Thermodynamic term contents such as energy, temperature, pressure
Trajectory file (.xtc)	Stores the positions of the atoms over a period
Checkpoint file (.cpt)	Recording simulation progress. Allows simulation to resume where it left off if failed.
Log file (.log)	Information and procedure for simulation error messages
Molecular dynamic parameter file (.mdp)	Include parameters
Topology file (.top)	It contains structural information of the molecular system, such as atoms, molecules, bonds, angles, dihedrals and nonbonded interaction parameters.
Trajectory parameter file(.tpr)	Includes details of the topology of the molecular system, the force field parameters, the simulation settings and the initial coordinates.

In the first equilibration phase, NVT ensemble (constant number of particles, volume, and temperature), the temperature of the system should be reaching the expected plateau level. For this purpose, the V-rescale thermostat was used as the temperature control method parameter. In the parameter file called "nvt.mdp" in Appendix A (figure

A.2), the temperature thermostat parameter was assigned with the expression "tcoupl = V-rescale". First equilibration step was set to 100 ps.

In the second equilibration phase, NPT ensemble (Number of particles, Pressure, and Temperature), Parrinello-Rahman pressure matching was used for pressure equilibration. In the file "npt.mdp" as Appendix A (figure A.3), this parameter appears as "pcoupl = Parrinello-Rahman". Second equilibration step was set to 100 ps.

The last step is MD run. In this step, the system dynamics were analyzed with the conditions specified in the "md.mdp" parameter file and specified up to this step. Each simulation was performed with 25,000,000 steps and 0.002 femtoseconds. It was set to record 5000 frames at each step. Therefore, a total of 500 frames and 50 nanoseconds were obtained.

3.6. MD Simulation Analysis and Evaluation

3.6.1. Cluster Analysis

Cluster analysis was performed for the apo state TKA enzyme after MD simulation. This analysis indicates the conformation in which atoms or proteins exhibit similar behavior when examined and shows the conformation in which they are most frequently found throughout the simulation and indicates how many frames or nanoseconds these conformations belong to in the aggregated state. Within the scope of this thesis, cluster analysis for apo state TKA was performed according to the cutoff value of 0.1. The cutoff value was determined according to the average RMSD value.

3.6.2. RMSD (Root Mean Square Deviation)

RMSD analyzes were obtained by using the "gmx rms" command by taking the protein as reference to the backbone and fitting it against the backbone. The equation for

RMSD calculations is defined as two structures corresponding to each N particle, with particle coordinates $\{r1^{\vec{}}$ and $\{r0^{\vec{}}$, denoted as 1 and 2, and the equation is like equation1 (Pitera 2014).

$$RMSD = \sqrt{\frac{1}{N} \sum_{i=1}^N (\vec{r}_{1,i} - \vec{r}_{0,i})^2}$$

3.6.3. RMSF (Root Mean Square Fluctuations)

The RMSF is a property of structures or scattering ensembles. It indicates the fluctuation of an atomic position around its mean value. Calculation is provided by with the programmed “gmx rmsf” in GROMACS. The RMSF calculated from an ensemble of structures T for a single site or atom i is:

$$RMSF = \sqrt{\frac{1}{T} \sum_{j=1}^T \left[\vec{r}_{j,i} - \frac{1}{T} \sum_{k=1}^T \vec{r}_{k,i} \right]^2}$$

The assumption is that the coordinates have been rotated and translated in order to minimize the RMSD (Pitera 2014).

3.6.4. Radius of Gyration

In order to have a rough measure of the compactness of a structure, it is possible to calculate the radius of gyration with the programmed “gmx gyrate” as follows:

$$R_g = \left(\frac{\sum_i \| r_i \|^2 m_i}{\sum_i m_i} \right)^{1/2}$$

In this equation, m_i is the mass of atom as well as the position of atom with respect to the center of mass of the molecule (Lemkul 2019).

3.6.5. Number of Hydrogen Bond Analysis

The number of hydrogen bonds was calculated between atoms at 0.35 nm in GROMACS with the "gmx hbond" command. Analyzes hydrogen bonds (H bonds) between all possible donors D and acceptors A in figure 3.3. A geometric criterion is used to determine whether an H bond is present (Lemkul 2019).

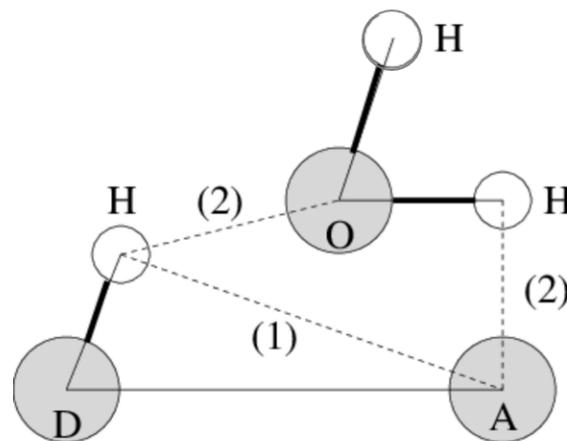


Figure 0.3. Hydrogen bond geometry in water. (1) shows H-bond between two residues. (2) indicates H-bonding bridge via water molecules.

3.6.6. SASA (Solvent Accessible Surface Area)

The solvent accessible area (ASA) defines the area where contact between the protein and the solvent can occur. In the SASA calculation, a probe sphere representing the solvent molecule is rolled over the van der Waals surface of the protein. This movement allows calculation of the surface area accessible by the solvent. The accessible surface is where the center of the probe is called the locus in figure 3.4.

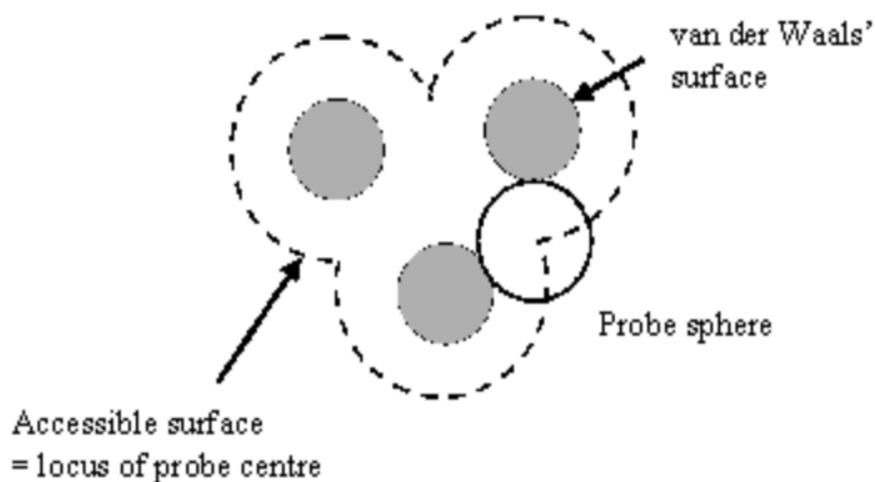


Figure 0.4. Accessible surface of a molecule with a probe sphere.

3.6.7. Protein Ligand Interaction Energy

Lennard-Jones interaction energy, which describes protein and ligand interaction energy, and Coulomb interaction energy, which defines electrostatic energy, were examined. Total energy was also obtained from this study for WT TkA and mutant TkA. Total energy was obtained by summing the Lennard-Jones and Coulomb interaction energy values. These total energy values were then compared for WT TkA and mutant TkA. TkA and the ligand complex with the highest energy value were evaluated as having the strongest interaction (Lemkul 2019).

3.6.8. RDF (Radial Distribution Functions)

RDF was calculated using the "gmx rdf" command. With the calculation, the probability of finding the ligand in the protein active site and the selected residues Thr11, Thr55, Thr85, Asp86 and Glu294 in the active site over time was examined.

The equation is:

$$g_{AB}(r) = \frac{\langle \rho_B(r) \rangle}{\langle \rho_B \rangle_{local}} = \frac{1}{\langle \rho_B \rangle_{local}} \frac{1}{N} \sum_{i \in A} \sum_{j \in B} \frac{\delta(r_{ij} - r)}{4\pi r^2}$$

$\langle \rho_B(r) \rangle$ represent the particle density of *B* at a distance *r* around particles *A*.

$\langle \rho_B \rangle_{local}$ represent the particle density of *B* average of all the spheres around the particles *A* with radius (Lemkul 2019).

3.6.9. MSD and Diffusion Coefficient (Mean Square Displacement)

MSD and D_A for the ligand was calculated with the "gmx msd" command. The data obtained from the MSD curve was used in Einstein equations and the diffusion coefficient was calculated. The results were compared for mutants and WT TkA. MSD equation is seen as:

$$\lim_{t \rightarrow \infty} \langle \| r_i(t) - r_i(0) \|^2 \rangle_{i \in A} = 6D_A t$$

$$D = 1/2d \log_{t \rightarrow \infty} \frac{d}{dt} \langle [\overrightarrow{r}(t) - \overrightarrow{r}(0)]^2 \rangle$$

Typically, an index file is used that contains the atom numbers and the MSD is then averaged over these atoms. r_i , are the positions of the molecular centers of mass (Lemkul 2019).

In overall, WT TkA and mutant TkA Md studies were evaluated comparatively together with the results of the analyses.

CHAPTER 4

RESULTS AND DISCUSSION

In the study, the behaviors, and conformational changes between the Apo state of the wild type (WT) TkA enzyme and the enzyme with its ligand asparagine were examined. Different mutations were modeled in the TkA active site residues. TkA and L-asparagine ligand were affected by mutations were examined while examining conformational changes and behavior relative to each other. The study was conducted with *in silico* analysis and methods. Various mutation experiments were carried out in the TkA enzyme, whose active site residues were examined. Conformational changes were observed through docking studies, protein ligand docking in the active site, and molecular dynamic simulation studies.

4.1. Molecular Docking Studies and Mutations Selection

A multiple sequence alignment was performed for TkA and L-ASNase from other different organisms in figure 4.1.

Table 4.1. L-asparaginase active site residues from different strains. The apostrophe at Tyr233 and Glu275 indicates active site residue from the B chain.

Name of Strain of L-asparaginase	Active Site Residues
Thermococcus kodakarensis L-Asparaginase (5OT0)	T11, Y21, S54, T55, T85, D86, K156 Y233', G275'
Erwinia chrysanthemi L-Asparaginase (1HG1)	T15, S62, Q63, T95, D96, A120
Escherichia coli L-Asparaginase (3ECA)	T12, Y25, S58, T89, D90, K162 N248', E283'

(cont. on next page)

The active site for the TkA as reported by Guo et al. These are quite similar to the L-ASNases obtained from other organisms. As shown in Table 4.1, the sequence alignment results for *Erwinia chrysanthemi* (1HG1), *Escherichia coli* (3ECA), *Pyrococcus horikoshii* (1WNF) and *Helicobacter pylori* (2WT4) strains, supported by the literature search, show that the active site residues in different organisms are similar (Dhavala and Papageorgiou 2009; Yao et al. 2005; Swain et al. 1993; Aghaiypour, Wlodawer, and Lubkowski 2001; Guo et al. 2017).

After sequence alignment and literature review, it was understood that active site residues for TkA were conserved as Thr11, Ser54, Thr85, Asp86, Lys156 from chain A Tyr233 and Glu275 from the neighboring chain.

4.2. Molecular Docking Studies with WT TkA and L-asparagine Ligand

Following sequence alignment and literature research, molecular docking studies were performed. Docking work was done as rigid, flex and blind docking. Rigid docking and blind docking work was done with Autodock Tools and flex docking work was done with Autodock Vina.

The wild-type TkA is known to be a thermostable enzyme that exhibits activity in a dimeric structure with six chains. According to RCSB data, TkA is provided together with ligands from PO_4^{3-} , PG_4 and EDO molecules. In the article by Guo et al., experiments with the PO_4^{3-} molecule indicated the potential active site region residue as Thr11, Ser54, Thr55, Thr85, Asp86, and Lys156, along with Tyr233 and Glu275 coming from the neighboring subunit. During molecular docking studies, these residues were considered while creating the grid box (Guo et al. 2017).

The suitable grid box size for flexible and rigid docking studies was determined as 20x20x20, and the grid box coordinates were set at (-29, 26, - 25). A grid box large enough to encompass the entire dimer protein was defined to observe how the ligand would move in an unrestricted manner, known as blind docking. However, blind docking studies showed that a box of a specific size and coordinates containing the active site

residues should be used, with a relatively low docking score. These data were not used in the continuation of the study.

Table 4.2. Best docking scores for rigid docking studies on the TkA. Studies were performed as 10 runs, 20 runs, 50 runs and 100 runs.

	10 run	20 run	50 run	100 run
Docking Scores (kcal/mol)	-4.93	-5.5	-5.30	-5.38

In the first set of flex docking studies, Thr11 and Thr 85 were set as flexible residues. In the second set of flex docking studies, all active site residues were set as flexible. Ultimately, however, no significant change in binding score was observed between the two studies. Rigid docking studies were performed using Autodock Tools and Autodock Vina and were performed in 10, 20, 50 and 100 runs as shown in Table 4.2, while flexible docking was performed in 10 cycles in Table 4.3. Flexible docking was performed using Autodock Vina. All rigid docking experiments were repeated using Chimera.

As a result, for the WT dimer TkA, with a rmsd value of zero, the best binding score for flexible docking was -5.87 kcal/mol, while the best binding score for rigid docking was -5.5 kcal, with a RMSD value of 0. After 3 repetitions of each run, the average value of the best binding score for wild-type TkA was determined to be -5.5 kcal/mol.

Considering these results obtained from docking studies, there may be several reasons why higher results are obtained from 50 run simulation than from 100 run simulation. First, the algorithms used during docking studies are sensitive and random. It contains random components such as the genetic algorithms used in these studies. This means that even if the same parameter standards are used, different results can be obtained regardless of the number of runs. Another reason is that in docking studies, the initial coordinates and orientations of the receptor and ligand are different, which causes randomness. The visual of TkA and ligand asparagine along with the active site residues

of the -5.5 kcal/mol docking result obtained from Ligplot+ and Chimera is shown in figure 4.3.

Table 4.3. Flex docking results according to the selection of active site residues in different combinations. -5.87 kcal/mol is chosen as best score.

Flex Residues	Dock scores (kcal/mol)
Thr11, Thr85	-5.56
T11, Y21, S54, T55, T85, D86, K156, Y233', G275'	-5.87

The results obtained were compared with literature data. The figure 4.2 shows the results of the comprehensive docking studies carried out by Baral et al. shown for the L-asparaginase family. The results obtained for TkA appear to be compatible with these studies (Baral et al., 2021).

Organism	Michaelis constant value from literature (mM)	Measured k_{cat} values from literature (s^{-1}) ^a	Binding affinity calculated from docking (kcal/mol)
<i>Bacillus licheniformis</i> 1	0.014	2.68×10^3	-4.8
<i>Escherichia coli</i> ^b	0.015	2.4×10^1	-5.1
<i>Deftia acidovorans</i> ^b	0.015	— ^c	-5.1
<i>Dickeya chrysanthami</i> 2 ^b	0.058	23.8×10^3	-5.0
<i>Azobacter vinelandi</i> ^b	0.11	—	-4.9
<i>Pseudomonas stutzeri</i> 2	0.14	—	-4.9
<i>Bacillus aryabhatai</i> ^b	0.257	—	-4.8
<i>Helicobacter pylori</i> 1 ^b	0.29	19.26 +/- 0.56	-4.8
<i>Bacillus subtilis</i> 1 ^b	0.43	—	-4.5
<i>Pectobacterium carotovorum</i> 1	0.657	2.751×10^3	-4.4

Figure 0.2. A comprehensive docking study and docking scores by Baral et al. for the l-asparaginase family that categorized as gene and type of l-asparaginase. ^B indicate ansb gene and 1 or 2 indicates type of l-asparaginase.

As seen in figure 4.3, the Ligplot+ image obtained from TKA analysis shows that the hydrogen bonds, and the ligand is observed far from the residues that cause nucleophilic attack in the active site (Thr11, Thr85). For *E. coli* l-asparaginase, which has more research in the literature, the events occurring in the active site were described by Anishkin et al. They proposed the following for the EcA active site: T12, S58 and T89 “clamp” and stabilize the ligand. Y25 loses contact with T12 and E283 over time. In fact, this indicates that Y25 is now away from the ligand. Figure 4.4 shows the image of Anishkin et al. Although K162 and N248 do not directly contact the ligand, these residues may contribute to the hydrogen bond network.

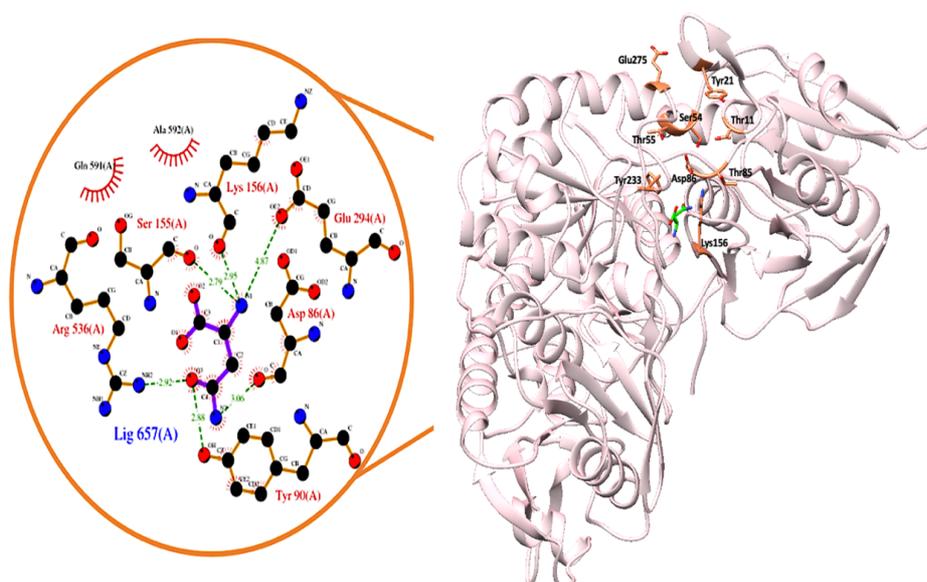


Figure 0.3. Three-dimensional structure of TkA and its ligand asparagine after docking. The green color indicates the L-asparagine ligand, and the orange parts represent active site residues. The image on the left was created with Ligplot+ and shows hydrogen bond formation.

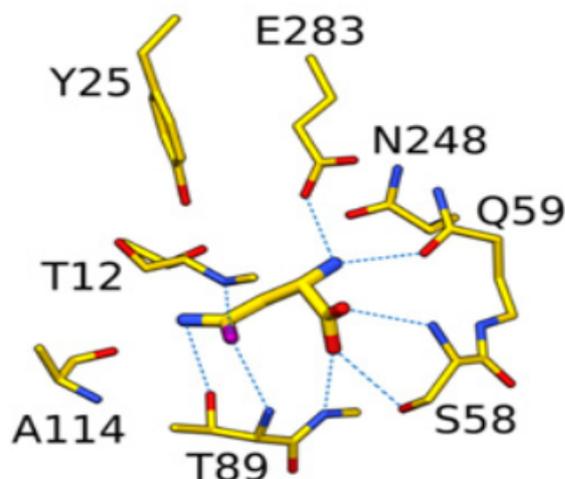


Figure 0.4. Active site residues and asparagine ligand of *E. coli* L-asparaginase. Showing Thr11 and Thr85 residues that are involved in the nucleophilic attack on the substrate.

4.3. Molecular Docking Studies with Mutated TkAs and L-asparagine Ligand and Mutation Selection

The information obtained about the active site amino acids and which regions are more conserved have shown that the Thr55 is not conserved in other members of the asparaginase family. Therefore, Thr55 was selected as a potential mutation candidate. Thr11 was also selected for mutation as it contributes to the nucleophilic attack. This mutation provided insight into behavior of the asparagine ligand in the active site. Docking studies have shown that Asp86 contributed to hydrogen bond formations. So, it was chosen as the third mutant for TkA enzyme.

New amino acids determined for the mutation sites were made using Hotspot Wizard and molecular docking studies. Alternative amino acid candidates for the Thr55 mutation point were also obtained with the HotSpot wizard in figure 4.5.

Table 4.4. Docking scores of WT TkA, WT EcA and mutant TkAs. T11E, T55E and D86S are used for MD simulation candidates.

Mutation Name	Binding Score (kcal/mol)
WT EcA (3ECA)	-5.1
WT TkA (5OT0)	-5.5
T11E	-5.2
T11Q	-5.1
T11A	-5.1
T11Y	-5.2
T11W	-5.1
T55A	-4.3
T55E	-5.5
T55E	-5.1
T55K	-4.8
T55D	-5.3
D86Q	-5.2
D86S	-5.3
D86H	-5.2
D86E	-5.0
D86F	-4.8

Mutants were also identified by comparing mutant TkA data obtained from molecular docking studies with WT TkA. Molecular docking studies performed by rigid

docking for mutant TkAs and asparagine ligand gave results close to docking studies with WT TkA and asparagine ligand. Table 4.4 shows some binding energy scores. Residues giving relatively higher binding energy scores were selected as mutant TkA in figure 4.6. These mutants were designated T11E, T55E, and D86S TkA. The binding energy scores of the selected mutant structures are -5.2, -5.5, -5.3 kcal/mol, respectively. These three selected mutants were then examined as established complexes in MD simulations.

Position	Amino Acids	Ala	Cys	Asp	Glu	Phe	Gly	His	Ile	Lys	Leu	Met	Asn	Pro	Gln	Arg	Ser	Thr	Val	Trp	Tyr
12	Ile	22,7	30,8	18,0	18,0	56,3	10,0	26,5	X	12,8	47,4	45,0	25,1	2,8	18,0	12,3	39,3	25,1	60,1	9,5	12,3
14	Ser	78,8	87,0	85,1	73,1	78,5	73,1	87,7	56,3	84,5	78,8	87,7	75,9	53,1	100,0	73,1	X	85,1	84,5	65,2	78,5
21	Tyr	87,7	85,1	87,0	84,5	87,7	73,1	85,8	72,5	78,8	87,7	87,0	81,0	50,2	87,7	84,5	87,7	84,5	87,7	60,1	X
22	Lys	100,0	87,0	81,0	78,5	87,7	73,1	84,5	73,1	X	87,7	87,7	78,5	47,4	85,1	75,9	100,0	81,0	87,7	65,2	75,9
23	Ala	X	100,0	87,0	78,5	85,1	87,7	85,1	60,1	73,1	81,0	81,0	84,5	85,1	81,0	73,1	87,0	81,0	81,0	39,3	60,1
55	Thr	85,1	87,0	87,0	76,9	87,7	75,9	87,7	50,2	84,5	87,0	87,7	75,9	65,2	87,7	78,8	87,7	X	84,5	60,1	78,5
83	His	10,0	30,8	38,9	10,0	18,0	10,0	X	9,5	22,7	12,8	18,0	18,0	2,8	18,1	18,0	71,2	45,0	22,7	5,2	42,7
112	Met	53,1	18,0	42,7	56,3	39,3	9,5	25,1	38,9	72,5	78,8	X	18,0	2,8	87,7	78,5	38,9	18,1	39,3	30,8	30,8
159	Ser	84,5	84,5	81,0	71,2	84,5	73,1	84,5	60,1	53,1	78,8	81,0	76,9	53,1	78,5	65,2	X	87,0	81,0	18,0	38,9
233	Tyr	47,4	78,5	73,1	47,4	87,7	39,3	84,5	65,2	65,2	72,5	71,2	73,1	12,8	65,2	50,2	100,0	85,8	84,5	73,1	X
235	Ala	X	100,0	84,5	71,2	87,7	73,1	78,5	65,2	47,4	87,7	87,7	78,5	47,4	78,5	60,1	87,7	84,5	87,7	12,8	26,5
237	Gly	53,1	76,9	81,0	53,1	60,1	X	85,1	65,2	50,2	50,2	60,1	87,7	26,5	60,1	50,2	85,8	76,9	73,1	26,5	47,4

Figure 0.5. Amino acid substitution matrix of the protein. The table shows the possible percentages of amino acid substitutions that can occur at certain positions in the protein. Each row represents a different amino acid combination (shown left), and each column represents a possible substituting amino acid (shown above). The color coding visualizes the level of the change score: Green indicates a higher match; red indicates a lower match. The 'X' sign is an indication that a change at that position is unlikely.

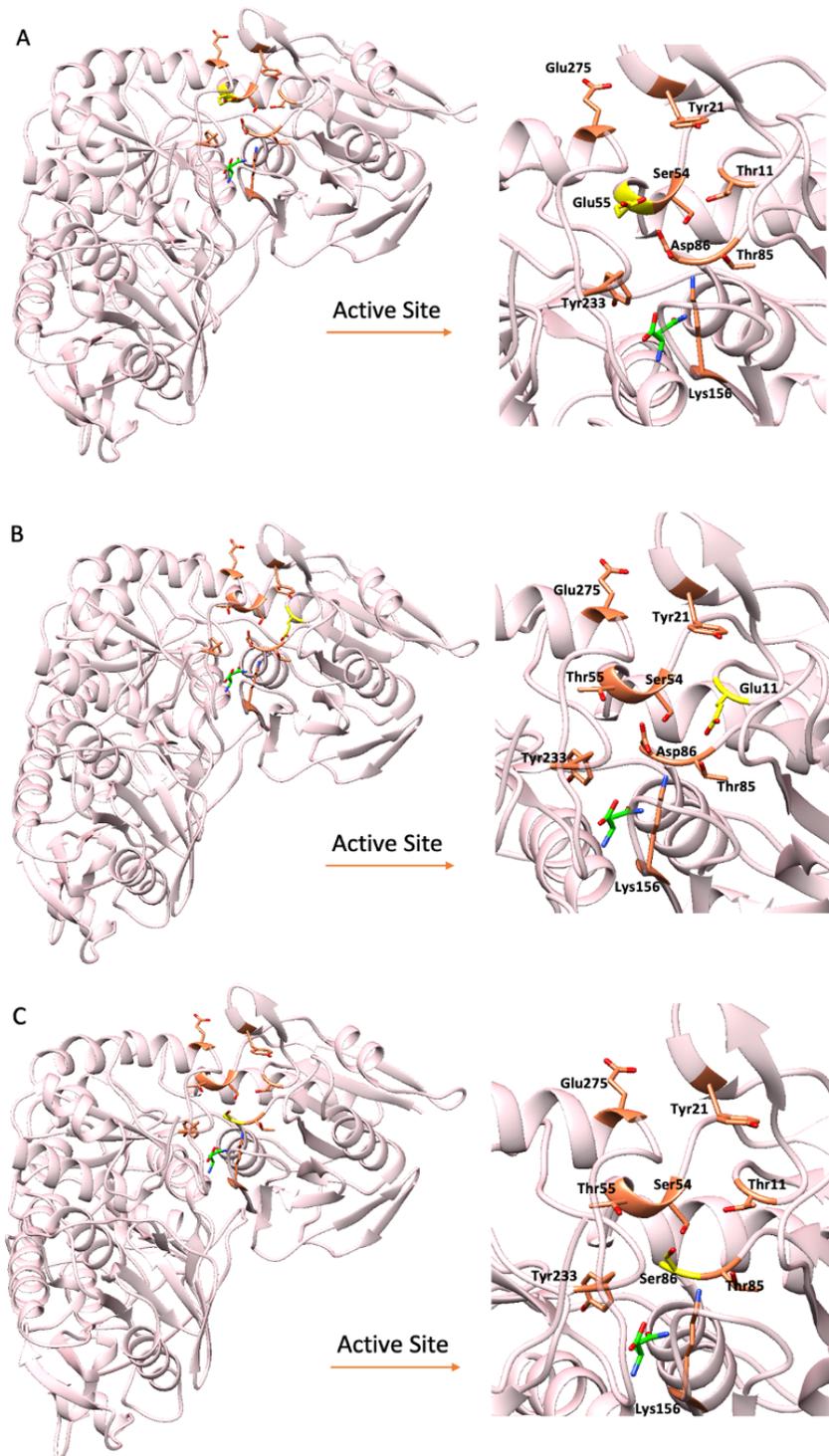


Figure 0.6. A) Structure of T55E TkA highlighted in orange indicate active site residues and green indicates the amino acid ligand. B) 3D TkA structure of the Asp86Ser mutation, structures highlighted in orange indicate active site residues and green indicates the amino acid ligand. C) 3D TkA structure of Thr11Glu mutation, structures highlighted in orange indicate active site residues and green indicates the amino acid ligand. Additionally, regions with mutations are shown in yellow.

4.4. Molecular Dynamics Simulations

Molecular docking studies were used to understand the binding between TkA L-asparaginase and its ligand asparagine and how the conformation of the ligand changes relative to the protein. They are typically useful for understanding potential areas of interaction between the ligand and the protein. However, molecular docking does not show the dynamics of binding or the interactions of atoms with each other over time. Molecular Dynamics (MD) simulations model the movement of the protein-ligand or just the protein in water molecules over time, providing an environment much closer to biological systems and more accurate results than docking studies.

For all MD simulations, a common set of parameters were chosen for all MD simulations at a temperature of 300 K, using the charmm36 force field, the Tip3p solvent and a cubic box environment. Many different analyses were carried out using the cubic box as a boundary of 1.0 nm. Analyses carried out after the simulations showed that periodic boundary condition errors were easier to deal with in the 2.0nm studies. 100 ps NVT was performed to determine the thermostat and balance the system, and 100 ps NPT was performed to select the barostat, control the density, and prepare the system for the production step. While 2 repetitions of 50 ns were performed for the MD simulation of the apo state TkA enzyme in aqueous environment, 3 repetitions of 50 ns were performed for the WT TkA enzyme and ligand complex obtained from docking models. Additionally, T55E, D86S and T11E TkA enzymes were studied with 2 repetitions MD simulations of 50 ns in the presence of asparagine ligand.

Initial parameters for each MD simulation were evaluated with potential energy, temperature, pressure, and density. Additionally, root mean square deviation (RMSD), root mean square fluctuation (RMSF), radius of gyration, hydrogen bonding (H-bond), accessible surface area (SASA), protein-ligand interaction energy, radial diffusion function (RDF), and diffusion coefficient analyzes were performed.

Equilibrium parameter graphs of all MD simulations such as potential energy, temperature, pressure, and density values are given in Appendix B. A comparison with the literature was also made for the accuracy of the initial conditions. Results for protein and protein-ligand complexes showed that, over 50 ns at a temperature of 300 K, systems

equilibrated between pressures of -200 to +200 bar and densities of 1020 kg/m³ to 1025 kg/m³. Additionally, the outcomes were in line with previous research (Lemkul 2019).

4.4.1. Apo State WT TkA in Water 50 ns Simulations

A 50 ns MD simulation was performed for apo TkA in aqueous media. Throughout this simulation, the TkA crystal structure was used according to the GROMACS tutorial. Backbone-to-backbone RMSD graphs for Apo TkA can be seen in figure 4.7. The protein backbone represents the skeleton of amino acids consisting of the C-alpha carboxyl group, ammonium group and hydrogen atoms. This structure plays a role in determining the 3D structure of the protein (Alberts, A, and Lewis J, 2002).

The two separate replication studies are compatible with each other. The average RMSD value can also be seen in figure 4.7.B.

The graph showed a slight increase from 0.1nm to 0.15nm in the first 1000 frames and then reached equilibrium. The small increases that occur initially represent conformational changes that the protein undergoes during the equilibrium phase. As a result, a balanced RMSD was obtained over 5000 frames. This showed that most of the structural changes in the protein were completed and the protein's equilibrium state was reached (Erva et al. 2016).

According to the literature, Guo et al. colleagues mentioned the similarity of the subunits of the TkA enzyme, which are positioned asymmetrically relative to each other in their dimer structure. This similarity argued that the subunits contained in each dimer structure were similar within the RMSD range of 0.08 to 0.15 Å. TkA was compared with E. coli l-asn I and E. coli l-asn II. The similarity value for E. coli l-asn I was found to be 1.11 Å and the similarity value for E. coli l-asn II was found to be 1.67 Å RMSD. Additionally, the apo EcA RMSD value is given in the literature as an RMSD value between approximately 0.2 and 0.4 nm. When this value is compared with apo TkA, the apo EcA RMSD value is higher than the apo TkA RMSD value (Guo et al., 2017; Erva et al., 2016).

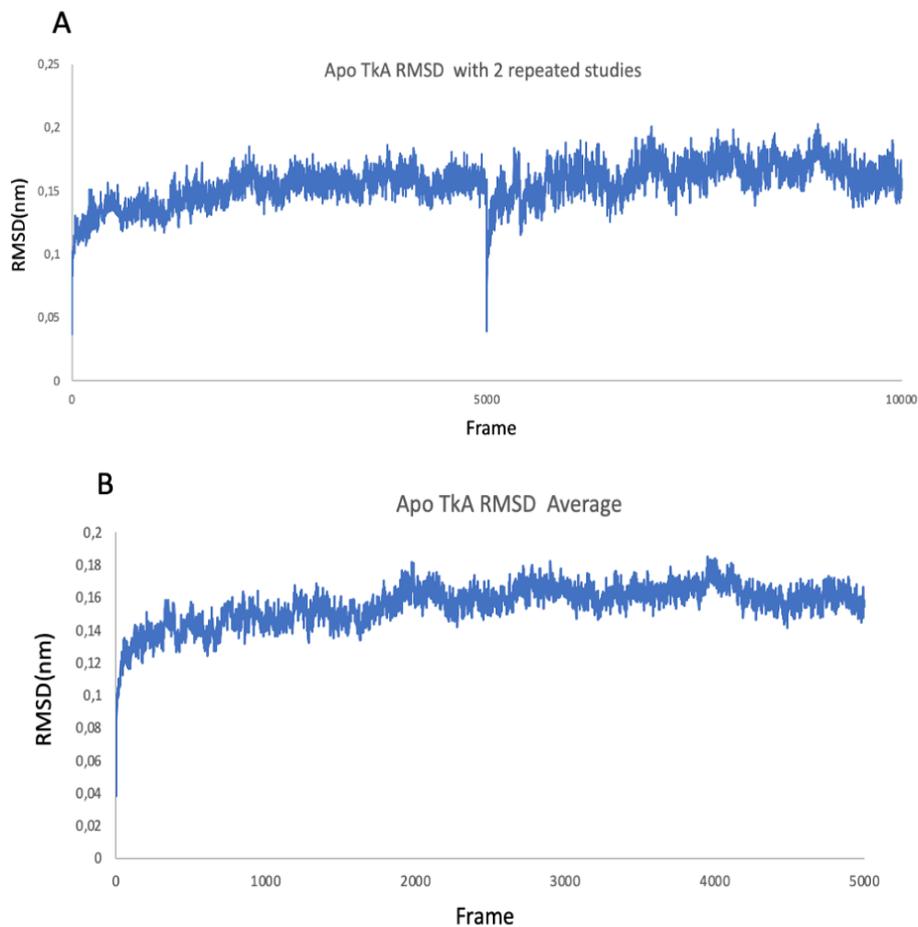


Figure 0.7. Shows RMSD graphs of WT Apo TkA. A) Shows the RMSD graphs of two separate repetitions, while B) shows to the average RMSD graph of two repetitions.

The RMSF graph of WT apo TkA is shown in figure 4.8.A and figure 4.8.B. To investigate the notable peaks in the RMSF graph of WT TkA, the protein was visualized as most frequent encountered and initial conformation in the chimera environment in figure 4.8.C. It was observed that the peaks occurring at certain residues between 37-40 and 189-192 in the protein sequence belonged to regions representing more flexible or exposed regions of the protein. These residues show that it is relatively more flexible and dynamic than the others (Chi et al. 2022; Zhang et al. 2023).

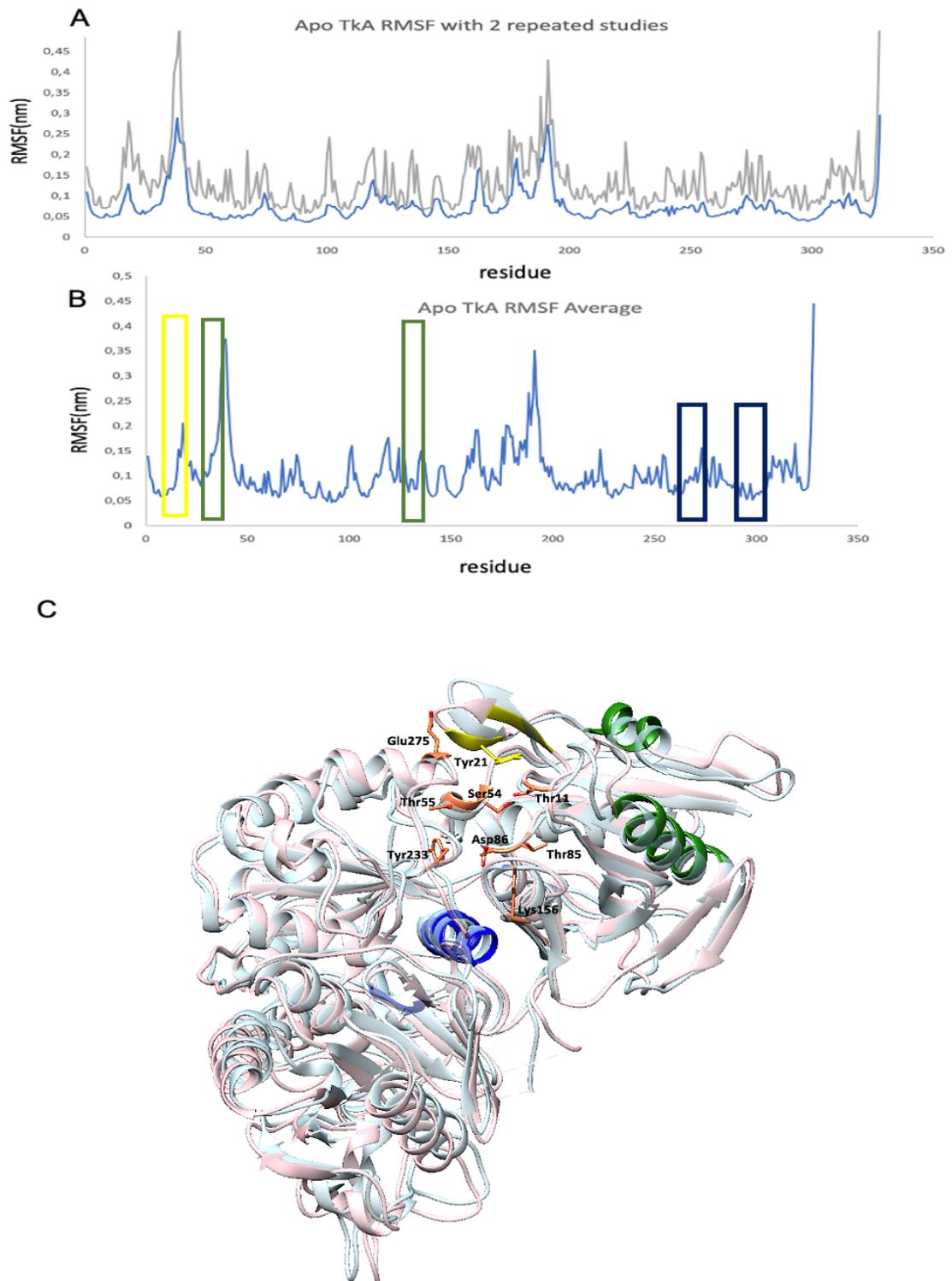


Figure 0.8. RMSF plot and Chimera images of Apo TkA. A) RMSF graphs taken from two replicate studies of Apo TkA, B) shows the average RMSF graph obtained from RMSF graphs. C) As a result of cluster analysis, comparison of the most frequent encountered conformation (pink) during the simulation and the apo TkA conformations at the starting point (blue).

Figure 4.8.C also shows the part shown in yellow here represents the hairpin: Residues: $\beta 2(14-16)$ - $\beta 3(20-22)$. Blue regions indicate allosteric site: Residues: $\alpha 8(293-306)$ - $\beta 15(258-262)$. The green regions $\alpha 1(27-33)$ – $\alpha 4(122-133)$ are the regions that tend to move towards the active site during the ligand and enzyme interaction. In previous studies by Guo et al. it was also stated that the orange region that is between residues $\alpha 1(27-33)$ – $\alpha 4(122-133)$ was moving towards the active region during the enzyme-ligand interaction period (Guo et al. 2017).

The radius of gyration provides important information about protein stability and compactness. The radius of gyration graph for WT apo TkA is seen in figure 4.9. In the graph, the average radius of gyration value of the protein is 2.59 nm. The compact structure of the protein was preserved throughout the simulation, with a very slight increase of 0.5 nm compared to the beginning. Minor fluctuations in the graph express the orientation of the protein but only affect the spatial orientation and do not change the compactness (Erva et al., 2016; Tou et al., 2013).

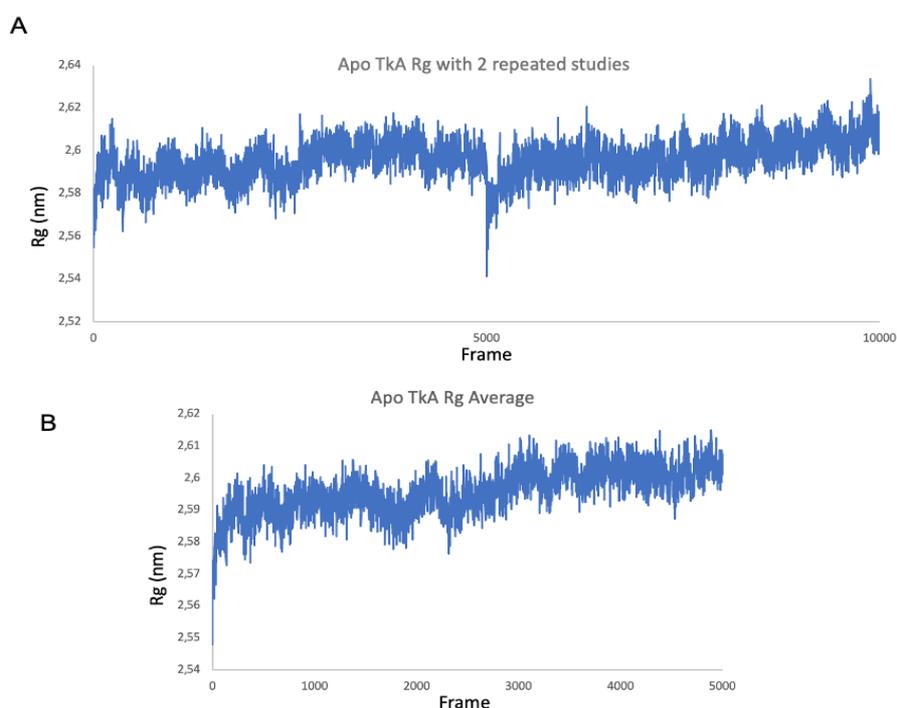


Figure 0.9. Radius of gyration plot to evaluate protein compactness. A) shows two repetitions for apo TkA, B) shows an average plot obtained from two repetitions of Apo TkA plot.

SASA reflects changes in the surface area of a protein interacting with the solvent and provides information regarding conformational contraction or expansion of the protein. It also provides information about its hydrophobicity. According to figure 4.10, the graph of the interaction of WT apo TkA with the solvent reached the value of 260 nm² from 255 nm² at the beginning. It shows that the hydrophilic parts of the protein stabilize at 260.38 nm², preserving its hydrophobicity after interaction with the solvent. This minor change also suggests changes in the structure of WT TkA leading to a slight expansion. (Tou et al. 2013; Erva et al. 2016; Hozoorbakhsh et al. 2023)

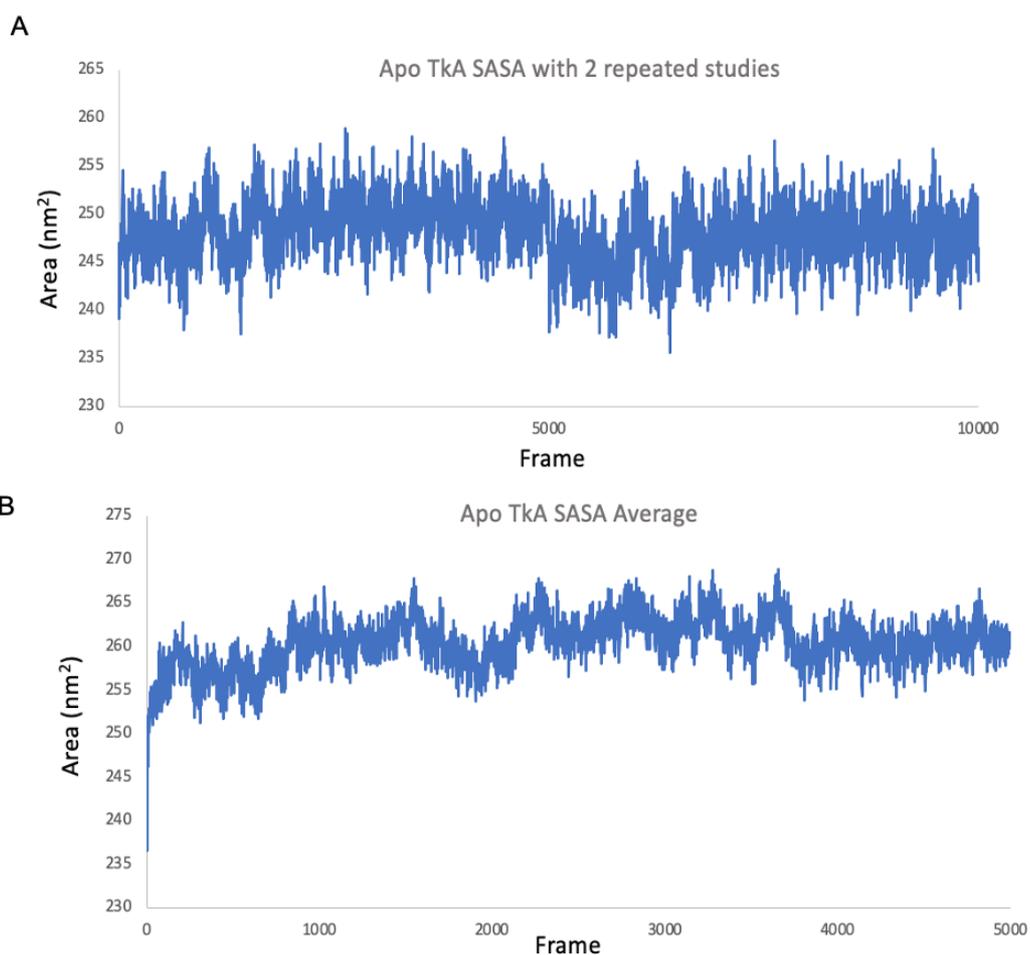


Figure 0.10. SASA analysis plot for Apo TkA. A) shows two repetitions studies for apo TkA and B) shows average SASA plot from two repetitions studies of apo TkA.

In general, while the RMSD graph showed that the WT Apo TkA had undergone minor conformational changes in the aqueous environment, SASA and radius of gyration graphs confirmed that these minor conformational changes resulted in the expansion of the protein. The corroborating data obtained with the RMSF graph and chimera showed that the loops on the protein surface were the parts of the protein that were more flexible than the other parts throughout the general simulation.

Stable graphs for WT apo TkA were compared with literature graphs. In 2017, Maggi et. al. conducted similar studies on EcAII. al found the RMSD value for WT apo EcAII to be 0.19 Å. Ardalan et al., 2021's RMSF analysis for WT EcA stated that residues with high RMSF values are more flexible (Ardalan et al., 2021; Maggi et al., 2017).

4.4.2. Molecular Dynamics Simulations for WT TkA and Asparagine Ligand Complex

MD simulations of WT TkA and its ligand asparagine complex were performed in triplicate for 50 ns.

The results were evaluated for RMSD, RMSF, radius of gyration, SASA, H-bond, RDF, MSD, protein-ligand interactions. In addition, images showing the H-bond interaction and conformational change of the ligand were obtained at time intervals of 0 ns, 15 ns, 30 ns, 45 ns and 50 ns using Chimera and Ligplot+.

Firstly, to understand the dynamic changes and conformational changes of the protein in the complex, the backbone-to-backbone RMSD values of the protein and ligand were examined in three replicates as shown in figure 4.11.A. The results were consistent from the starting point to the equilibrium point.

Backbone-to-backbone RMSD analysis was performed for WT apo TkA and WT TkA with ligand, as shown in figure 4.11.B. Both graphs gave similar results at 0.15 nm and maintained their stability throughout the simulation.

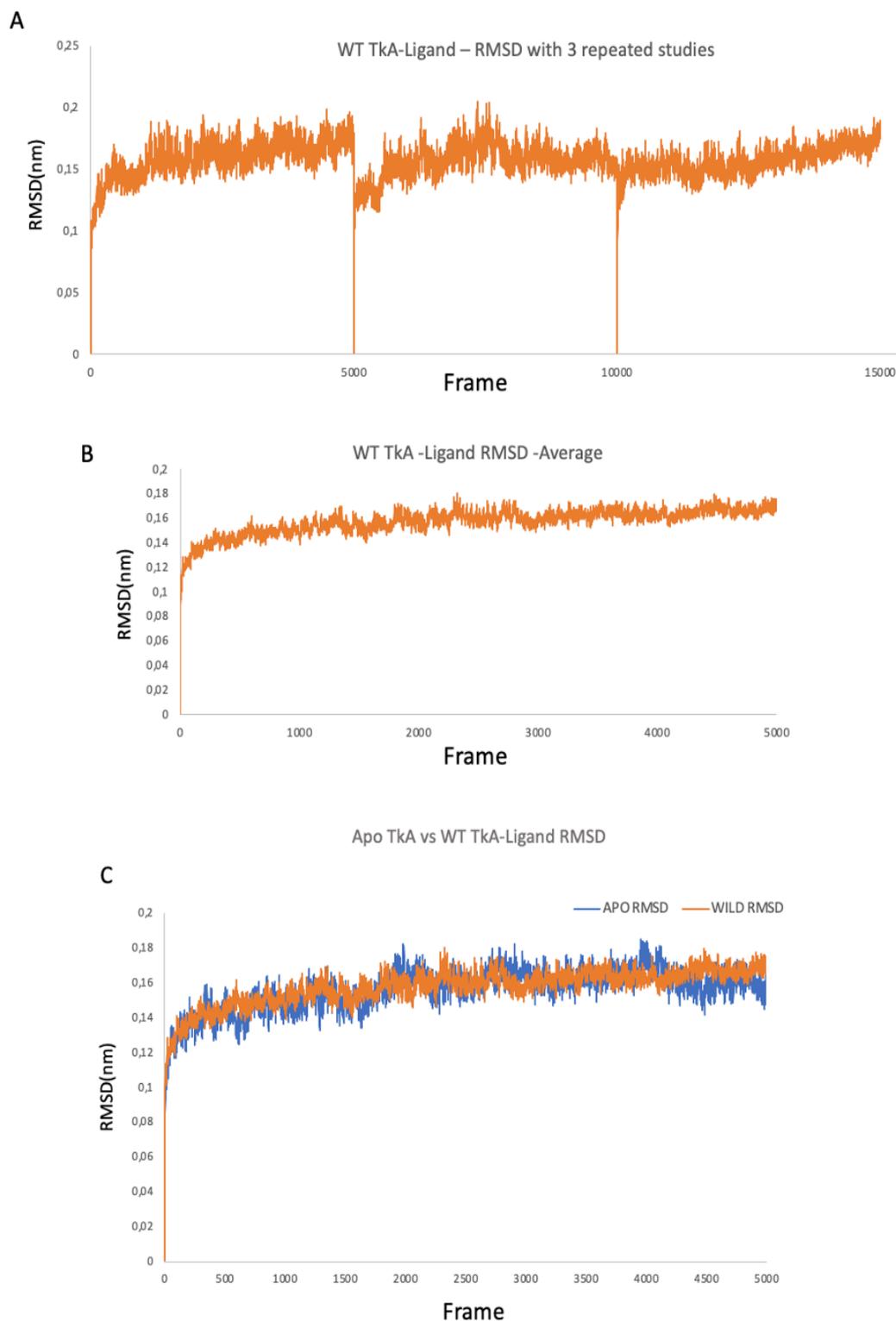


Figure 0.11. A) Backbone-to-backbone RMSD plot for WT TkA and asparagine ligand complex with 3 repetitions. B) Average RMSD plot from 3 repetitions for WT TkA with ligand. C) Comparing RMSD data for WT apo TkA and WT TkA with ligand.

System dynamics change with ligand and protein binding. The protein adapts its conformation to the ligand. The fact that this production state reaches equilibrium indicates that it has a balanced conformational change relative to each other. Erva et al.'s 1NNS pdb-encoded EcA study compared WT EcA in the presence and absence of ligand. RMSD values for apo EcA were found to be higher than those for the ligand and enzyme complex. Hozoorbakhsh et al. performed similar work on thermostable l-asparaginase and compared it to asparaginase with thermal stability in the presence and absence of ligand. This time, asparaginase with thermal stability in the apo state and ligand-bound asparaginase gave RMSD results at close values (Ardalan et al., 2021; Chi et al., 2022; Jiao et al., 2022; Hozoorbakhsh et al., 2023).

Figure 4.12.A is the RMSF graph of three different studies obtained for RMSF WT TkA and its ligand asparagine. The part shown here with the yellow box represents the hairpin: Residues: β 2(14-16) - β 3(20-22). Blue boxes belong to the residue region of the allosteric site: Residues: α 8(293-306) - β 15(258-262). Green regions α 1(27-33)– α 4(122-133) indicate residues from the region known to move towards the active site during ligand and enzyme interaction in figure 4.12.C. RMSF graphs obtained from three different studies gave different peaks at some points, but in general, all three are compatible with each other. The reason for this is explained by Meher et al.: It is expected to be acceptable as long as each simulation does not change significantly, because the initial conformation of the protein and ligand may be different in each simulation or the behavior of a water molecule during the simulation may cause similar conformational changes (Meher and Wang 2015).

Figure 4.12.B is a comparison of the RMSF plots of WT apo TkA and WT TkA with its ligand. Compared to the RMSF in the ligand-free protein MD study, the RMSF plot for the protein–ligand complex shows more regions and higher peaks of RMSF values. This indicates that the ligand contributes to the conformational changes of the protein after interacting with it. Currently, the peaks in the RMSF plot for WT apo TkA indicate loops and more flexible regions on the outer surface of the protein. When the 3D structure of the protein was compared with the chimera and RMSF graphs, it was determined that the peaks occurring at 0.2, 0.3 and 0.4 nm indicate the rings in the outer region of the protein. Greater flexibility occurred in these regions after the interaction between the protein and the ligand. The RMSF value of active site residues Thr11, Tyr21, Ser54, Thr55, Thr85, Asp86 and Lys156, as well as Tyr233 and Glu275 from the B chain,

were also examined in Table 4.5. Comparison of these regions with their numerical values.

Table 4.5. RMSF values of active site residues before ligand binding (Apo State) and after ligand binding.

Residue Name	WT Apo TkA	WT TkA with L-asparagine
Thr11	0.0497	0.0818
Tyr21	0.0639	0.2780
Ser54	0.0486	0.0848
Thr55	0.0486	0.1035
Thr85	0.0451	0.0569
Asp86	0.0576	0.0729
Lys156	0.0594	0.0554
Tyr233'-Bchain	0.0434	0.0679
Glu275'-Bchains	0.0525	0.1756

A comparative table of active site residues when WT TkA is in the presence of l-asparagine ligand and when WT TkA is in the apo state is given in Table 4.5. Previous work on *Bacillus Velezensis* l-asparaginase evaluated the flexibility of calculated RMSF values at active site residues. The team argued that since active site residues are less flexible, RMSF values should decrease after binding with the ligand. They also stated that this situation supports more interaction with the ligand in the active site. Contrary to this study, the RMSF graph of WT TkA and l-asparagine ligand showed higher peaks in the active site residues (Srivastava et al. 2022; Townsend et al. 2015; Hozoorbakhsh et al. 2023).

Figure 4.13 shows the changes in the structure of the mobile part of the protein (orange), the allosteric site (blue) and the hairpin (yellow) during the enzyme-ligand interaction at certain time intervals. When these moving parts are compared with the RMSF graphs, it is seen that they give higher peaks in the RMSF graph of WT TkA.

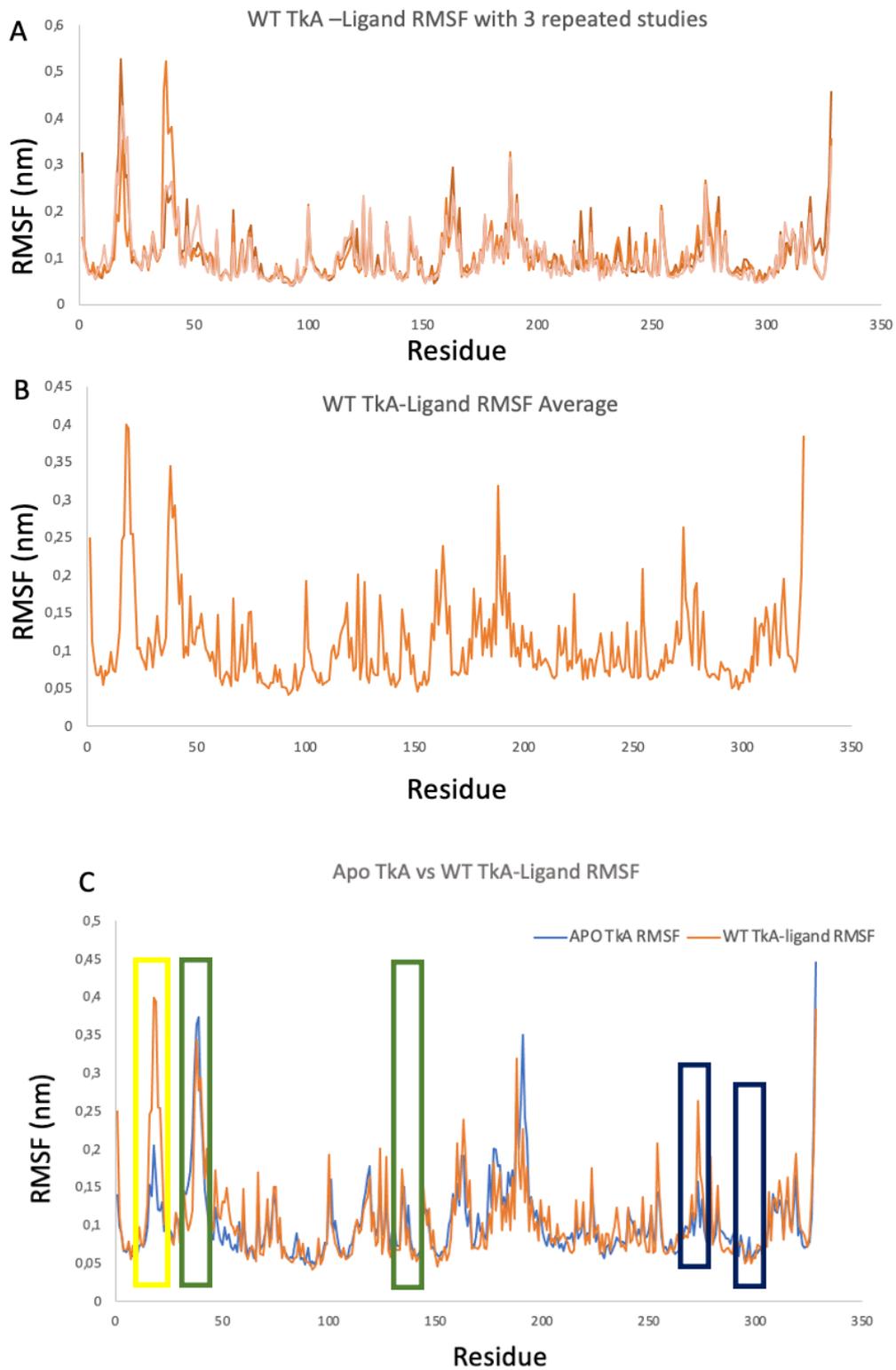


Figure 0.12. A) RMSF graphs of three replicates for WT TkA with l-asparagine ligand. B) Average RMSF graph for WT TkA with l-asparagine ligand. C) Comparing RMSF graphs between WT TkA and its l-asparagine ligand and WT apo TkA.

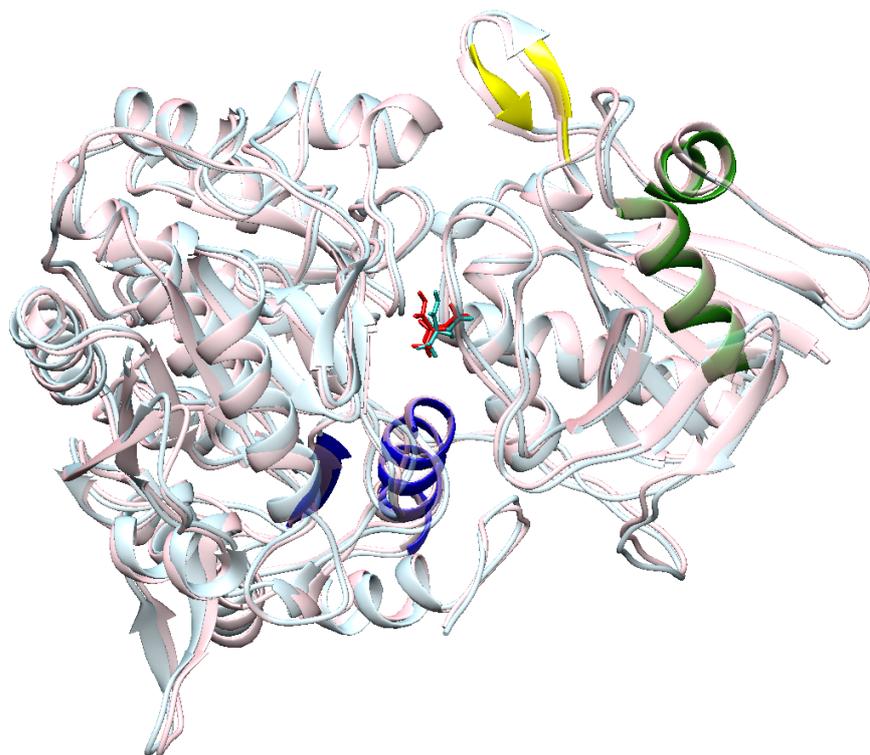


Figure 0.13. Initial position image of WT TkA (light pink) and l-asparagine ligand (red) and image of WT TkA (light blue) and l-asparagine ligand (cyan) at 10ns. The part shown in yellow here represents the hairpin: Residues: $\beta 2(14-16)$ - $\beta 3(20-22)$. Blue regions indicate the allosteric site: Residues: $\alpha 8(293-306)$ - $\beta 15(258-262)$. Green regions $\alpha 1(27-33)$ – $\alpha 4(122-133)$ are regions that tend to move towards the active site during ligand and enzyme interaction.

Figure 4.14.A belongs to the radius of gyration graph of the protein-ligand complex obtained from three separate experiments. These values stabilized around 2.60 nm in all three simulations. Consistency was observed between the three analyses.

When comparing WT TkA with ligand and WT apo TkA, the radius of gyration curves values is shown in the graph in figure 4.14.B. WT TkA radius of gyration value was observed at higher values than Apo TkA radius of gyration value. When evaluated according to the starting points, it can be deduced that the radius of gyration value of the liganded protein increases after ligand binding. This is because, as Hozoorbakhsh and colleagues explain, a less folded form and less compactness are expected as the entropy of the protein increases (Hozoorbakhsh et al. 2023).

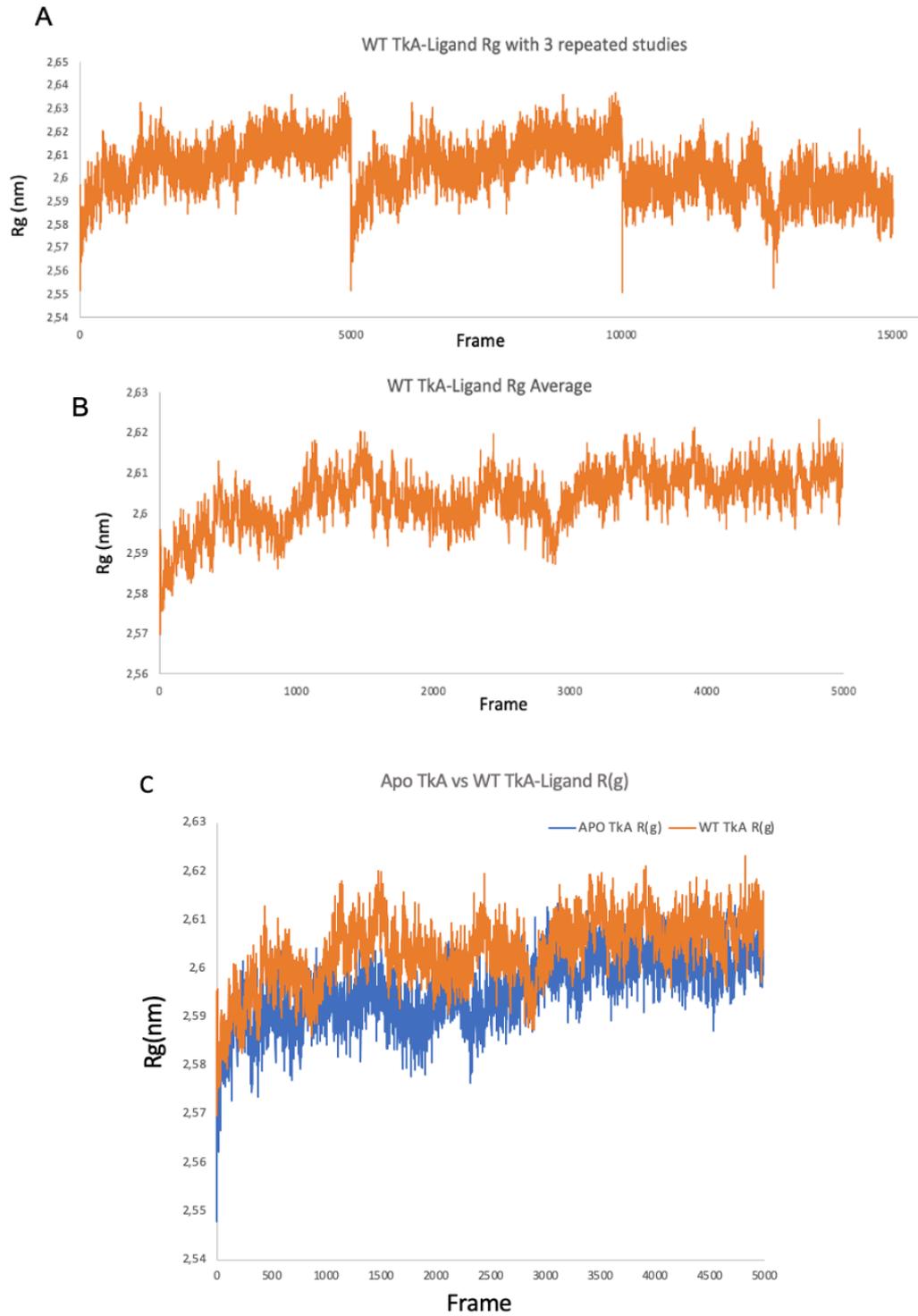


Figure 0.14. A) Radius of gyration graphs of WT TkA with l-asparagine ligand from three different replicate runs. B) Average radius of gyration from three replicate runs. C) Comparative radius of gyration graph of WT apo TkA and WT TkA l-asparagine ligand complex.

The analysis of the number of hydrogen bonds between WT TkA and l-asparagine ligand is shown in figure 4.15. In addition, the average hydrogen bond values obtained from three separate replicate studies were 3.6, 3.4 and 3.12 for run_1, run_2 and run_3, respectively. Ligplot+ images showing the hydrogen bonds of the ligand at different times for each 0 ns, 15ns, 30ns, 45ns and 50 ns can be seen in figure (4.16).

When the graph in figure 4.15 is examined in detail, it shows the number of hydrogen bonds between atom pairs within 0.35 nm. When you look at the ligplot+ images obtained, hydrogen bond formation between L-asparagine can be seen in the Asp86 and Glu294 residues.

The change in hydrogen bond interactions was thought to result from the enzyme's behavior to maintain stability throughout the simulation period. Additionally, water molecules in the active site bound to the simulation may have caused competition in H-bond formation. The fact that MD simulations provide a more realistic dynamic environment compared to docking studies explains the reason for this difference. It is planned to investigate whether the ligand is in competition with water or not in more detail in future studies (Nakagawa & Tamada et al., 2021).

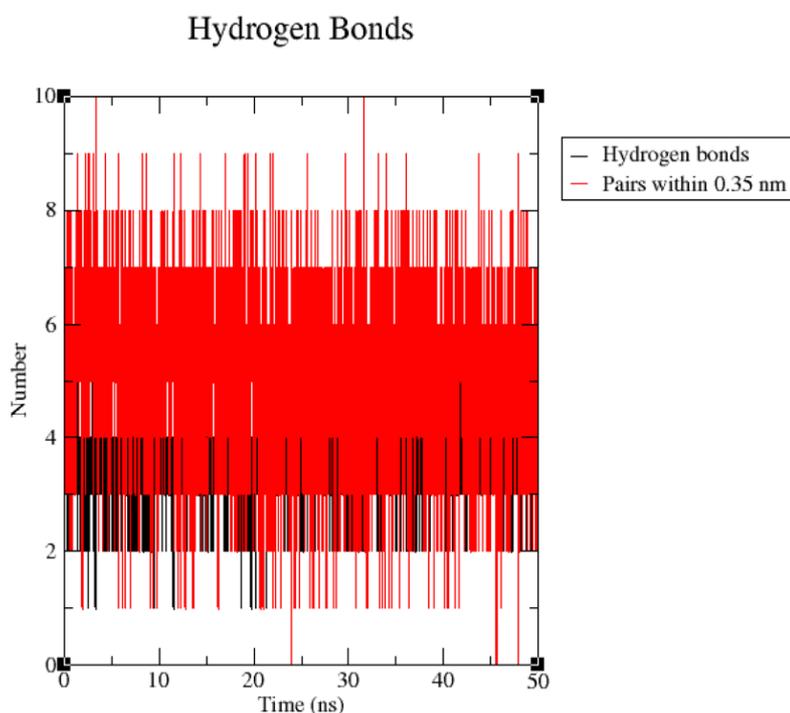


Figure 0.15. The number of hydrogen bonds between atom pairs within 0.35 nm distance.

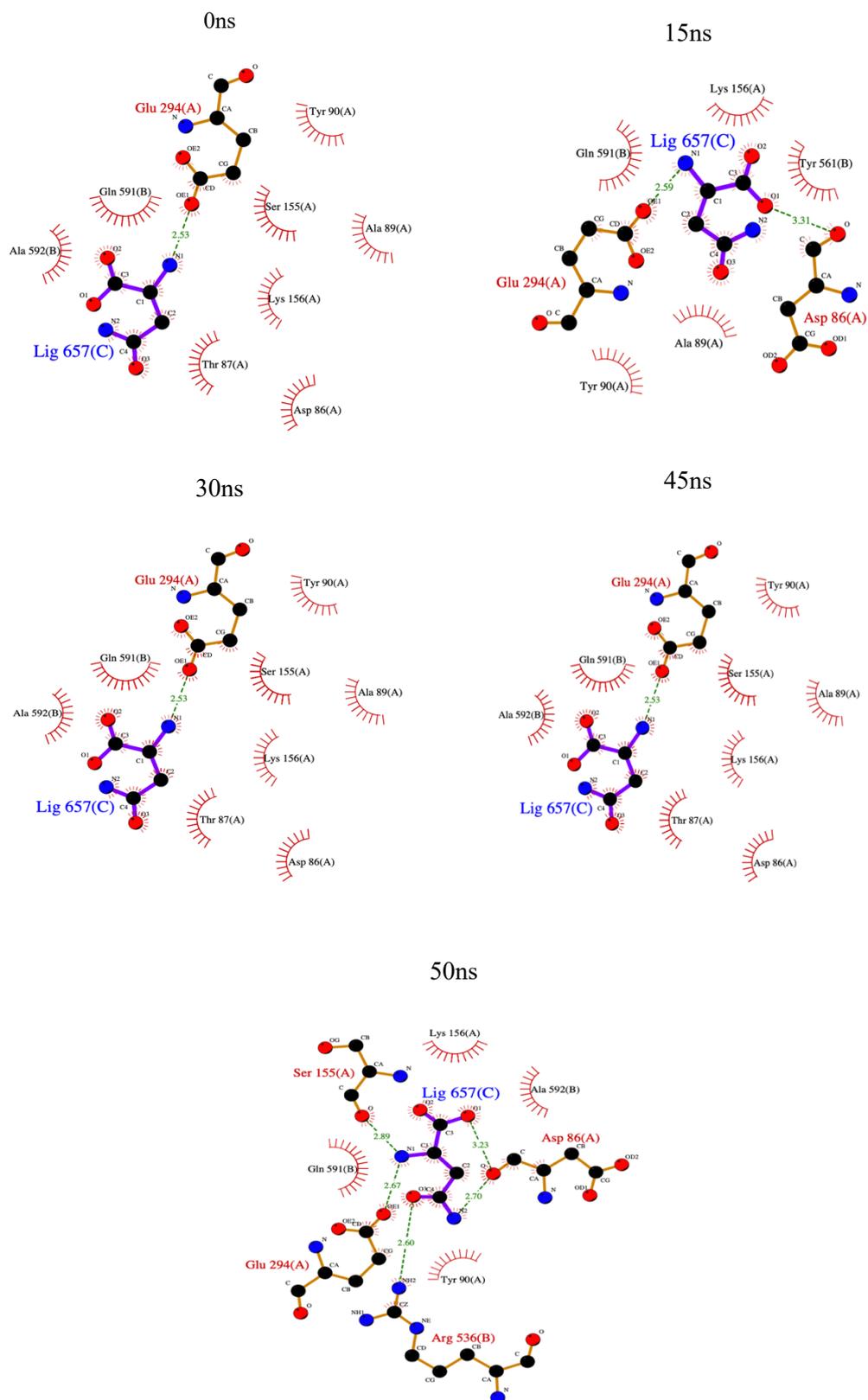


Figure 0.16. The images show hydrogen bond formation between l-asparagine ligand and some residues that are in active site of WT TkA.

Erva et al. In their similar analysis for *E. coli* l-asparaginase, they observed a similar decrease in the number of hydrogen bonds after MD. In addition, the table shows the average number of hydrogen bonds obtained by the team from their studies with Erwinase chrysanthemi and EcA as 5.5 and between 4 and 5, respectively (Erva et al. 2016).

In a study by Chi et al. to improve the thermal stability of thermostable *Bacillus licheniformis* l-asparaginase, they suggested that weak hydrogen bond formation and hydrophobicity contribute to thermostability while maintaining enzyme flexibility (Chi et al. 2022).

The comparative graph of SASA values of three repetitions of WT and L-asparagine ligand is shown in figure 4.17.A. Distribution throughout the simulation and the equilibrium to 258.92 nm² made us think that these three analyzes were consistent. The current imbalance at some points in this graph indicates that the conformation at the starting point of the ligand in the protein is different from the others in the initial condition, which affects the SASA value due to the conformational change of the protein relative to the ligand.

Surface accessibility analysis of WT TkA and L-asparagine ligand bound is shown in figure 4.17.B. A decreasing SASA value generally indicates that the protein is contractin over time and indicates that the molecular surface interacts less with the solvent as well as conformational tightening. When SASA values are compared with WT TkA in ligand-free and ligand-bound states in solvent, an increase in SASA value is observed after the protein interacts with the ligand. This indicates that the protein expands upon ligand binding and the accessibility of its surfaces to solvent increases. The SASA value for wild TkA with ligand still resulted in a higher SASA value than the apo TkA. Consequently, this means more exposure of the hydrophilic regions of the protein. (Trisciuzzi et al. 2018; Du et al. 2020; Durham et al. 2009).

In the literature, a SASA value of 263.4±4.0 nm² was obtained from the study of Jiao et al. *Acinetobacter soli* with L-ASNase. Another study by Chi et al. SASA value for *Bacillus licheniformis* L-ASNase was found to be 233 nm². When WT TkA and L-asparagine ligand were compared with these values, results close to Jiao et al. were obtained, and consistent with both reference studies, WT TkA and L-asparagine ligand showed a stable profile (Chi et al. 2022; Jiao et al. 2022).

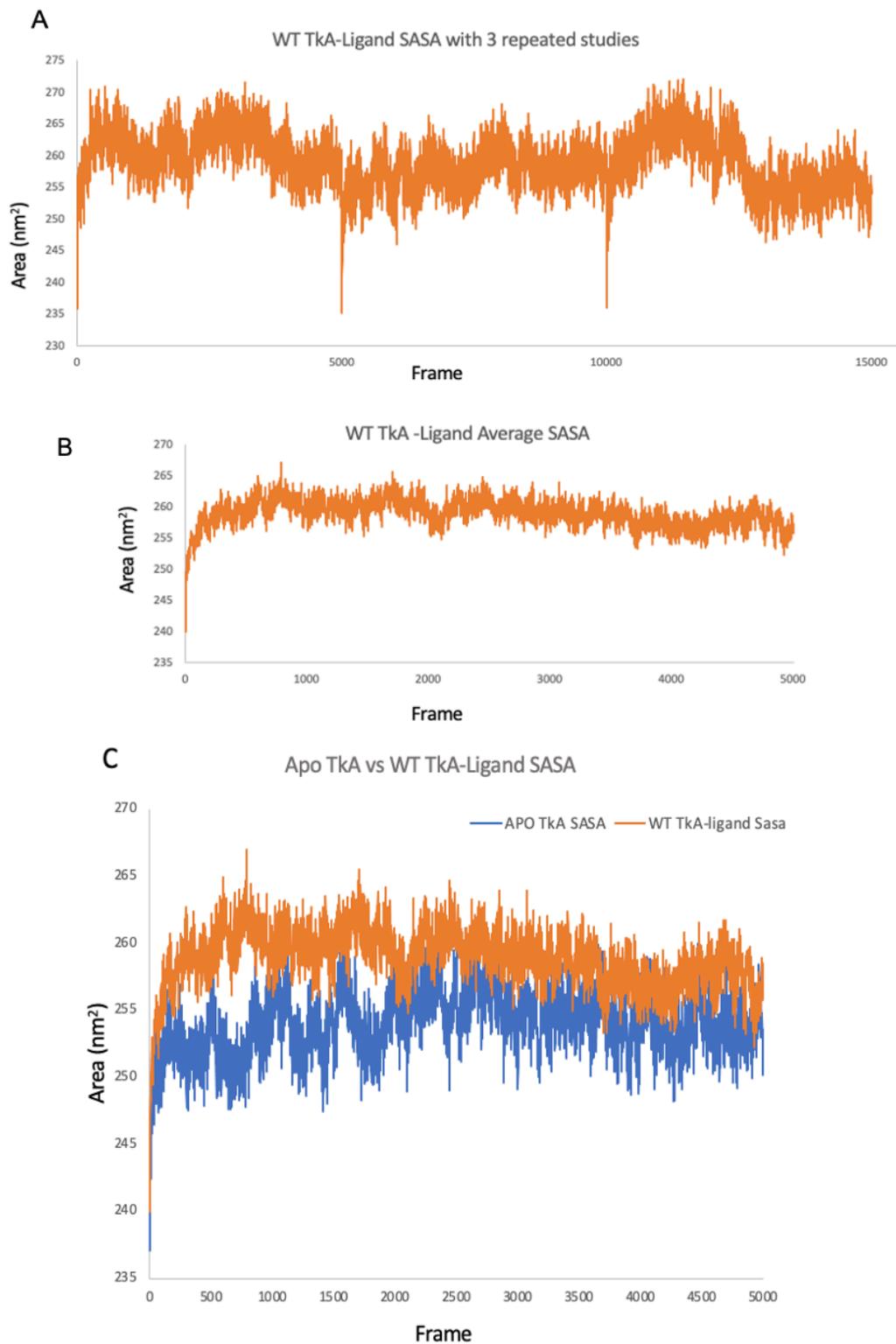


Figure 0.17. A) SASA graphs of WT TkA with l-asparagine ligand from three different replicate runs. B) Average SASA from three replicate runs. C) Comparative SASA graph of WT apo TkA and WT TkA l-asparagine ligand complex.

Focused on two important energy terms and analyzed protein and ligand interaction energy: LJ-SR (short-range Lennard-Jones interactions) and Coulomb-SR (short-range Coulomb interactions). The red line defines the LJ-SR potential, which fluctuates around -45.40 kJ/mol. It represents Van der Waals interaction energies. Negative value indicates attractive van der Waals interactions. Therefore, it is seen that the protein and ligand are under the influence of attractive forces within 3 runs throughout the simulation.

The black line reflects the Coul-SR values, and the average is -292.58 kJ/mol. This indicates that electrostatically attractive interactions between protein and ligand are more intense in all three runs. Negative values indicate the presence of these attractive forces and a stable electrostatic interaction profile. The LJ-SR potential exhibits a lower interaction energy compared to Coul-SR.

Briefly, it shows that the stability of the protein–ligand complex during simulation largely depends on Coul-SR interactions. In addition, the average total energy value was calculated as -337.98 kJ/mol that is the value coming from the sum of two energies, and this value was compared with TkA mutants (Erva et al. 2016; Reddy et al. 2016; Bepari and Reza 2021; Londhe et al. 2019).

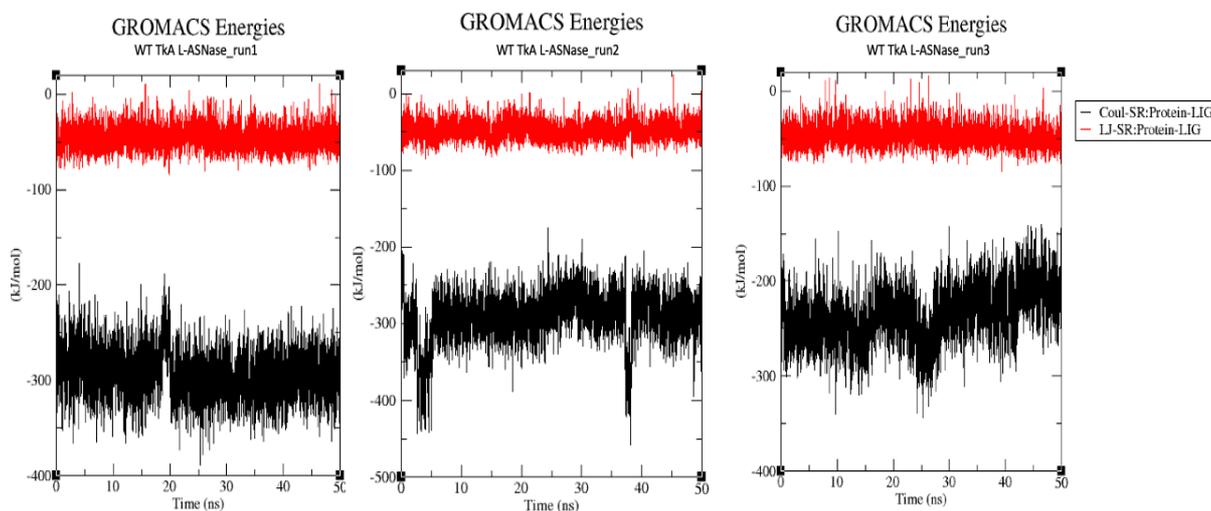


Figure 0.18. WT TkA and l-asparagine ligand interaction energies plots with three repeated studies.

The RDF graph created by referencing the ligand relative to the protein was calculated for three different runs in figure 4.19.A. The RDF plot of wild-type TkA L-asparaginase and its ligand is shown in figure 4.19.B. The reference ligand was taken, and RDF analysis was performed according to different residues of the active site. The graphs express the radial distance in nanometers (nm) on the horizontal axis (x), and the radial distribution function in $g(r)$ on the vertical axis. RDF plots provide indirect information about how residues in the active site, which play an important role during ligand-protein binding, are distributed in the protein cavity, surface, or interior. RDF analyzes were used in protein ligand interaction studies. Interactions between molecules forming H-bonds in the active site were evaluated. Thus, the proximity of the ligand to this residue and the probability of its occurrence frequency were evaluated (Shamim, Abbasi, and Azam 2015).

In the graph, the ligands and residues that form the hydrogen bond formation between the protein and the ligand, these Glu294 (black peak) and Asp86,(red peak) as well as Thr11(blue peak), Thr85(green peak) residues known to initiate nucleophilic attack, and the Thr55(violet peak) residue in which the mutation attempt was made, were presented in figure 4.19.B.

According to the figure 4.19.B, Glu294 (black peak) residue interacts with the ligand at shorter distances than other residues. However, the fluctuations at the peak point indicate that the interaction of this residue with the ligand is not stable at the measured distances. This indicates that the H-bond formed between the ligand and the Glu294 residue is unstable.

Asp86 (red peak) is seen as the residue with the lowest peak and largest area in the graph. In Ligplot+ and molecular docking studies, especially this residue and the ligand were in hydrogen bond formation. This situation suggested that the hydrogen bond interaction formed between the ligand and Asp86 was not stable.

Thr11(blue peak) and Thr55(violet peak) have a more defined as well as narrow peak in the graph and have more possibilities to be found with the ligand at long distances compared to other residues. It is in a more stable interaction state, while ligand shows activities.

Thr85 (green peak) has the highest probability of being near to the ligand during its activity. Indicating that the ligand spends more time in the vicinity of this residue.

In general, ligand with the highest probability and stabilities of its activities occur near to residue Thr55 and Thr85. When more defined states of the peaks were identified,

the most stable residue was determined as Thr55 (Shamim, Abbasi, and Azam 2015; Ahmad et al. 2017; Karataş et al. 2017).

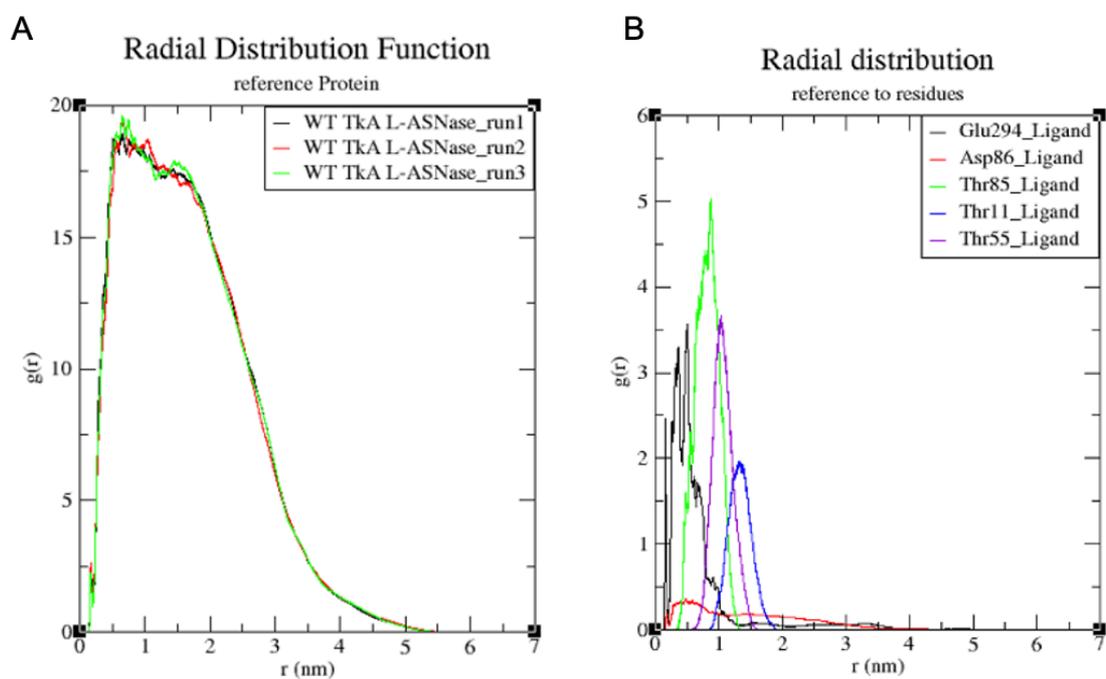


Figure 0.19. A) RDF plots are a comparative plot of 3 replicates. B) is the RDF graph of 5 different residues that may be important.

In addition, diffusion coefficient ratios of the ligand for complex were calculated by MSD (Mean Square Displacement) analysis and Einstein relationship (Karataş et al. 2017; Michalet 2010).

As a result of the average of diffusion coefficients was obtained, and this value was compared with mutants TkA values. Average diffusion coefficient value is $3.3 \times 10^{-7} \text{cm}^2/\text{s}$.

4.4.3. Molecular Dynamics Simulations for T55E TkA and Asparagine Ligand Complex

MD simulations, RMSD, RMSF, SASA radius of gyration, number of hydrogen bonds, RDF, diffusion coefficient, protein-ligand interaction analyzes of the T55E TkA and its ligand asparagine complex were performed. Additionally, images showing the hydrogen bond interaction and conformational change of the protein were acquired at time periods of 0 ns, 15 ns, 30 ns, 45 ns and 50 ns using ligplot+ as well as Chimera.

RMSD studies were performed to understand stability. Figure 4.20 represents the backbone-to-backbone RMSD plot for the T55E TkA. The protein backbone showed a slight increase between 2500 and 3500 frames. Compared to WT TkA, T55E TkA showed a different release from WT TkA in the frame between 1500 and 2000. However, in general, the two graphs gave values close to each other in figure 4.20.B. When T55E TkA and WT TkA were compared, WT TkA and its ligand remained more stable throughout the simulation (Erva et al. 2016; X. Zhang et al. 2021; Lubkowski et al. 2020).

RMSF analysis was performed to examine the fluctuations in the residues of the T55E TkA or the maintenance of the stable state in figure 4.21.A. RMSF of wild-type TkA was compared to the mutant-type protein in figure 4.21.B. RMSF values generally fluctuated quite close to each other. However, the mutant protein strain showed higher peaks at some points. When these regions were examined, it was seen that they generally contained rings belonging to the outer surface of the protein. When we look at the RMSF value of the region belonging to the T55E position, the pre-mutation RMSF value of Thr was found to be 0.1035, while the post-mutation value of the Glu residue was found to be 0.10985. The minor increase in this residue indicates that it has become more rigid after mutation. It is indicated with a black box in figure 4.21.C. Also, the part shown here with the yellow box represents the hairpin: Residues: β 2(14-16) - β 3(20-22). Blue boxes belong to the residue region of the allosteric site: Residues: α 8(293-306) - β 15(258-262). Green regions α 1(27-33)– α 4(122-133) indicate residues from the region known to move towards the active site during ligand and enzyme interaction in figure 4.20.C. And the black box belongs to the T55E region. This small increase in value meant that this residue became less flexible after the mutation. Additionally, images obtained from Chimera were examined in figure 4.22. As observed after the interaction of WT TkA with the

ligand, the hairpin, and allosteric regions of the mutant TkA were observed to move towards the active site after the interaction with the ligand (X. Zhang et al. 2021; Erva et al. 2016).

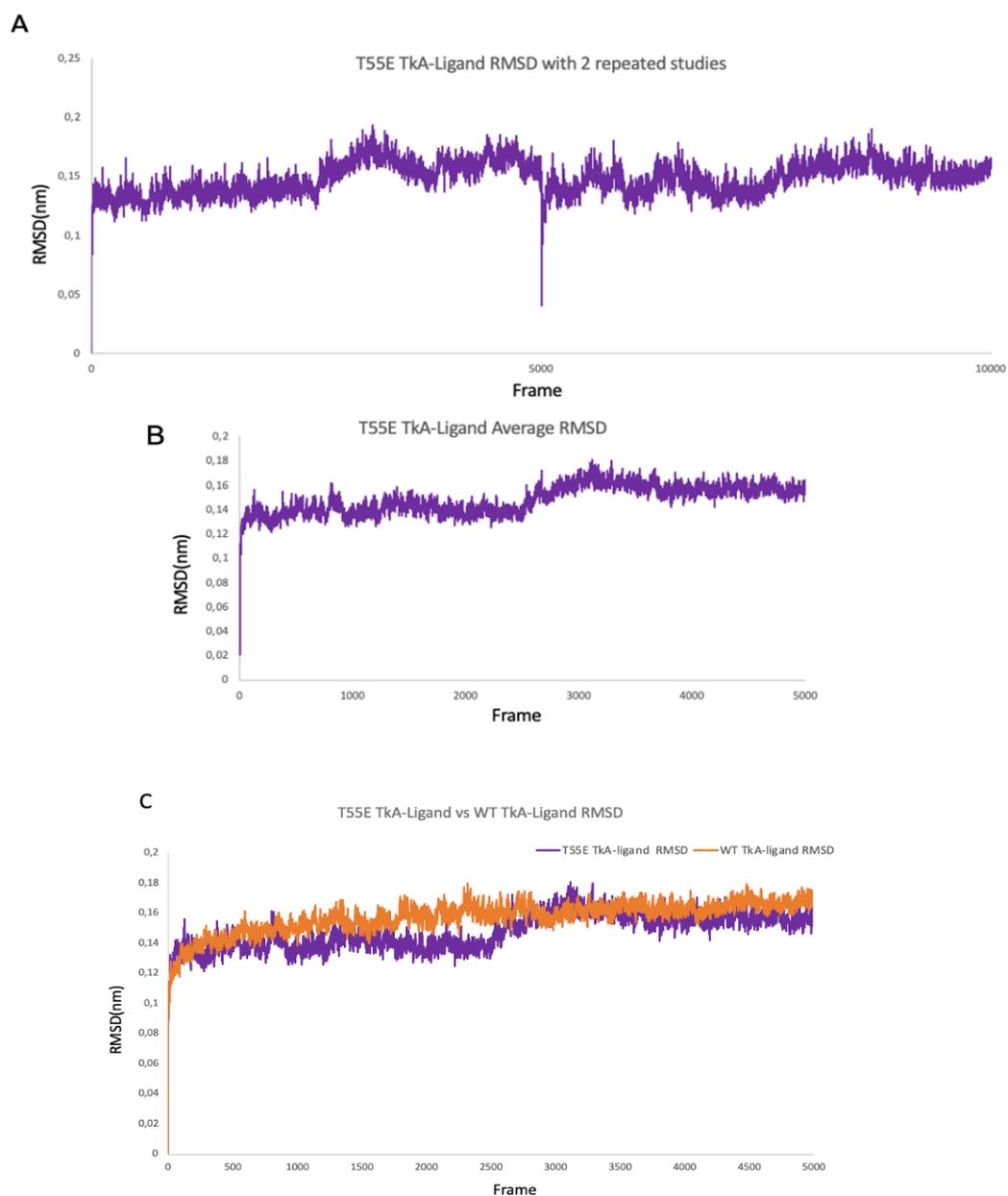


Figure 0.20. A) RMSD graphs of T55E TkA two replicates. B) T55E TkA average RMSD graph obtained from two repetitions. C) RMSD analysis; Comparison with WT TkA and T55E mutant TkA.

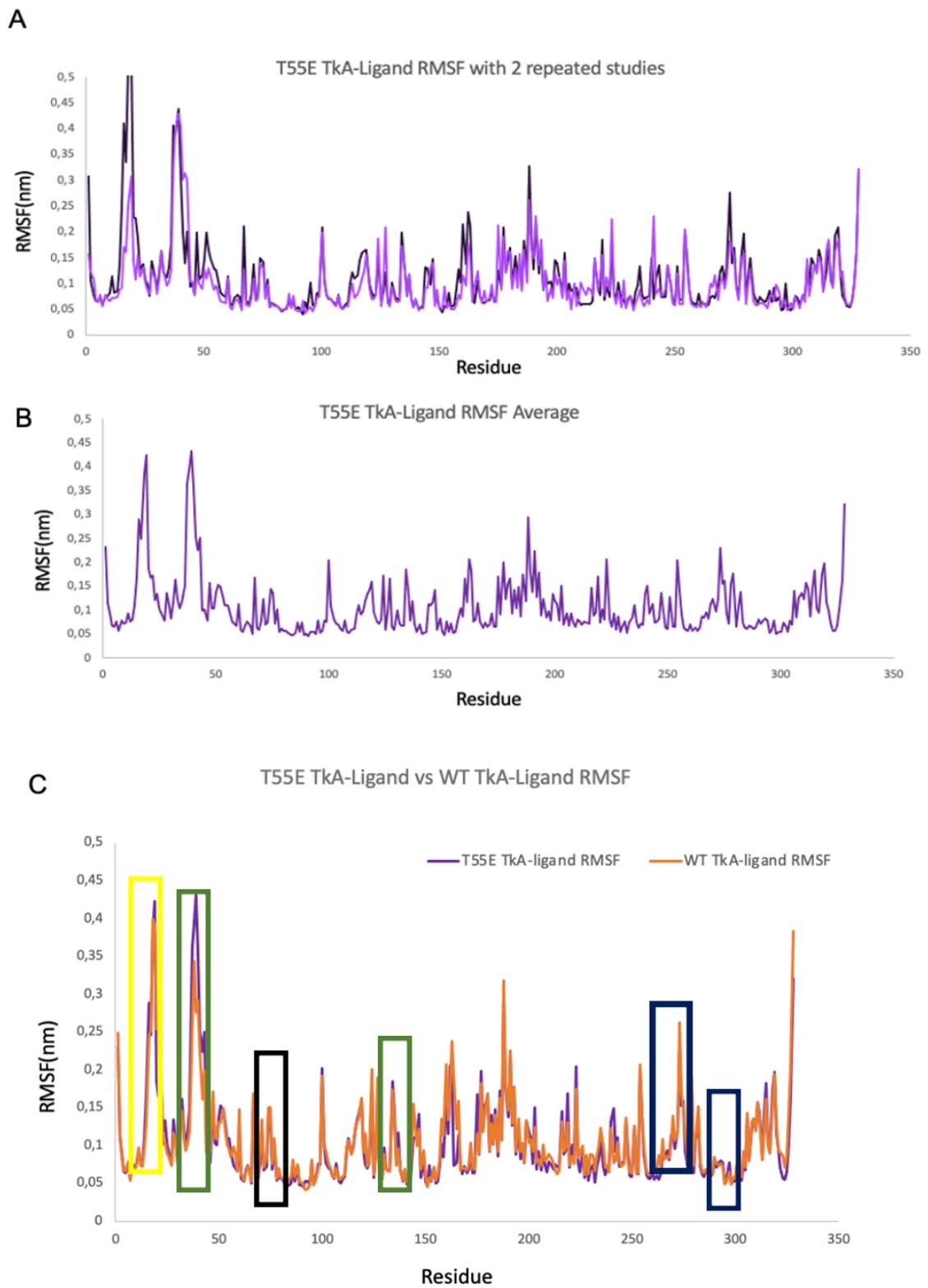


Figure 0.21. A) RMSF graphs of T55E TkA two replicates. B) T55E TkA average RMSF graph obtained from two repetitions. C) RMSF analysis; Comparison with WT TkA and T55E mutant TkA.

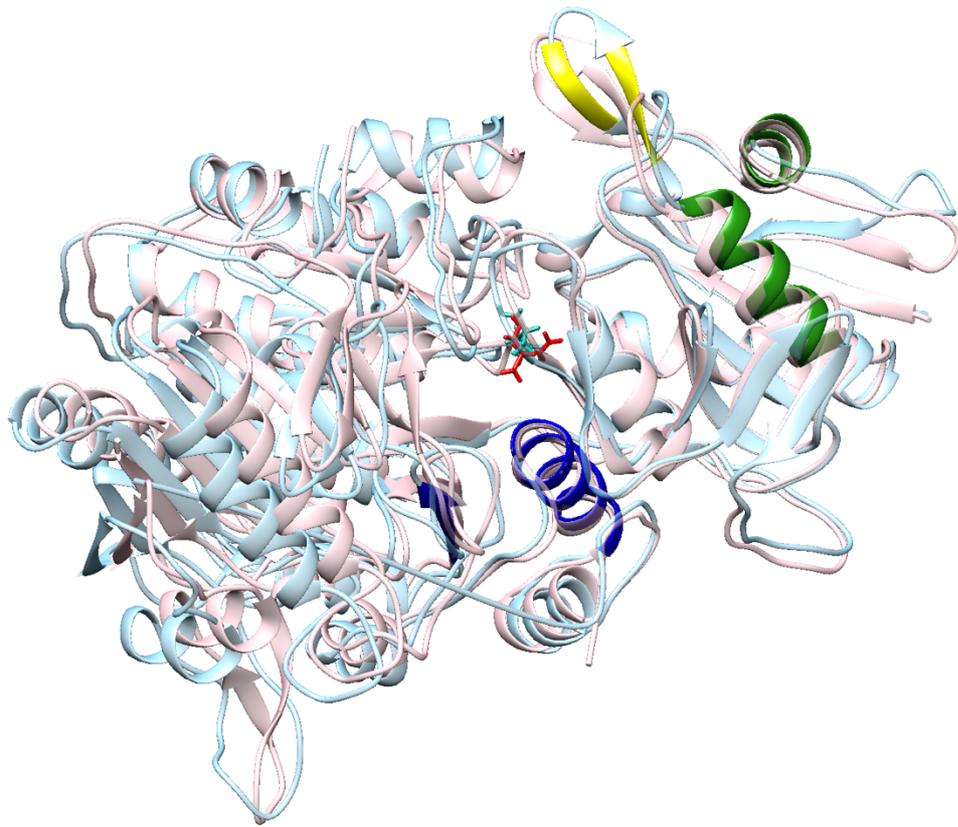


Figure 0.22. Initial position image of T55E TkA (light pink) and l-asparagine ligand (red) and image of T55E TkA (light blue) and l-asparagine ligand (cyan) at 10ns. The part shown in yellow here represents the hairpin: Residues: $\beta 2(14-16)$ - $\beta 3(20-22)$. Blue regions indicate the allosteric site: Residues: $\alpha 8(293-306)$ - $\beta 15(258-262)$. Green regions $\alpha 1(27-33)$ – $\alpha 4(122-133)$ are regions that tend to move towards the active site during ligand and enzyme interaction.

T55E TkA was lower with an average radius of gyration value of 2.59 and WT TkA with an average radius of gyration value of 2.61. This indicates that the T55E TkA is more compact than the WT TkA, implying that the structure exhibits a much tighter conformation after mutation (Erva et al. 2016).

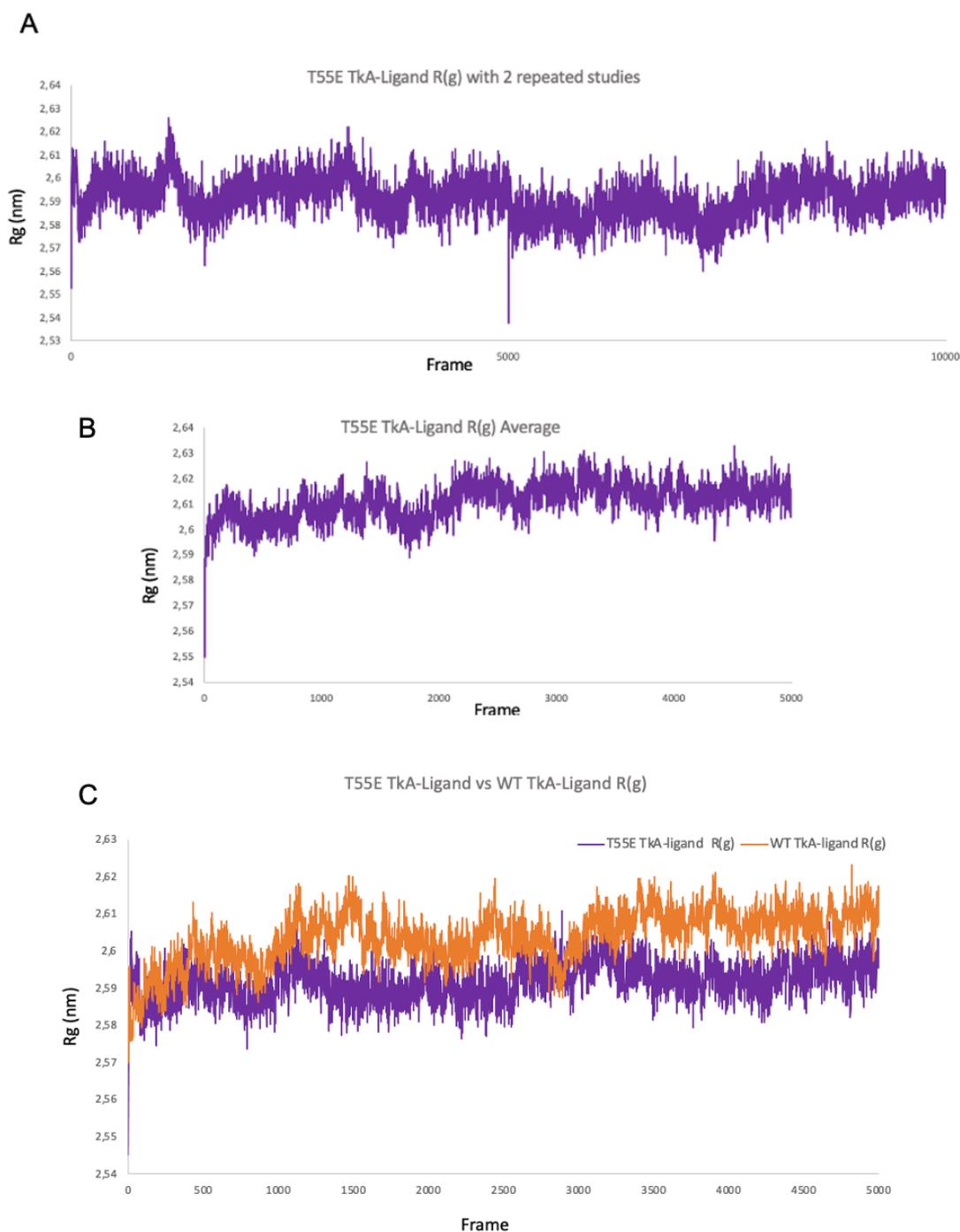


Figure 0.23. A) Radius of gyration plot for T55E TkA with two repeats. B) Average plot of radius of gyration for T55E TkA and its ligand. C) Comparative radius of gyration for T55E TkA and WT TkA.

The graph of the number of hydrogen bonds between the T55E TkA and the asparagine ligand is seen in Figure 4.24. The graph shows the number of H-bonds

between pairs of atoms within 0.35nm. Additionally, images obtained from LigPlot+ showing which atoms the hydrogen bonds are formed between are in figure 4.25. the values obtained from two separate MD analyzes were 3.3 and 3.1, respectively. The average value was determined as 3.2.

This value is slightly less compared to wild TkA (3.4). As in the WT TkA, hydrogen bonding was seen in the Asp86, Glu294 and extra Lys156 regions in the T55E TkA. After MD simulation, it was observed that some H-bond formations decreased and disappeared as in the wild type. It was thought that the change in hydrogen bond interactions resulted from the stability-maintaining behavior of the enzyme. Additionally, water molecules in the active site bound to the simulation may have caused competition in hydrogen bond formation (Jiao et al. 2022).

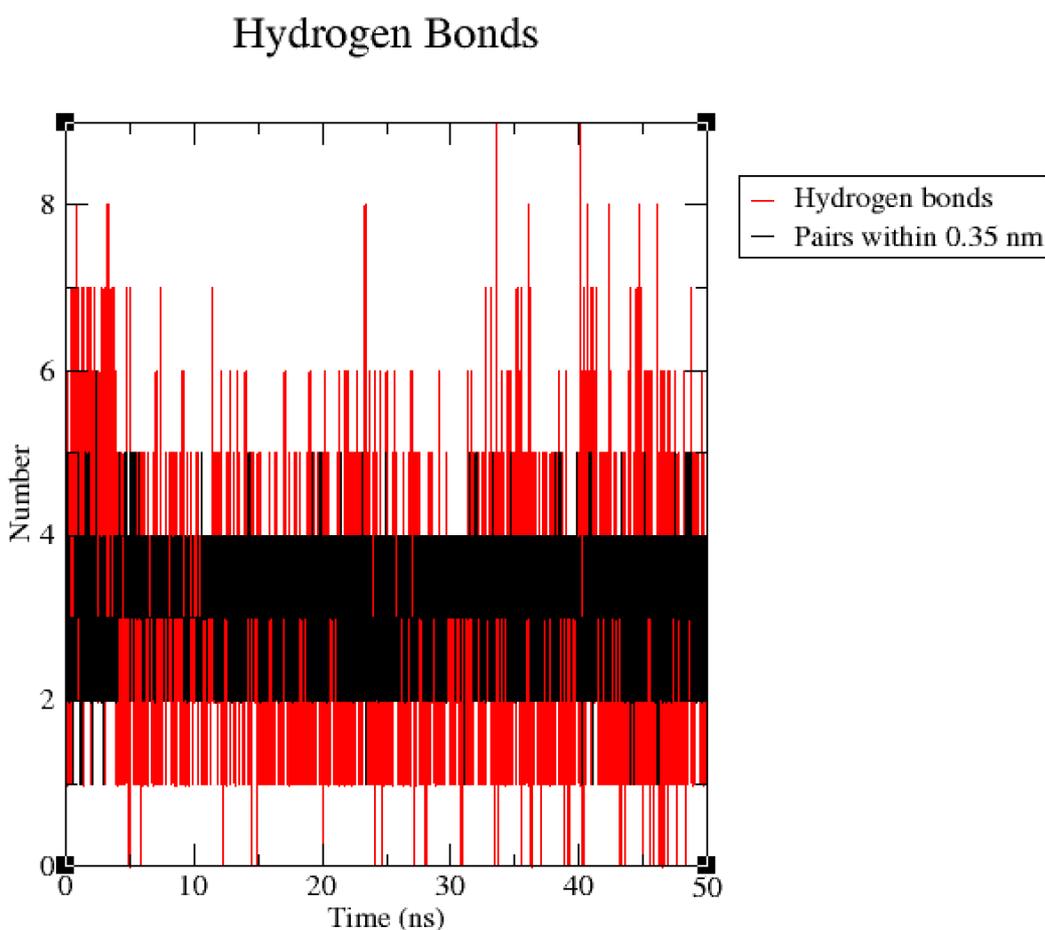


Figure 0.24. The number of hydrogen bonds between atom pairs within 0.35 nm distance.

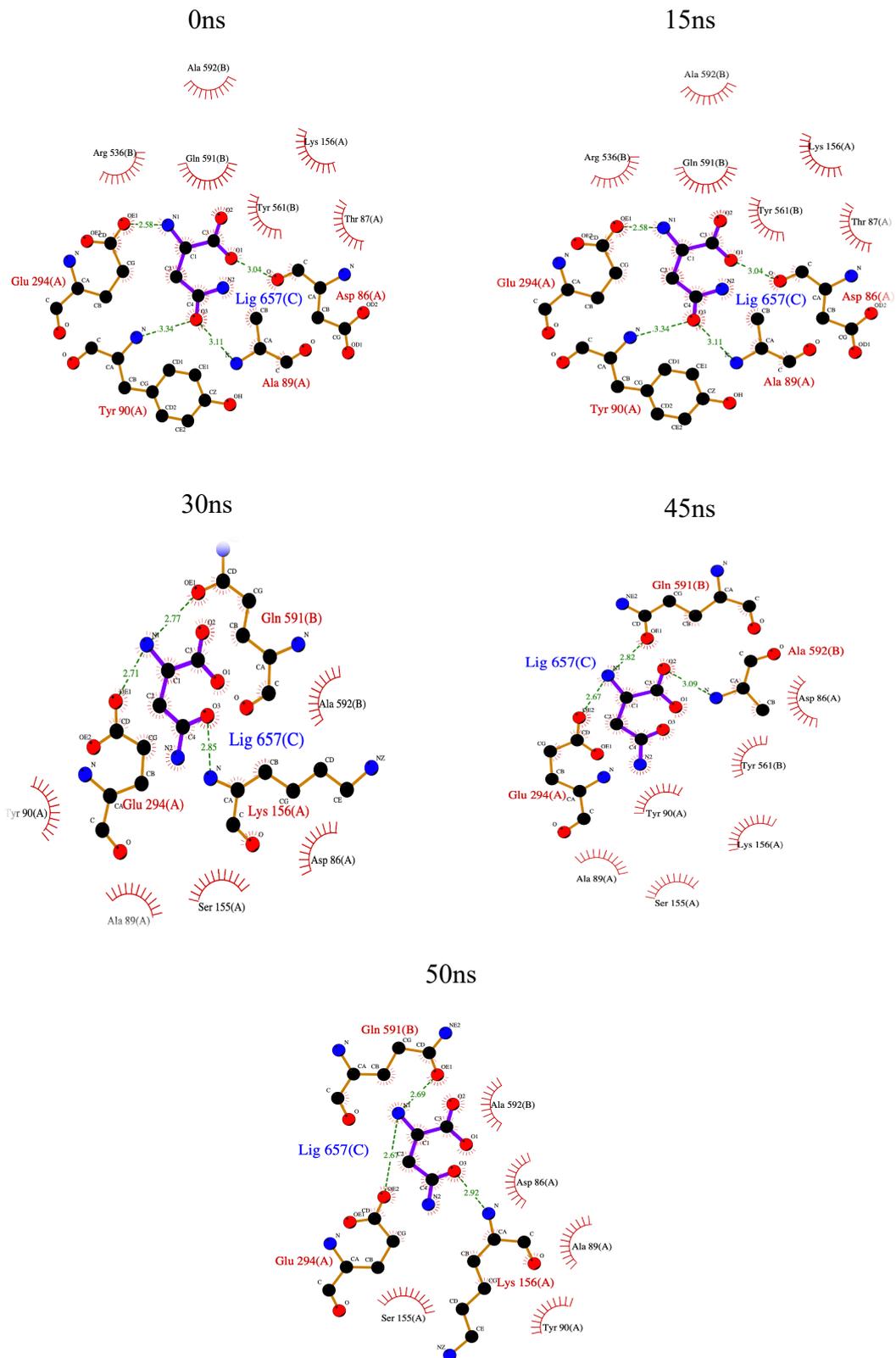


Figure 0.25. The images show hydrogen bond formation between 1-asparagine ligand and some residues that are in active site of T55E TkA.

SASA plot that derived for protein and ligand obtained from MD simulation of T55E TkA is shown in figure 4.26.A. When the SASA graph is compared with the SASA graph of WT TkA, in figure 4.26.B although the graphs have close values, the SASA value of the mutants of protein increased after 2000-2500 frame interval compared to the Wt TkA.

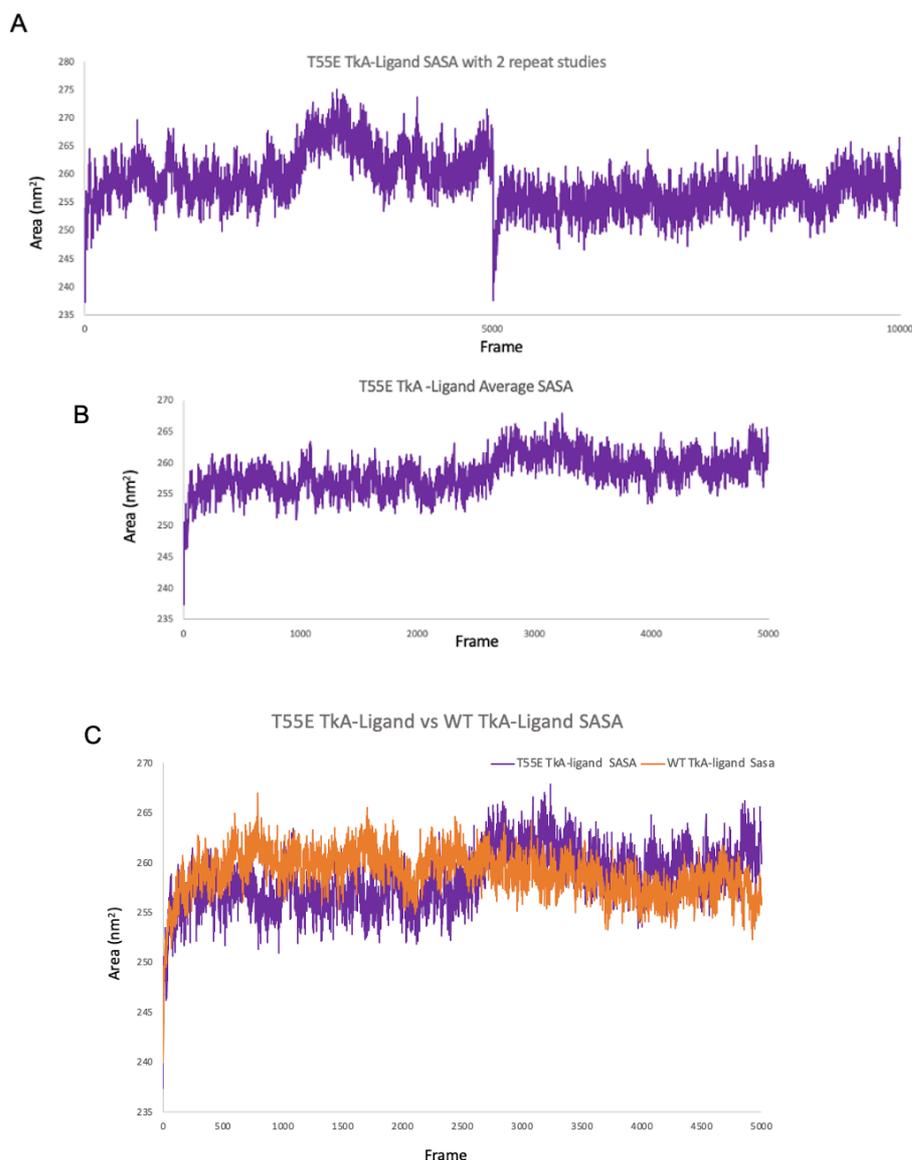


Figure 0.26. Images A show SASA plots with two repeated studies and average SASA plot respectively. Image B shows T55E TkA and WT TkA are compared on the same plot.

The SASA value was calculated as 258.40 nm²/ns for T55E TkA. This value was 258.92 nm²/ns for WT TkA. This increase in SASA indicates that especially the hydrophilic regions of the molecule interact more with the solvent as well as, the mutants means that the overall hydrophobicity of the protein is reduced after 2500 frame. Therefore, protein has become more expandable and can be occupied by ligand more. In contrary of T55E TkA, SASA value of WT TkA is decreased after 2500 frame (Durham et al. 2009; Trisciuzzi et al. 2018; Hozoorbakhsh et al. 2023; Topham and Smith 2019).

Coulomb-SR and LJ-SR interactions were examined for T55E TkA and its ligand asparagine in figure 4.27. The red line describes the LJ-SR van der Waals interactions and remained around -66.48 kJ/mol throughout the simulation. This value is very close but there is a small increase compared to WT TkA.

The black line reflects the Coul-SR electrostatic energy, which is with an energy value of around -179.54 kJ/mol. Additionally, the total energy value was calculated as -246.02 kJ/mol. Total energy values of WT TkA L-ASNase and T55E TkA were compared. WT TkA L-ASNase had higher protein ligand energy with a total energy value of -337.98 kJ/mol.

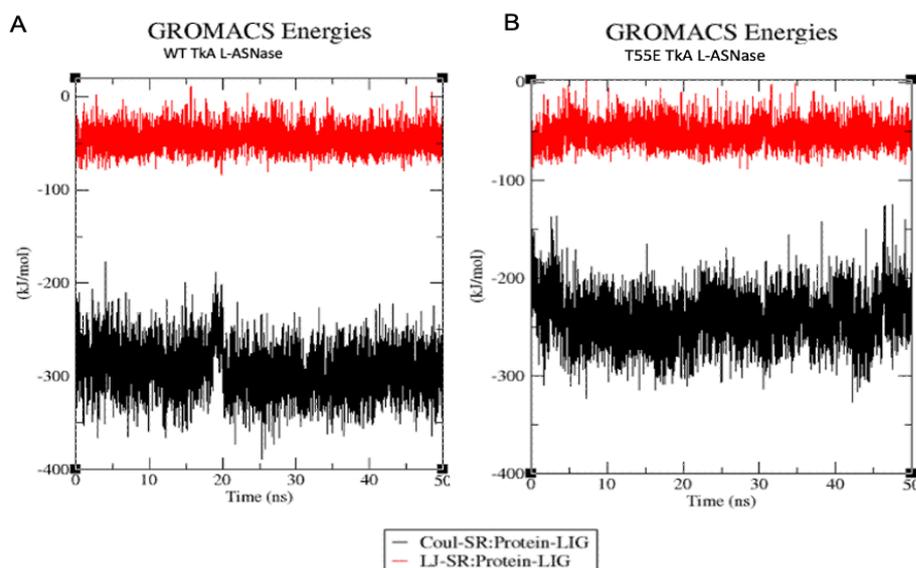


Figure 0.27. A) shows WT TkA an l-asparagine ligand interaction energy plot. B) shows T55E TkA interaction energy graph. Red regions represent LJ-SR and black regions represent Coul-SR energy.

A stronger van der Waals interaction is observed in the T55E TkA and L-asparagine ligand interaction energy, while a weaker electrostatic interaction dominates (Rampogu et al. 2022; Reddy et al. 2016; Erva et al. 2016; Londhe et al. 2019).

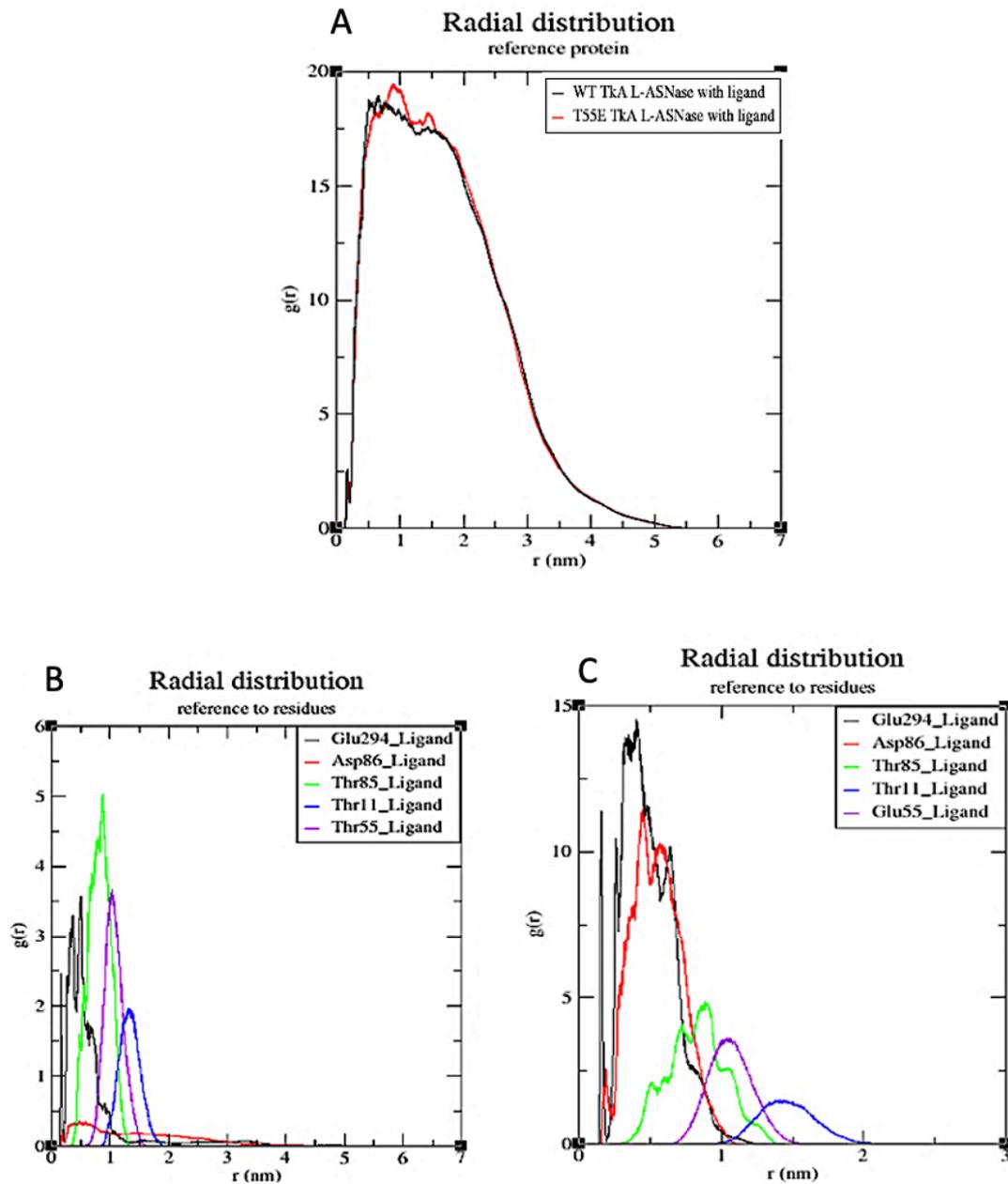


Figure 0.28. A) shows comparison of WT TkA RDF values and T55E TkA RDF values. B) RDF analysis is given specifically for residues for WT TkA. C) RDF analysis is given specifically for residues for T55E TkA.

The comparative RDF plot for WT TkA and T55E TkA with the ligand as reference is shown in figure 4.28. Considering the protein-ligand RDF plot as an overall comparison of the two simulation results, no significant differences were observed. However, only the sharpening and slight increase in the mutant protein peak suggests that there is a more likely interaction between the mutant protein and the ligand at certain distances. For a more effective analysis, the residues Thr11, Thr55, Thr85, Asp86 and Glu294 were analyzed for their efficacy in WT TkA and ligand interaction in figure (4.28.C).

Here the residues that stood out after mutation were Asp86 and Glu294, which gave much higher peaks. However, the fact that the high probability of presence leads to a fluctuation in the peaks shows that this presence is not stable.

On the contrary, for Thr85 and Thr11, the ligand was less likely to interact with these residues during the protein-ligand interaction. In addition, the Thr85 residue showed a highly variable peak, indicating that its stability had decreased significantly compared to the previously defined peaks. This situation was thought to be caused by Glu at position 55, whose activity decreased after the mutation (Ahmad et al. 2017; Shamim, Abbasi, and Azam 2015).

The diffusion coefficient for the ligand from T55E mutant TkA and ligand complex was calculated with MSD and Einstein equation. and the diffusion coefficient of the ligand was calculated as $5 \times 10^{-7} \text{cm}^2/\text{s}$ in this pair. This value, when compared to the ligand diffusion coefficient interacting with WT TkA, i.e. $3.3 \times 10^{-7} \text{cm}^2/\text{s}$, shows that the ligand of the mutant complex moves faster on the protein surface and exhibits more conformations, and expresses its ability (Chen et al., 2014; Karataş et al., 2017; Ma et al., 2005; Samanta & Roccatano, 2013; Tou et al., 2013).

4.4.4. Molecular Dynamics Simulations for D86S TkA and Asparagine Ligand Complex

MD simulations, RMSD, RMSF, SASA radius of gyration, number of hydrogen bonds, RDF, diffusion coefficient, protein-ligand interaction analyzes of D86S TkA and its ligand asparagine complex were performed. Additionally, images depicting the

hydrogen bond interaction and conformational change of the protein were acquired at time periods of 0 ns, 15 ns, 30 ns, 45 ns, and 50 ns using Chimera and ligplot+.

Figure 4.29.A represents the backbone-to-backbone RMSD plot for the D86S TkA. When the average RMSD value of the two repeat studies for mutant TkA was evaluated, it increased from the starting point to 2000 frames and then remained stable at an average value of 0.2 nm. Looking at figure 4.29.B, it can be seen that the RMSD values for WT TkA and TkA maintain similar stability despite minor differences throughout the simulation. However, D86S TkA is slightly higher than the WT TkA and ligand RMSD value (Ardalan et al., 2021; Chi et al., 2022; Jiao et al., 2022; Hozoorbakhsh et al., 2023).

RMSF values were evaluated to examine the fluctuations in the residues of the D86S TkA protein or the maintenance of the stable state in figure 4.30.A. RMSF of WT TkA was compared with the mutants in figure 4.30.B. While RMSF values generally fluctuated quite close to each other, D86S TkA showed higher peaks at some points.

The higher peaks seen at residues 100-125 and 225-250 in the RMSF graph are the allosteric region of the regions mentioned in the article of Guo et al. and determined for WT TkA under normal conditions, and the helix structures that move to the active site during the enzyme interaction with the ligand. These parts shown with the yellow box represents the hairpin: Residues: β 2(14-16) - β 3(20-22). Blue boxes belong to the residue region of the allosteric site: Residues: α 8(293-306) - β 15(258-262). Green regions α 1(27-33)– α 4(122-133) indicate residues from the region known to move towards the active site during ligand and enzyme interaction. And the black box belongs to the D86S region in figure 4.30.C.

In figure 4.31, the regions are shown in blue for the allosteric site and in orange for helix structures. When the mutation site was examined, the RMSF value of Asp before the mutation was 0.0729, but after it was mutated with serine, its value became 0.0856. The minor increase in this residue indicates that it has become more rigid after mutation. It is indicated with a black box in figure 4.30.C. In general, D86S TkA is more flexible than WT TkA (X. Zhang et al. 2021; Erva et al. 2016).

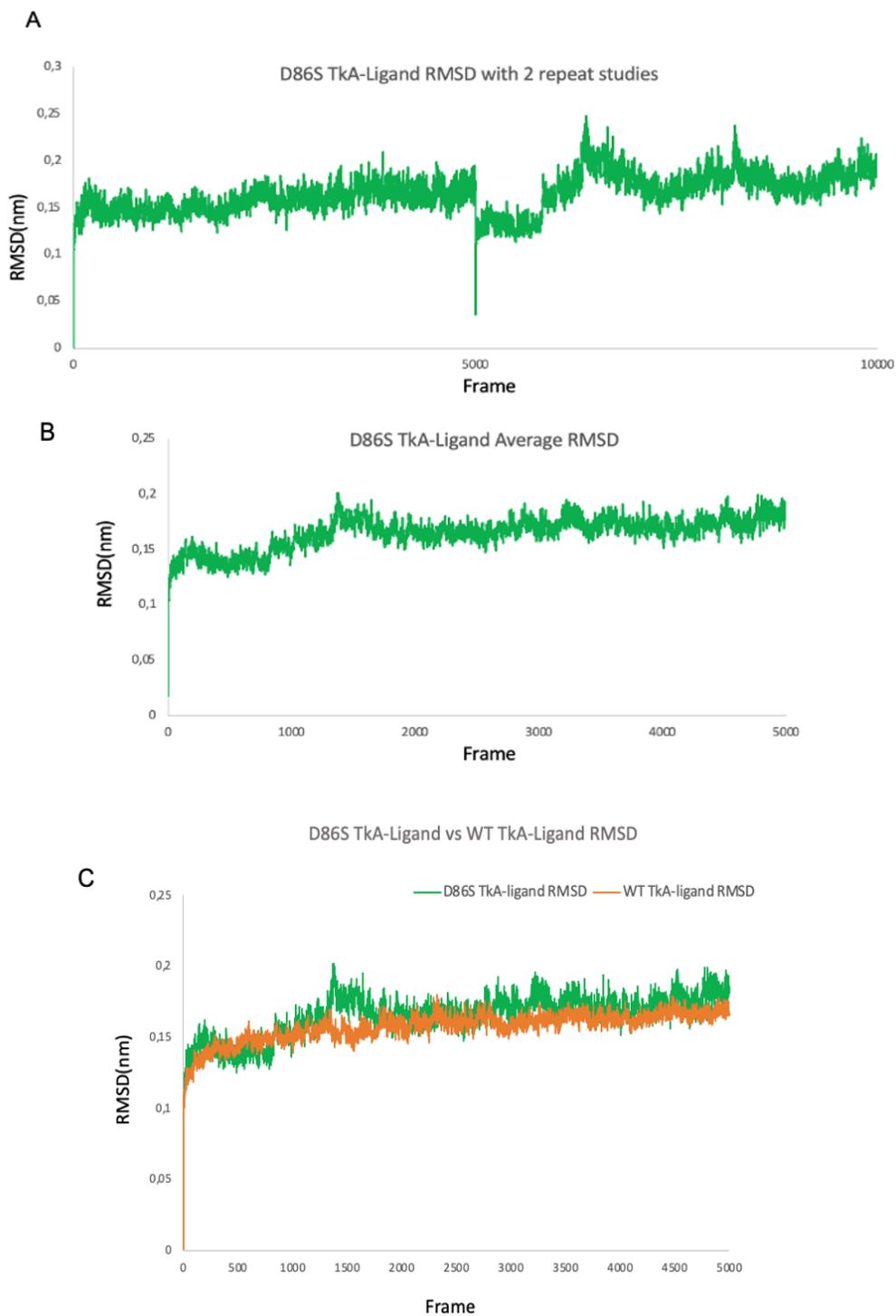


Figure 0.29. A) RMSD graphs of D86S TkA two replicates. B) D86S TkA average RMSD graph obtained from two repetitions. C) RMSD analysis; Comparison with WT TkA and D86S TkA with l-asparagine ligand.

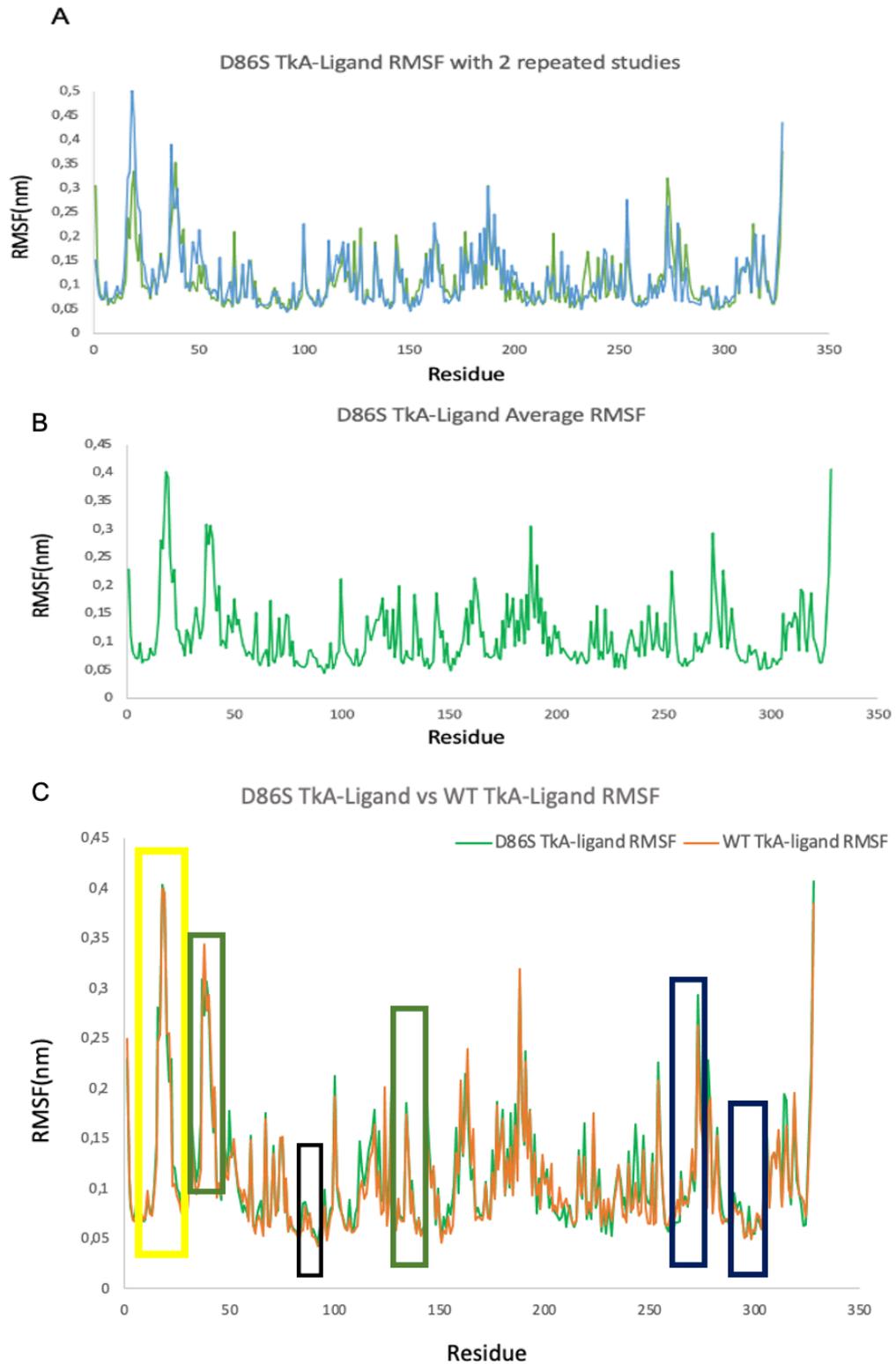


Figure 0.30. A) RMSF graphs of D86S TkA two replicates. B) D86S TkA average RMSF graph obtained from two repetitions. C) RMSF analysis; Comparison with WT TkA and D86S TkA.

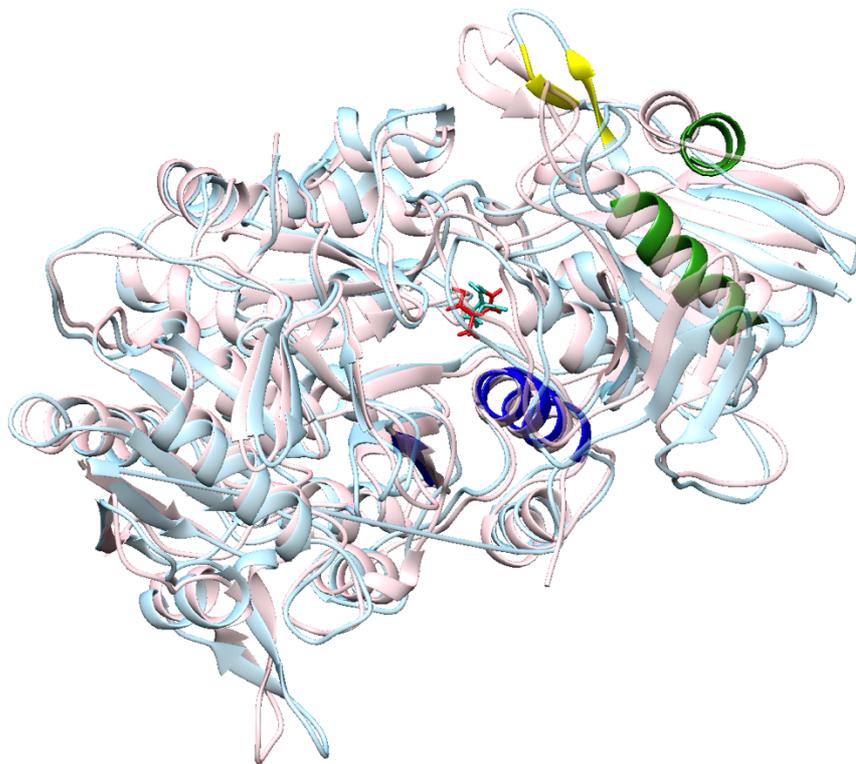


Figure 0.31. Initial position image of D86S TkA (light pink) and l-asparagine ligand (red) and image of D86S TkA (light blue) and l-asparagine ligand (cyan) at 10ns. The part shown in yellow here represents the hairpin: Residues: $\beta 2(14-16)$ - $\beta 3(20-22)$. Blue regions indicate the allosteric site: Residues: $\alpha 8(293-306)$ - $\beta 15(258-262)$. Green regions $\alpha 1(27-33)$ – $\alpha 4(122-133)$ are regions that tend to move towards the active site during ligand and enzyme interaction.

Radius of gyration graphs of two separate replicate runs are given in figure 4.32.A for D86S mutant TkA and asparaginase ligand. The average value for the D86S TkA enzyme, which showed a balanced radius of gyration throughout the simulation, is around 2.61 nm. When this value is evaluated with the graph compared to WT TkA given in figure 4.32.B, it is seen that D86S TkA has a higher radius of gyration value. WT TkA was lower with an average radius of gyration value of 2.61. This indicates that the D86S TkA enzyme is less compact than the WT TkA, implying that the structure exhibits a much tighter conformation after mutation (Hozoorbakhsh et al. 2023).

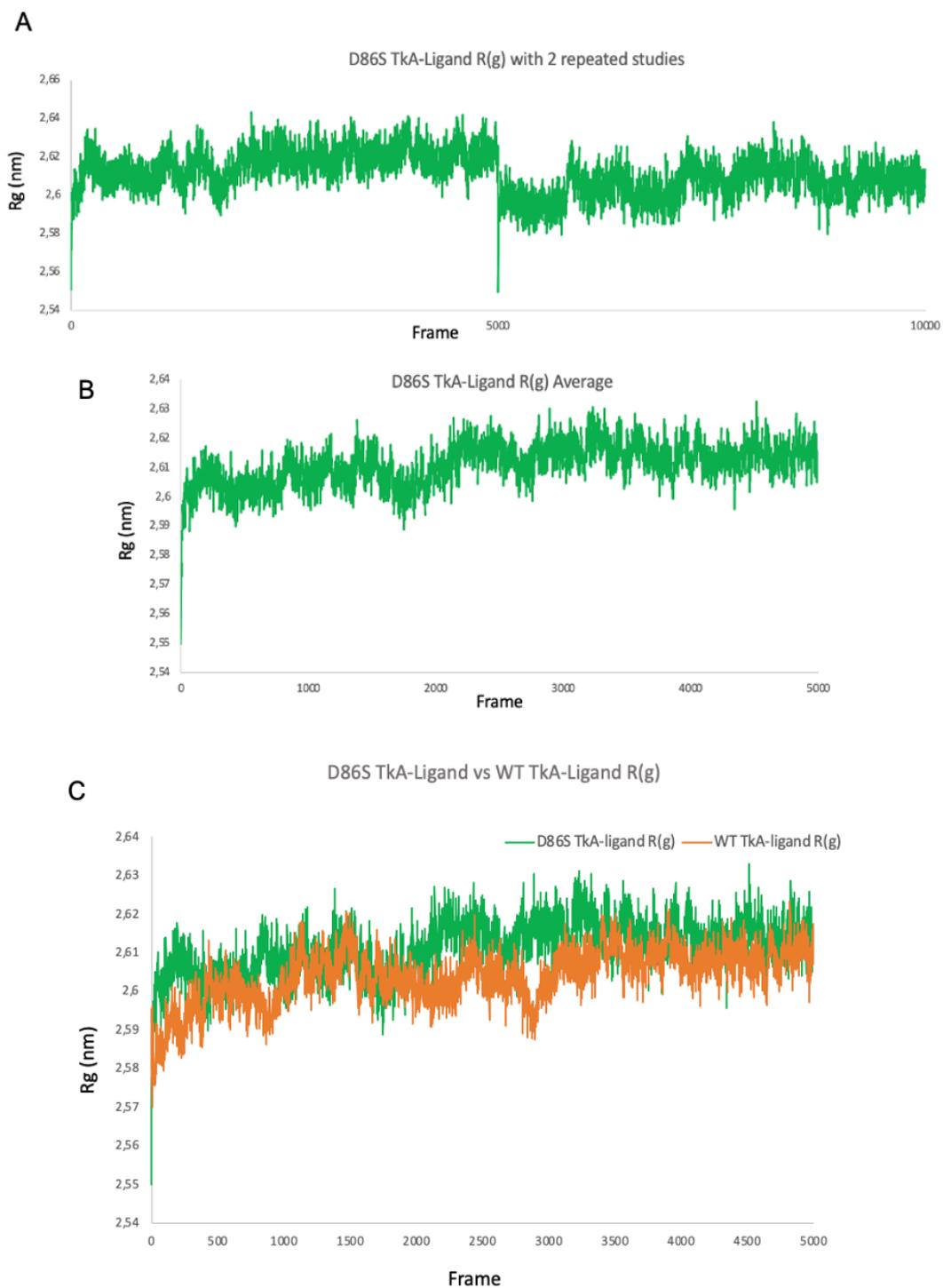


Figure 0.32. A) Radius of gyration plot for D86S TkA and its ligand l-asparagine with two repeats. A) Average radius of gyration plot for D86S TkA with l-asparagine ligand. C) Comparative radius of gyration for D86S TkA and WT TkA

The graph of the number of hydrogen bonds between the D86S TkA and the asparagine ligand is shown in figure 4.33. The graph shows the number of hydrogen bonds between pairs of atoms within 0.35 nm. The average of the hydrogen bond numbers obtained from two separate MD analyses is 3.1.

This value is lower compared to WT TkA (3.4). As with WT TkA, hydrogen bonding was observed at Ser86, Glu294 and an additional Lys156 site in D86S TkA. In addition, hydrogen bonding was observed in the Thr85 region in ligplot+ images (4.34). After MD simulation, it was observed that some H-bond formations decreased and disappeared as in the wild type. It was suggested that the change in hydrogen-bonding interactions resulted from the enzyme's stability-maintaining behavior (Nakagawa and Tamada, 2021).

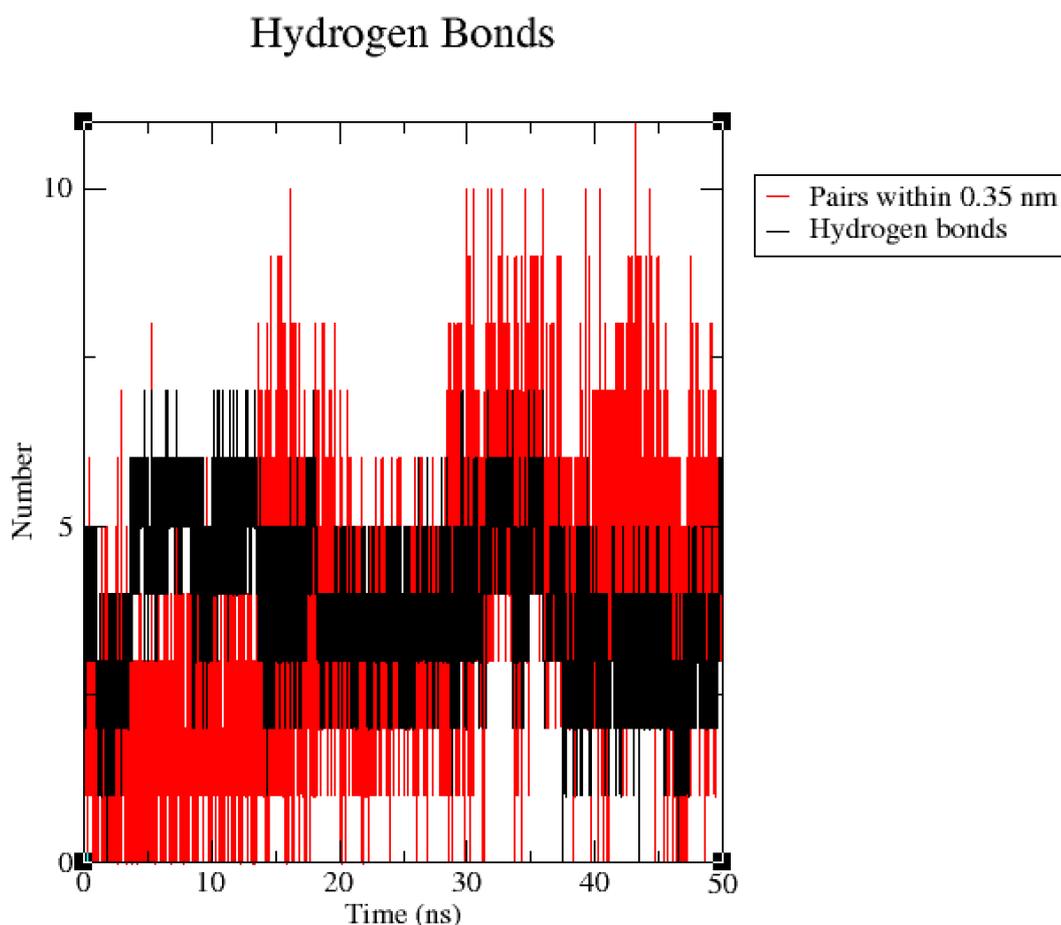


Figure 0.33. The number of hydrogen bonds between atom pairs within 0.35 nm distance.

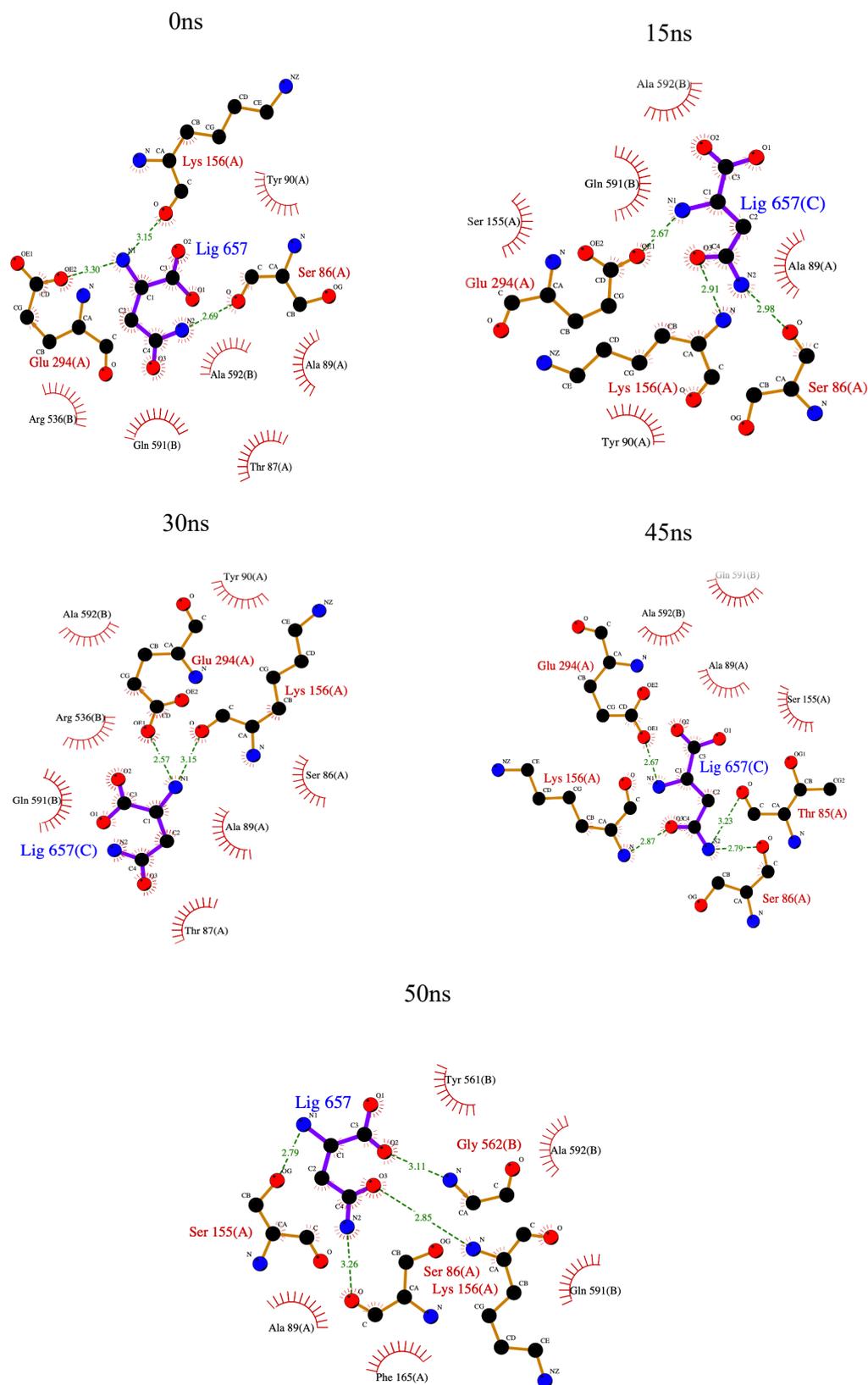


Figure 0.34. The images show hydrogen bond formation between l-asparagine ligand and some residues that are in active site of D85S TkA.

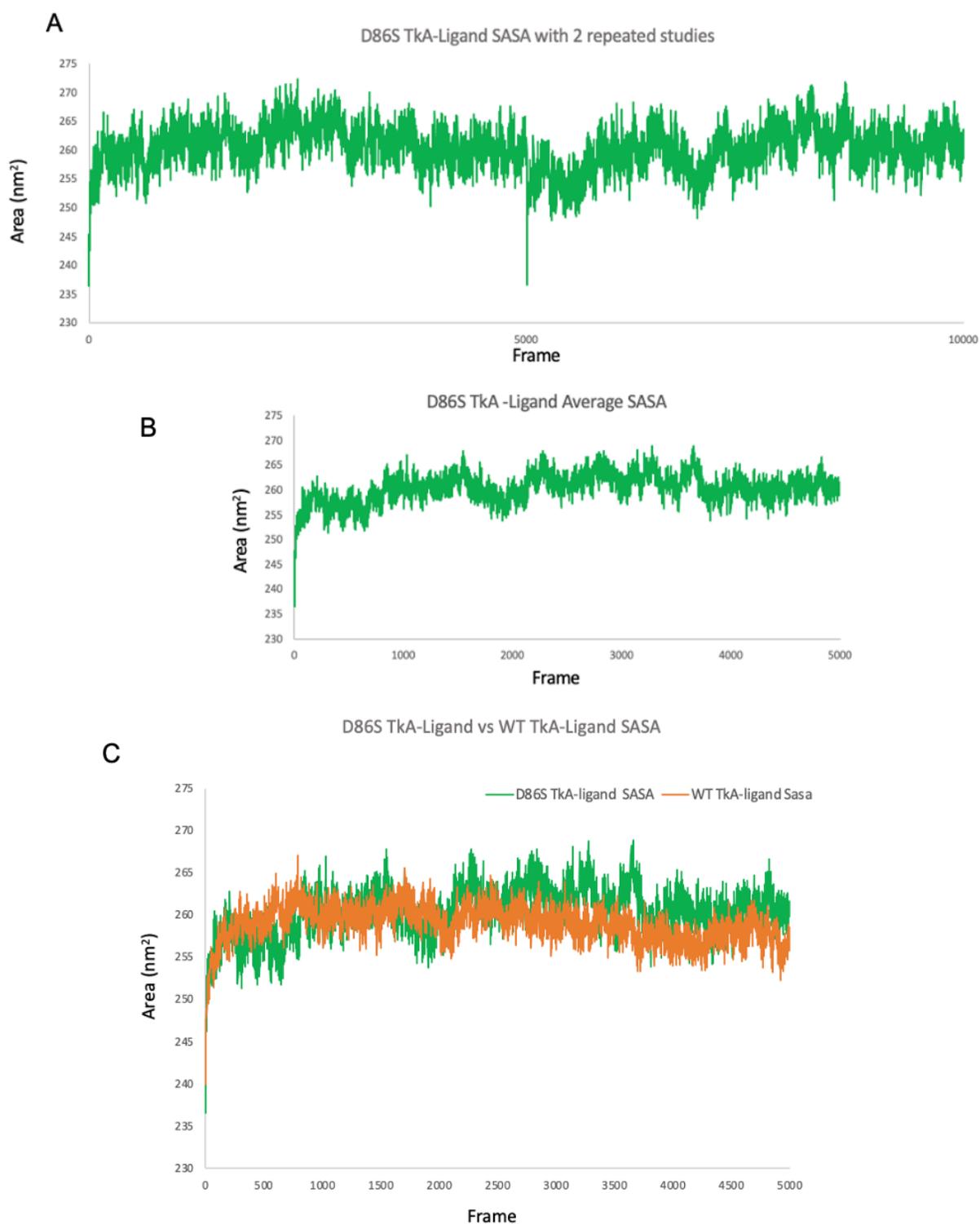


Figure 0.35. A) SASA plot for D86S TkA and asparagine ligand with two repetitions. B) Average SASA plot for D86S TkA and l-asparagine ligand. C) SASA plot for comparison of D86S TkA and WT TkA.

The SASA plot of D86S TkA derived for protein and ligand is shown in figure 4.35.A. Comparison of the SASA graph with the SASA graph of WT TkA is in figure 4.35.B. Total SASA value for D86S TkA was calculated as 260.39nm²/ns. For WT TkA, this value was 258.32 nm²/ns. SASA values for WT and D86S TkA enzymes are very close to each other. After 2500 frames, D86S TkA increased by a small margin, while WT TkA decreased. Based on this point, it can be said that the WT TkA enzyme exhibits less hydrophobic behavior with a lower SASA value (Chi et al. 2022; Jiao et al. 2022).

Coulomb-SR and LJ-SR interactions were examined for D86S TkA L-ASNase and its ligand asparagine in figure 4.36. The red line describes the LJ-SR van der Waals interactions and remained around -55.65 kJ/mol throughout the simulation. The black line indicates Coulomb electrostatic energy that is -307.37 kJ/mol. There is fluctuation on the plot that also represents an unbalanced energy state. The total energy value was evaluated in both studies and calculated as -363.02 kJ/mol. Total energy values of WT TkA and D86S TkA were compared. WT TkA had less total energy value, with a total energy value of -337.98 kJ/mol.

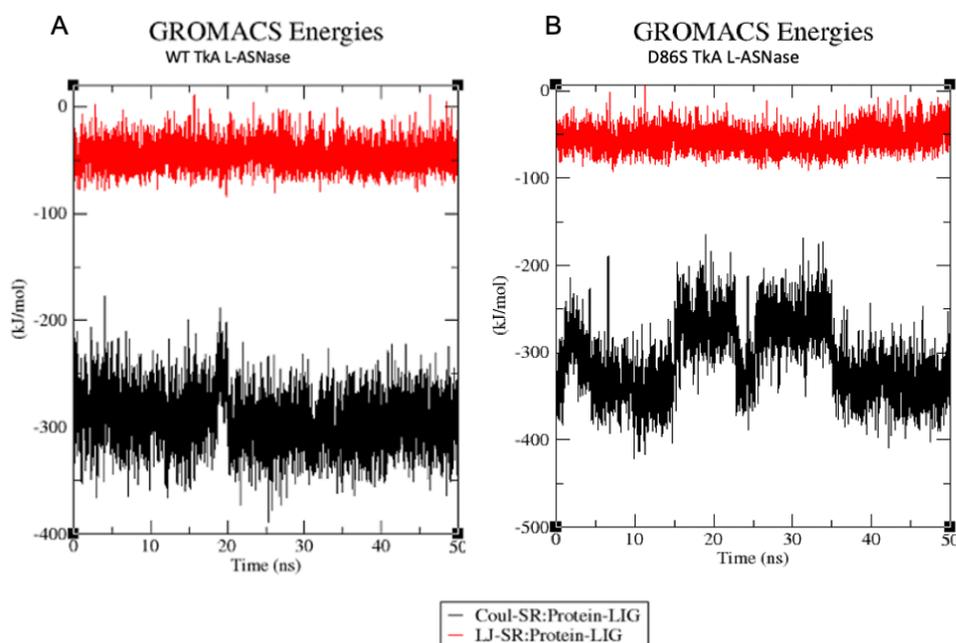


Figure 0.36. A) WT TkA and l-aspragine ligand interaction energy plot. B) D86S TkA and l-asparagine ligand interaction energy plot. Red regions represent LJ-SR and black regions represent Coul-SR energy.

A stronger van der Waals interaction is observed in the D86S TkA and ligand interaction energy, while a weaker electrostatic interaction dominates (Rampogu et al. 2022; Reddy et al. 2016; Erva et al. 2016; Londhe et al. 2019).

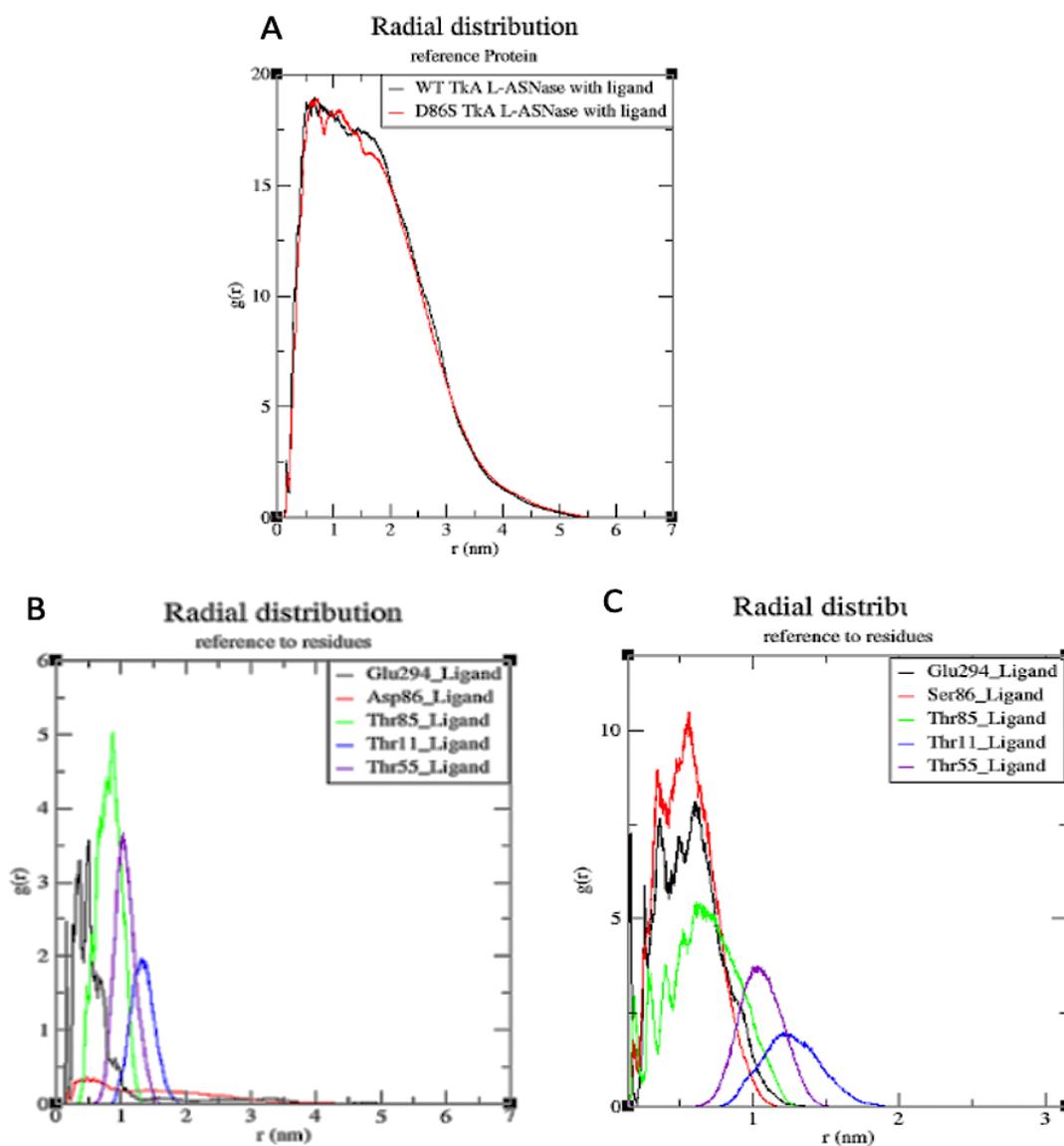


Figure 0.37. A) shows comparison of WT TkA RDF values and D86S TkA RDF values. B) RDF analysis is given specifically for residues for WT TkA. C) RDF analysis is given specifically for residues for D86S TkA.

The comparative RDF graph with the ligand as reference for WT TkA and D86S TkA is shown in figure 4.37. Considering the protein–ligand RDF plot as an overall comparison of the two simulation results, no significant differences were observed. For a more effective analysis, residues Thr11, Thr55, Thr85, Asp86 and Glu294 were analyzed for their activities in WT TkA and ligand interaction in figure 4.37.C.

As a result, after the mutation, the ligand was unstable towards the mutated residue and gave the highest peak, and this created a significant difference compared to WT TkA. Glu294 stand out with high peaks; But there is no stability. Thr85, Thr55 and Thr11 shows less interact with ligand, as well as decreased stability at Thr85 (Ahmad et al. 2017; Shamim, Abbasi, and Azam 2015).

The diffusion coefficient for the ligand from D86S TkA and ligand complex was calculated with MSD and Einstein equation. and the diffusion coefficient of the ligand was calculated as $10 \times 10^{-7} \text{cm}^2/\text{s}$ in this pair. This value, when compared to the ligand diffusion coefficient interacting with WT TkA, $3.3 \times 10^{-7} \text{cm}^2/\text{s}$, shows that the ligand of the mutant complex moves faster on the protein surface and exhibits more conformations, and expresses its ability to interact with the protein.

4.4.5. Molecular Dynamics Simulations for T11E TkA and Asparagine Ligand Complex

Finally, two repeat MD simulations were performed for T11E TkA and its ligand asparagine. Figure 4.38.A shows the backbone-to-backbone RMSD plot for the T11E TkA protein. Figure 4.38.B shows a graph comparing the RMSD values for WT TkA and T11E TkA. Generally, the graph of T11E TkA and ligand exhibited stable behavior. However, compared to WT TkA, the RMSD value of T11E TkA is seen to be less. In addition, both graphs show a stable behavior in conformational changes throughout the simulation (Ardalan et al., 2021; Chi et al., 2022; Jiao et al., 2022; Hozoorbakhsh et al., 2023).

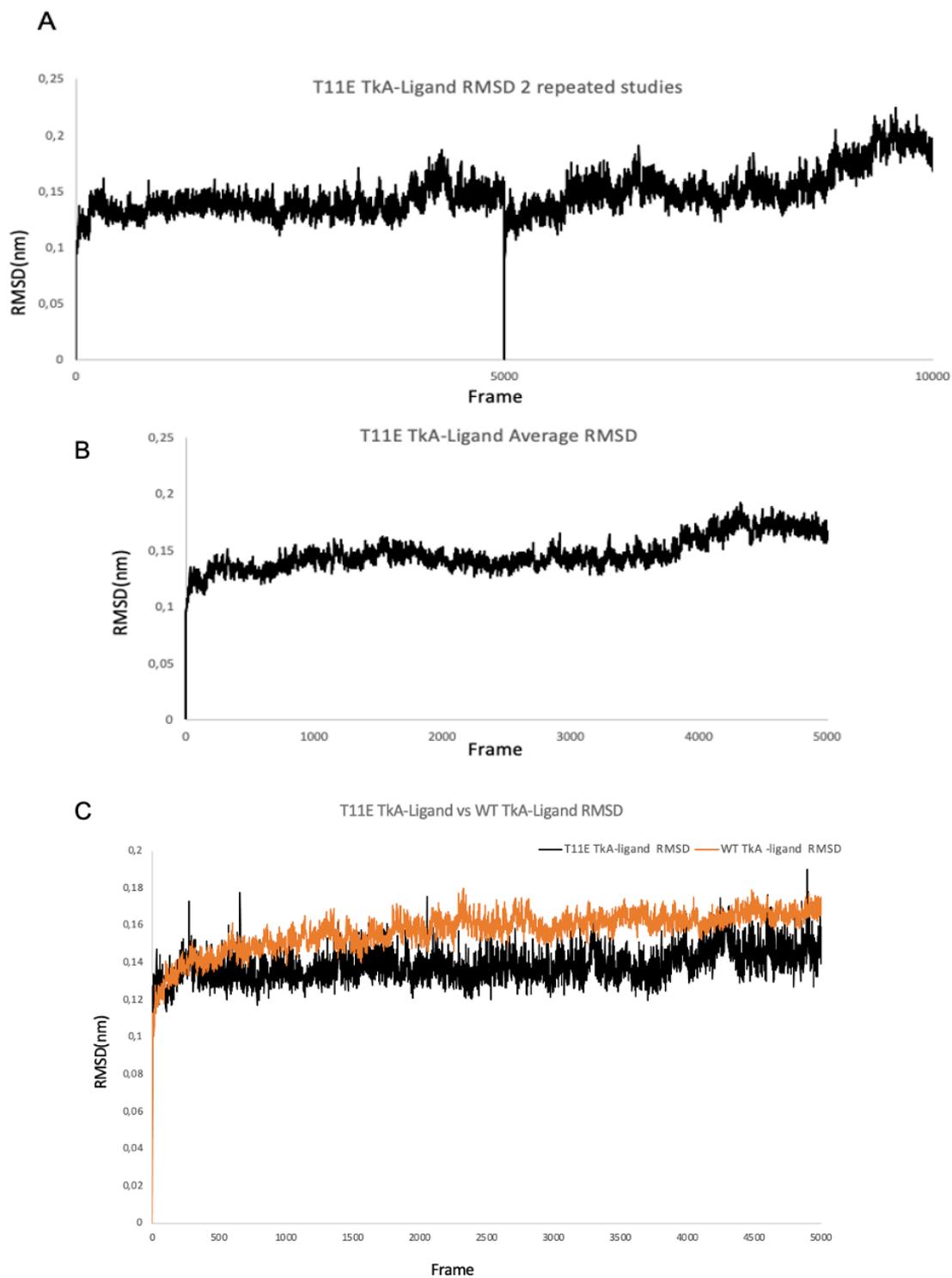


Figure 0.38. A) RMSD graphs of T11E TkA two replicates. B) T11E TkA average RMSD graph obtained from two repetitions. C) RMSD analysis; Comparison with WT TkA and T11E TkA with l-asparagine ligand.

RMSF graphs were obtained for T11E TkA and aspragine ligand. Figure 4.39.A contains two separate repetitions of the study and figure 4.39.B contains the comparative RMSF graph of WT TkA and T11E mutant TkA.

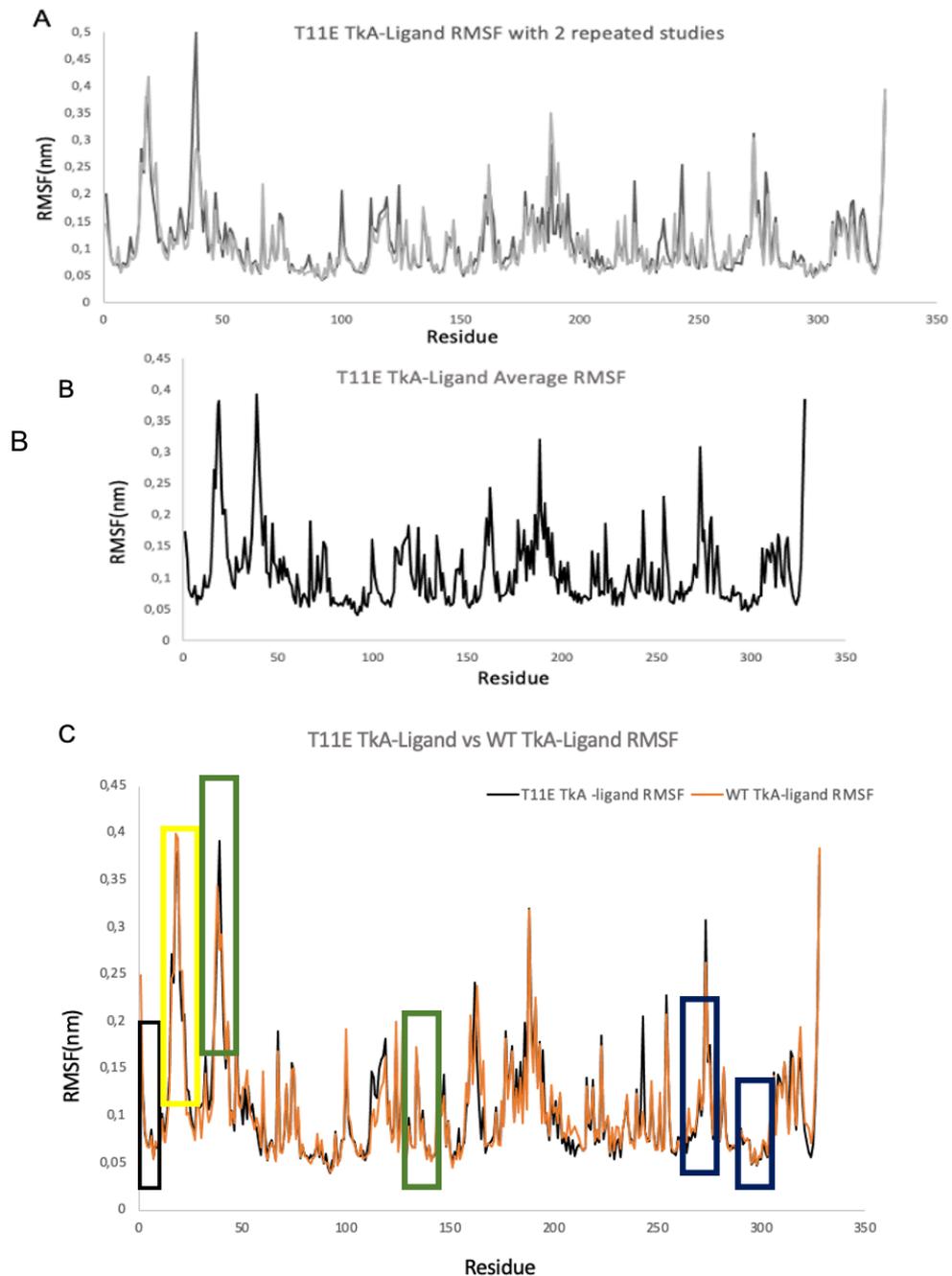


Figure 0.39. A) RMSF graphs of T11E TkA two replicates. B) T11E TkA average RMSF graph obtained from two repetitions. C) RMSF analysis; Comparison with WT TkA and T11E TkA.

RMSF plots showed fluctuations close to each other in enzymes before and after mutation. Residue: 120 -125 that is known as moved to active site in figure 4.40, T11E TkA exhibited higher peaks than WT TkA at positions 247-249 and 275. In general, these regions were in the outer region of the protein, and residues on the outer surface of the protein are expected to be more flexible than other regions. Also figure 4.39.C shows some parts of enzyme with the yellow box represents the hairpin: Residues: β 2(14-16) - β 3(20-22). Blue boxes belong to the residue region of the allosteric site: Residues: α 8(293-306) - β 15(258-262). Green regions α 1(27-33)– α 4(122-133) indicate residues from the region known to move towards the active site during ligand and enzyme interaction.

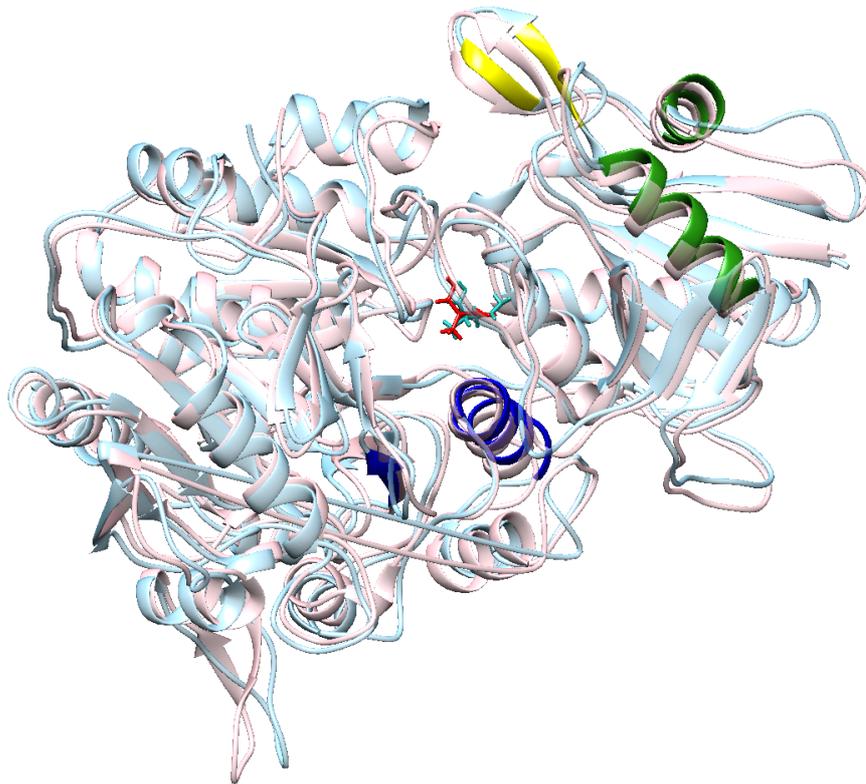


Figure 0.40. Initial position image of T11E TkA (light pink) and l-asparagine ligand (red) and image of T11E TkA (light blue) and l-asparagine ligand (cyan) at 10ns. The part shown in yellow here represents the hairpin: Residues: β 2(14-16) - β 3(20-22). Blue regions indicate the allosteric site: Residues: α 8(293-306) - β 15(258-262). Green regions α 1(27-33)– α 4(122-133) are regions that tend to move towards the active site during ligand and enzyme interaction.

It was thought that these fluctuations on the outer surface of the protein were greater for mutant TkA than for WT TkA. When the mutation site was examined, the RMSF value of Thr before the mutation was 0.0818, but after it was mutated with glutamic acid, this value became 0.1005, the average of both studies. In this residue indicates that it has become more rigid after mutation. It is indicated with a black box in figure 4.39.C. And the black box belongs to the T11E region. After the mutation, this region behaved more flexibly (Ardalan et al., 2021; Chi et al., 2022; Jiao et al., 2022; Hozoorbakhsh et al., 2023).

To evaluate the compactness of the protein after mutation, the radius of gyration graphs seen in figure 4.41 were evaluated. The two repeat studies were consistent with each other. Additionally, throughout the simulation, mutant TkA exhibited a radius of gyration profile that increased from 2.58 nm to 2.60 nm, maintaining its stability. As seen in figure 4.41.B, while mutant TkA compactness exhibited a stability close to WT TkA, it had lower levels of radius of gyration than WT TkA. This mutant TkA showed to have a more compact profile than WT TkA (Hozoorbakhsh et al., 2023).

The graph of the number of hydrogen bonds between the T11E TkA and the asparagine ligand is seen in figure 4.42. The hydrogen bonds formed between pairs of atoms at a distance of 0.35 nm were 3.116, as the average of both studies. This value was slightly lower than the WT TkA of 3.4. In addition, the hydrogen bonds obtained from LigPlot+ were in the aspartic acid, lysine, and glutamic acid regions, as in WT TkA and other mutants, in figure 4.43.

Hydrogen bonds indicate the stability of the interaction between the ligand and the enzyme, and since MD simulations are a dynamic formation, it was thought that the variation in bond formation was caused by the stability-maintaining behavior of the enzyme. Additionally, water molecules in the active site bound to the simulation may have caused competition in hydrogen bond formation (Jiao et al., 2022).

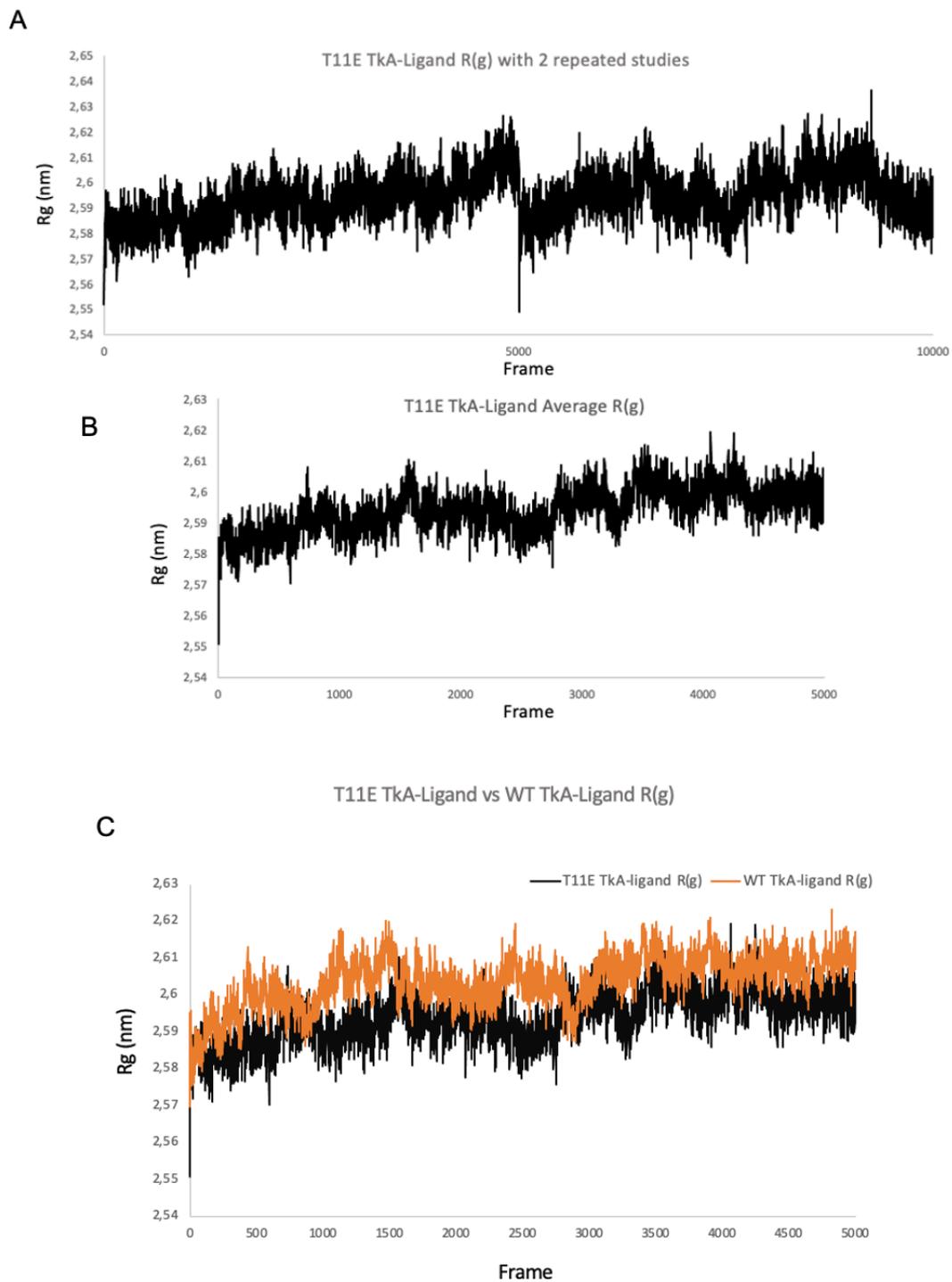


Figure 0.41. A) Radius of gyration plots for T11E TkA and ligand l-asparagine with two repeat studies. B) Average RMSF plot for T11E TkA and l-asparagine ligand. B) Comparison of radius of gyration WT TkA and T11E TkA.

Hydrogen Bonds

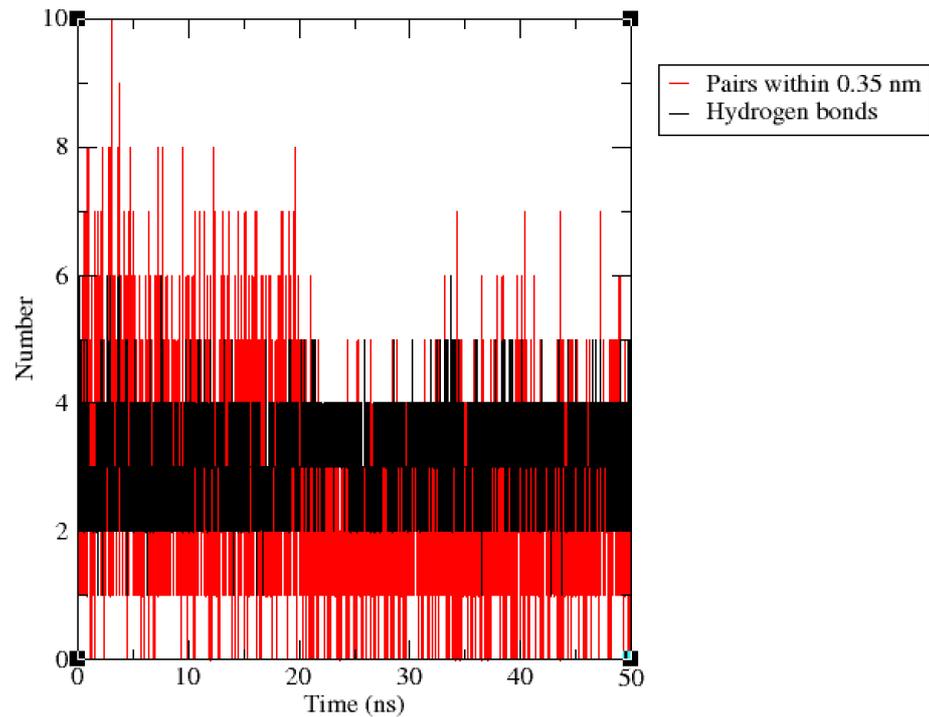


Figure 0.42. The number of hydrogen bonds between atom pairs within 0.35 nm distance.

The SASA graph derived for the protein and ligand obtained from the molecular dynamic simulation of T11E TkA is shown in figure 4.44.A. When the SASA graph was compared with the SASA graph of WT TkA, the graphs were close to each other in figure 4.44.B. The SASA value for T11E TkA was calculated as $259.75\text{nm}^2/\text{ns}$. SASA value of WT TkA was calculated as $258.92\text{nm}^2/\text{ns}$. However, after 2500 frames, WT TkA experienced a minor decrease, while mutant TkA remained stable.

This increase in SASA indicates that especially the hydrophilic regions of the molecule interact more with the solvent. It is understood that the hydrophilic parts of the enzyme change similarly to WT TkA during its interaction with the ligand and remain constant throughout the simulation (Chi et al. 2022; Jiao et al. 2022).

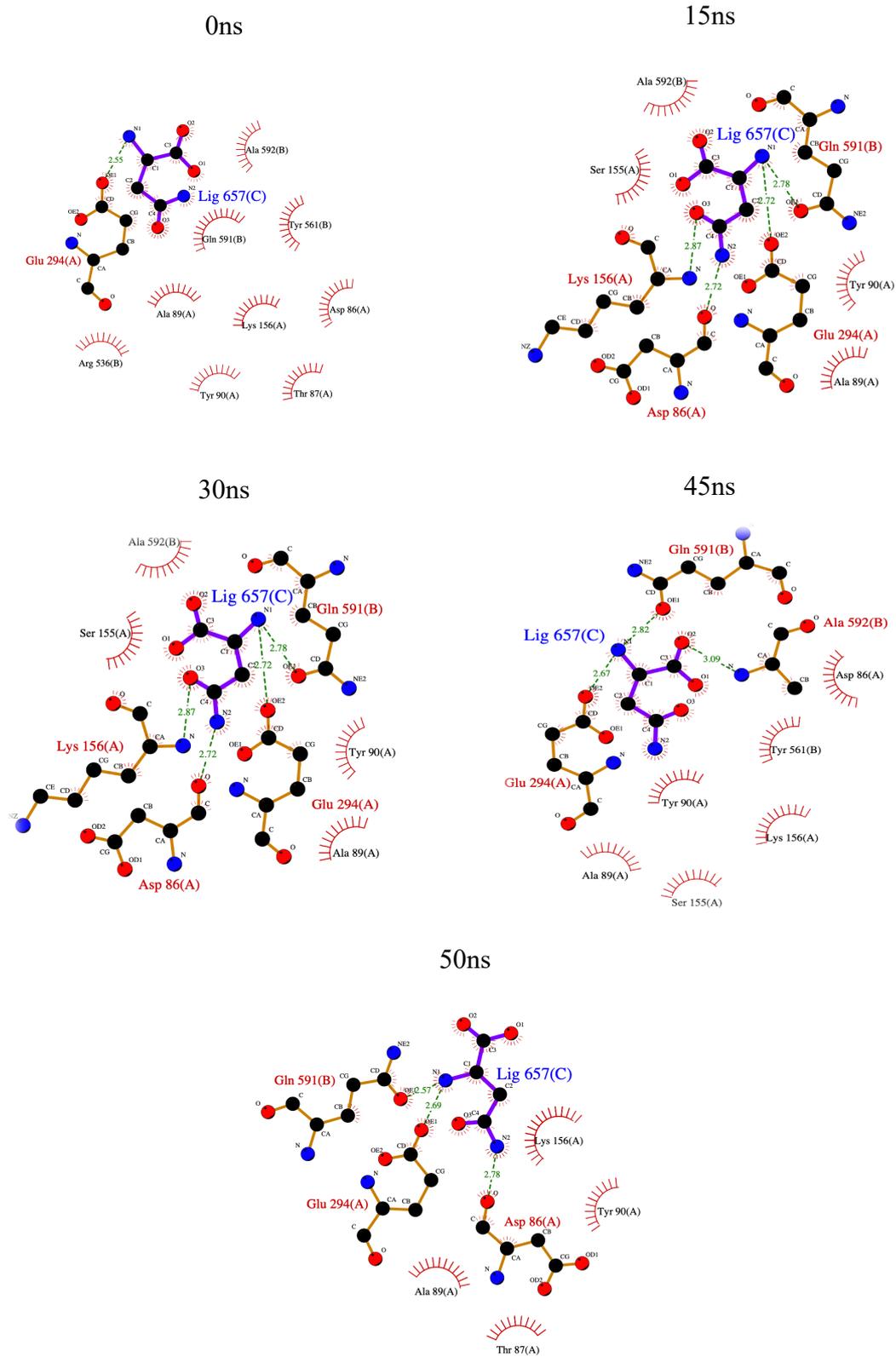


Figure 0.43. The images show hydrogen bond formation between l-asparagine ligand and some residues that are in active site of T11E TkA.

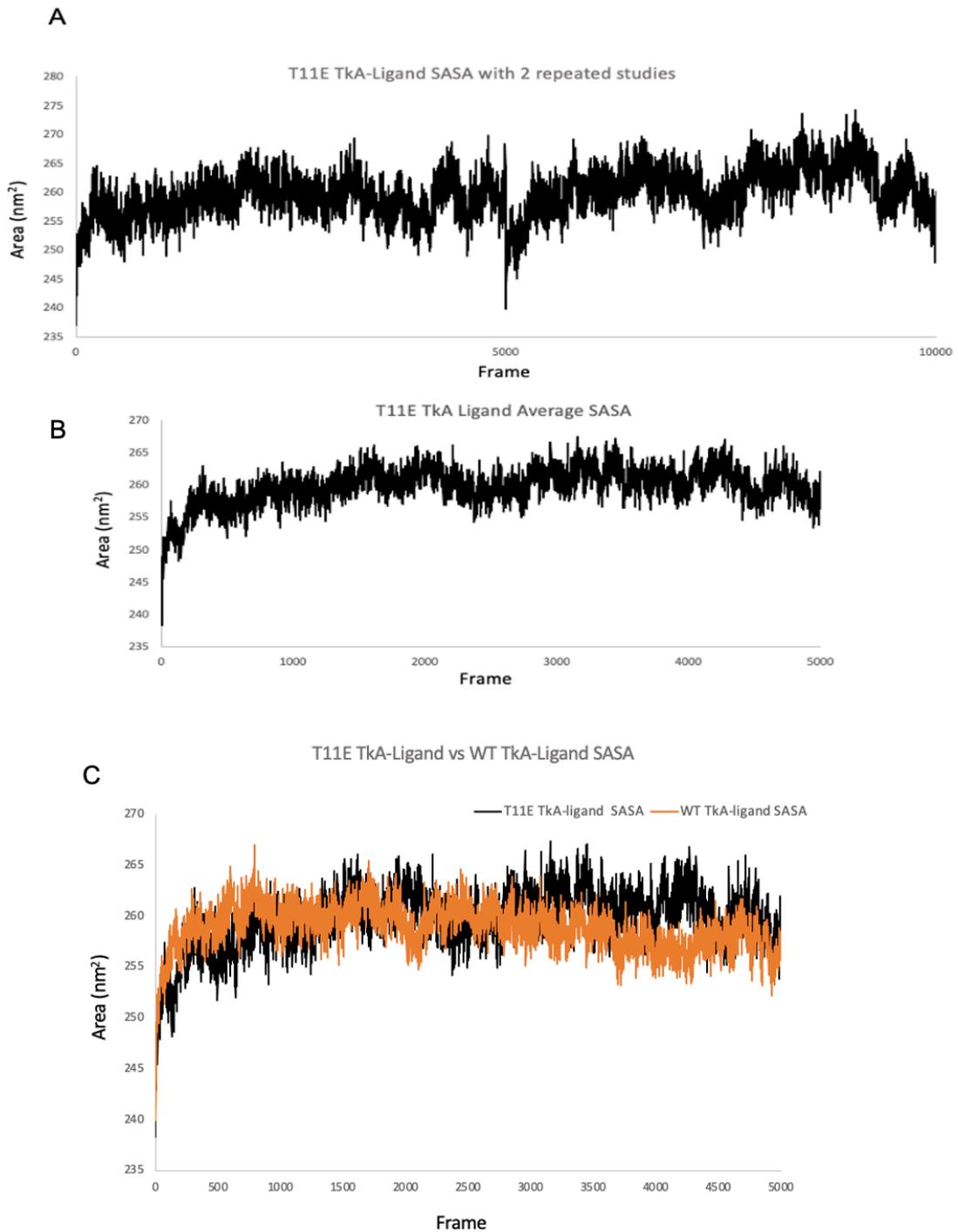


Figure 0.44. A) SASA plots of T11E TkA and its ligand l-asparagine is shown with two repeats. B) Average SASA plots of T11E TkA and ligand l-asparagine. C) WT TkA with l-asparagine and T11E TkA with l-asparagine ligand comparison via SASA plots.

Figure 4.45 was evaluated to evaluate the interaction energy between the enzyme and the ligand. In the comparison of WT TkA and T11E mutant TkA, van der Waals and electrostatic energy values were very close to each other. However, when the average values were examined, small deviations in the T11E resulted in the total energy value of mutant TkA being lower than WT TkA with -337.98kJ/mol. The total energy of T11E TkA value was calculated as an average of -314.88 kJ/mol after being obtained from both repeat studies. This value includes both van der Waals with -43.52 kJ/mol and electrostatic energy -271.36 kJ/mol. Total energy values obtained from literature studies show similar profiles for EcA and *Erwinaze chrysantemi* asparaginase (Rampogu et al. 2022; Reddy et al. 2016; Erva et al. 2016; Londhe et al. 2019).

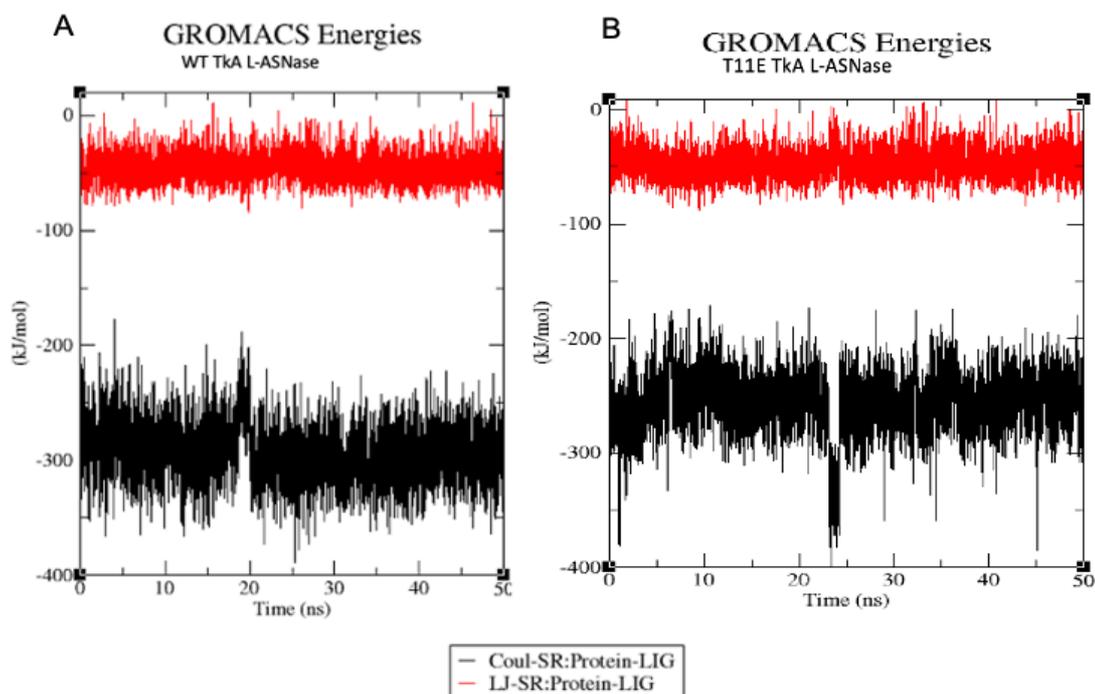


Figure 0.45. A) WT TkA and l-asparagine ligand interaction energy plot. B) T11E TkA and l-asparagine ligand interaction energy plot. Red regions represent LJ-SR and black regions represent Coul-SR energy.

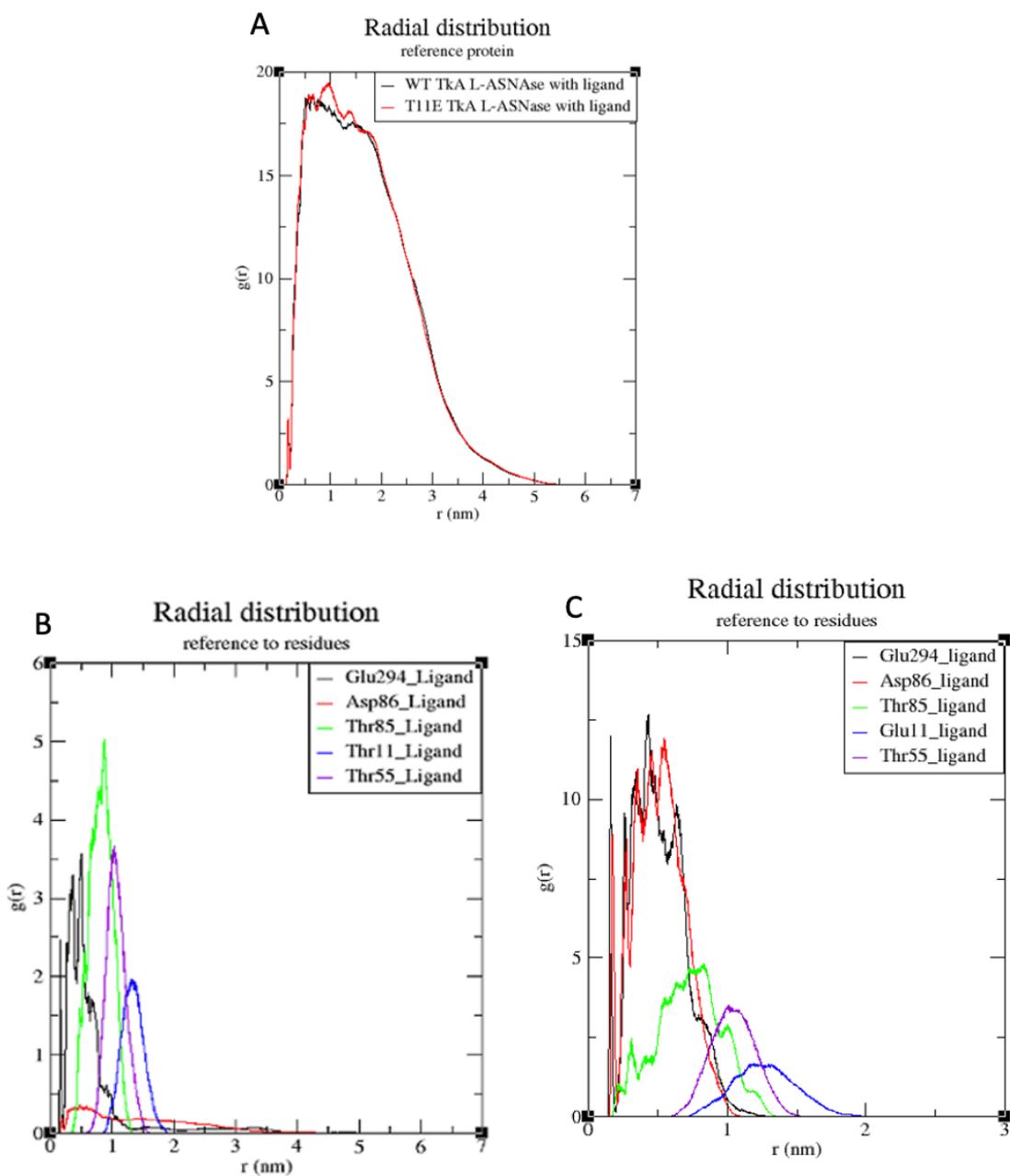


Figure 0.46. A) RDF analysis between WT TkA's ligand and its active site as well as T11E TkA's ligand and its active site. B) shows RDF analysis between some important residue on the active site and WT TkA' ligand. C) shows RDF analysis between some important residue on the active site and T11E TkA' ligand.

As seen in the RDF general evaluation figure 4.46.A, peaks with similar heights were obtained between WT TkA and mutant TkA. When similar regions in both graphs are examined in detail, in figure 4.46.B-C, the mutation region gave a lower peak as a

result of MD simulations, indicating that after the mutation, the probability of the ligand being found in the T11E region decreases. In addition, the clear peak and smooth image were replaced by an unclear peak, showing that this region had unstable interactions after the mutation.

On the other hand, Glu294 and Asp86, which contribute to hydrogen bond formation, appear to show a higher peak but more irregular interaction than WT TkA. In addition to Thr11, Thr85, which is thought to initiate another nucleophilic attack, again gave an irregular and lower peak after mutation (Ahmad et al. 2017; Shamim, Abbasi, and Azam 2015).

The Diffusion coefficient for the ligand from T11E mutant TkA and ligand complex was calculated with MSD and Einstein equation. and the diffusion coefficient of the ligand was calculated as $1 \times 10^{-7} \text{cm}^2/\text{s}$ in this pair. This value, when compared to the ligand diffusion coefficient interacting with WT TkA, i.e. $3.3 \times 10^{-7} \text{cm}^2/\text{s}$, shows that the ligand of the mutant complex moves slower on the protein surface and exhibits fewer conformations, and expresses its ability to interact with the protein.

4.4.6. Comparison of All Mutants and WT TkA

Overall comparisons shown in table 4.6, when comparing the WT TkA and the T55E TkA in terms of stability and dynamic balances, the mutants showed different stabilities. Although the RMSD values evaluated in terms of protein-ligand interaction was more stable than the backbone-to-backbone RMSD values of the protein, the WT TkA generally showed better results in terms of stability.

The differences in the radius of gyration values show that the T55E TkA is more compact and tighter than the wild type, which is thought to be an unstable region for the ligand in the protein-ligand interaction region. While increasing SASA values support the radius of gyration results, the fact that the T55E TkA interacts less with the solvent and becomes more conformationally tight may indicate specific interactions with the ligand. This also points to the more hydrophilic regions of the wild type. The slight decrease in the average H-bond number (0.3) and the decreasing values in the protein-ligand

interaction energy analysis were thought to indicate that the T55E TkA would have a weaker interaction than the WT TkA.

Table 4.6. Overall Results from MD studies and comparisons. According to the RDF value, the change in the mutation region is indicated as increase or decrease. The NA value corresponds to the values for which the data in the table is not available.

Enzyme	Radius of Gyration (nm/ns)	H-bond	SASA (nm ² /ns)	Total Energy (kJ/mol)	Difussion Coefficients ($\times 10^{-7}$ cm ² /s)	RDF
Apo WT TkA	2.59	NA	260.38	NA	NA	NA
WT TkA with L-asparagine	2.60	3.4	258.92	-337.98	3.3	NA
T55E TkA with L-asparagine	2.61	3.21	258.40	-246.02	5	Slightly decrease
D86S TkA with L-asparagine	2.61	3.1	260.38	-363.02	10	Highly increased
T11E TkA with L-asparagine	2.59	3.116	259.75	-314.88	1	Highly decreased

Guo et al. as stated, when specific RDF analyses are evaluated considering all analyses of Thr11 and Thr85 nucleophilic attack initiation and T55E TkA, it is noteworthy that the probability of interaction in the T55E mutation region and Glu294 and ASp86 regions increased, but their increase was not stable. The value at Thr11 and

Thr85 was not stable. The decrease supported the mutant structure being in a less stable conformation with the ligand.

The D86S TkA showed similar RMSD profile as WT TkA, and showed the similarity of stability in conformational change between these two enzymes. The higher peaks in RMSF values of the D86S mutant showed that it was more flexible than WT TkA. This flexible state of the mutant can also be associated with its higher radius of gyration value, which represents the compact state. It appears that the D86S TkA, which has a lower total SASA than WT TkA, has a more hydrophobic profile. The fact that the mutant region gives a very high peak in the RDF analysis but draws a less stable profile suggests that the effectiveness of the interaction in the mutation region has decreased or the hydrogen bonds formed in this region may be broken more easily. Additionally, the D86S mutant showed a higher total interaction energy profile than WT TkA and other mutants.

Finally, the RMSD profile of the T11E mutant TkA remained at a lower value than that of WT TkA, and when the RMSF values were compared, it was seen that these two enzymes gave results closer to each other than the other mutants. The mutant, which was more compact than WT TkA by radius of gyration analysis, did not show any difference in the hydrogen bond number analysis compared to other mutants. Again, in SASA analysis, WT TkA followed a similar profile as other mutations, but in the last 2500 frames, WT TkA, unlike WT TkA, remained stable. Total energy value was lower than WT TkA and showed instability at some points. RDF analysis resulted in higher peaks in regions known to form hydrogen bonds after mutation, while lower and irregular peaks for both residues that initiated the nucleophilic attack.

CHAPTER 5

CONCLUSION

In the thesis, molecule docking studies were carried out on WT *Thermococcus kodakarensis* L-asparaginase with L-asparagine ligand. Then, molecular docking studies were carried out for mutation experiments, and mutation sites for TkA were determined according to the best binding scores obtained. These were T55E, D86S and T11E mutants. Then, molecular dynamics simulation studies were performed and RMSD, RMSF, radius of gyration values were expected, and then H-bond and protein-ligand interaction energy analyzes were performed. In addition, RDF analyzes were performed based on the probability calculation of the ligand's tendency towards residues in the active site, and the diffusion coefficient of the ligand was calculated for each enzyme. Additionally, active site residues were identified by literature review and sequence analysis. To our knowledge, such detailed structural analysis has not yet been performed with TkA and its asparagine ligand. In our docking studies, docking results obtained from *E. coli* L-ASNase studies were taken as reference. However, it was seen in WT TkA and mutants that the L-asparagine ligand interacted with the enzyme at a point slightly farther from the active site residues, contrary to what was reported in the *E. coli* L-ASNase enzyme. This situation, previously stated by Guo et al. in their study on the TkA enzyme and L-asparagine ligand, is thought to be due to the unexpected electron densities created by the L-asparagine ligand when it interacts with the L-asparagine ligand in the active site of the TkA enzyme. The stability and dynamic properties of WT TkA and its mutants were investigated in aqueous environment with and without ligand. WT TkA gave compatible results when compared to the literature and showed that it mostly maintained its balanced structure.

As a result of the MD analysis performed comparatively in every respect with the WT TkA enzyme, no significant differences were observed in the mutations tested. However, it was thought that these mutations caused some minor changes in bond formation and interaction with the ligand. Based on docking analysis, the three mutant structures were selected: T55E, D86S, and T11E. However, since the asparagine ligand

is a small molecule, docking studies did not yield the desired results, and close values were obtained for mutants and WT TkA. Overall, this study provided a better understanding of the structure of the TkA protein and more information about its dynamics. Although the mutations did not have radical effects on the protein structure, minor changes were observed in the protein and ligand interaction processes. We investigated residues known to initiate nucleophilic attack in the active site, such as Thr 11 and Thr85, with RDF analysis. A disruption of the steady state of the TkA enzyme was observed due to the T11E mutation, but overall, the enzyme had a profile that was not affected by the mutations. This showed that Thr11 and Thr85 mostly retained the expected differential stability.

When all the results were taken together, D86S TkA enzyme showed higher hydrophilic properties with a SASA value of 260.38 nm²/ns higher than WT TkA, a SASA value of 258.92 nm²/ns. As a result of radius of gyration comparisons, the high value of the D86S mutant, albeit at minor levels, were in agreement with SASA analysis. Among WT TkA and mutant enzymes, the highest interaction energy value belonged to D86S with -363.02 kJ/mol. Subsequently, the most accelerated movement of the ligand was also observed in the D86S mutant, with a diffusion coefficient of $10 \times 10^{-7} \text{ cm}^2/\text{s}$.

In future work, docking studies can be performed with structures obtained from MD simulations for WT TkA and its ligand asparagine. To see clearer differences between mutated structures and wild-type TkA, we should focus on double-site mutations instead of single point mutations. It is also recommended to conduct detailed studies on the allosteric region.

REFERENCES

- Aghaiypour, Khosrow, Alexander Wlodawer, and Jacek Lubkowski. 2001. "Structural Basis for the Activity and Substrate Specificity of *Erwinia Chrysanthemi* L-Asparaginase." *Biochemistry* 40 (19): 5655–64.
- Ahmad, Sajjad, Saad Raza, Reaz Uddin, and Syed Sikander Azam. 2017. "Binding Mode Analysis, Dynamic Simulation and Binding Free Energy Calculations of the MurF Ligase from *Acinetobacter Baumannii*." *Journal of Molecular Graphics and Modelling* 77:72–85.
- Alberts, B, Johnson A, and Lewis J. 2002. "Analyzing Protein Structure and Function." *Molecular Biology of the Cell*. Pages
- Ambrose, Jenifermallavarpu, Malathi Kullappan, Shankargouda Patil, Khalidj Alzahrani, Hamsajameel Banjer, Fadis I. Qashqari, A. Thirumal Raj, et al. 2022. "Plant-Derived Antiviral Compounds as Potential Entry Inhibitors against Spike Protein of SARS-CoV-2 Wild-Type and Delta Variant: An Integrative In Silico Approach." *Molecules* 27 (6).
- Ardalan, Noeman, Abbas Akhavan Sepahi, and Ramazan Ali Khavari-Nejad. 2021. "Development of *Escherichia Coli* Asparaginase II for the Treatment of Acute Lymphocytic Leukemia: In Silico Reduction of Asparaginase II Side Effects by A Novel Mutant (V27F)." *Asian Pacific Journal of Cancer Prevention* 22 (4): 1137–47.
- Baral, Adesh, Ritesh Gorkhali, Amit Basnet, Shubham Koirala, Btech ; Hitesh, Kumar Bhattarai, and Hitesh Kumar Bhattarai. 2021. "Selection of the Optimal L-Asparaginase II Against Acute Lymphoblastic Leukemia: An In Silico Approach." *JMIRx Med* 2(3): e298442.
- Bepari, Asim Kumar, and Hasan Mahmud Reza. 2021. "Identification of a Novel Inhibitor of SARS-CoV-2 3CL-PRO through Virtual Screening and Molecular Dynamics Simulation." *PeerJ* 9; e11261.
- Bissantz, C., G. Folkers, and D. Rognan. 2000. "Protein-Based Virtual Screening of Chemical Databases. 1. Evaluation of Different Docking/Scoring Combinations." *Journal of Medicinal Chemistry* 43 (25): 4759–67.
- Chen, Kuan Chung, Kuen Bao Chen, Hsin Yi Chen, and Calvin Yu Chian Chen. 2014. "In Silico Investigation of Potential Pyruvate Kinase M2 Regulators from

Traditional Chinese Medicine against Cancers.” *BioMed Research International* 189495.

Chen, Xianfeng, Irene Weber, and Robert W. Harrison. 2008. “Hydration Water and Bulk Water in Proteins Have Distinct Properties in Radial Distributions Calculated from 105 Atomic Resolution Crystal Structures.” *Journal of Physical Chemistry B* 112 (38): 12073–80.

Chi, Huibing, Yilian Wang, Bingjie Xia, Yawen Zhou, Zhaoxin Lu, Fengxia Lu, and Ping Zhu. 2022. “Enhanced Thermostability and Molecular Insights for L-Asparaginase from *Bacillus Licheniformis* via Structure and Computation-Based Rational Design.” *Journal of Agricultural and Food Chemistry* 70 (45): 14499–509.

Covino, Clelia, Angela Sorrentino, Prospero Di Pierro, Alessandra Aiello, Raffaele Romano, and Paolo Masi. 2023. “Asparaginase Enzyme Reduces Acrylamide Levels in Fried and Wood Oven Baked Pizza Base.” *Food Chemistry Advances* 2.

Derst, Christian, Johannes Henseling, and Klaus-Heinrich Röhm. 2000. “Engineering the Substrate Specificity of *Escherichia Coli* Asparaginase II. Selective Reduction of Glutaminase Activity by Amino Acid Replacements at Position 248.” *Protein science : a publication of the Protein Society* 9(10), 2009–2017.

Dhavalala, Prathusha, and Anastassios C. Papageorgiou. 2009. “Structure of *Helicobacter Pylori* L-Asparaginase at 1.4 Å Resolution.” *Acta Crystallographica Section D: Biological Crystallography* 65 (12): 1253–61.

Du, Juan, Jianjun Dong, Songjie Du, Kun Zhang, Junhong Yu, Shumin Hu, and Hua Yin. 2020. “Understanding Thermostability Factors of Barley Limit Dextrinase by Molecular Dynamics Simulations.” *Frontiers in Molecular Biosciences* 7.

Durham, Elizabeth, Brent Dorr, Nils Woetzel, René Staritzbichler, and Jens Meiler. 2009. “Solvent Accessible Surface Area Approximations for Rapid and Accurate Protein Structure Prediction.” *Journal of Molecular Modeling* 15 (9): 1093–1108.

Egler, Rachel A., Sanjay P. Ahuja, and Yousif Matloub. 2016. “L-Asparaginase in the Treatment of Patients with Acute Lymphoblastic Leukemia.” *Journal of Pharmacology and Pharmacotherapeutics*. Medknow Publications 7(2): 62–71.

Erva, Rajeswara Reddy, Satish Babu Rajulapati, Chandrasai Potla Durthi, Mayuri Bhatia, and Madhuri Pola. 2016. “Molecular Dynamic Simulations of *Escherichia Coli* L-Asparaginase to Illuminate Its Role in Deamination of Asparagine and Glutamine Residues.” *3 Biotech* 6 (1): 1–7.

- Fresch, Elisa, and Elisabetta Collini. 2023. “The Role of H-Bonds in the Excited-State Properties of Multichromophoric Systems: Static and Dynamic Aspects.” *Molecules*. MDPI 28(8):3553.
- González-Torres, Iván, Ernesto Perez-Rueda, Zahaed Evangelista-Martínez, Andrés Zárate-Romero, Angélica Moreno-Enríquez, and Alejandro Huerta-Saquero. 2020. “Identification of L-Asparaginases from *Streptomyces* Strains with Competitive Activity and Immunogenic Profiles: A Bioinformatic Approach.” *PeerJ* 8.
- Guo, Jingxu, Alun R. Coker, Steve P. Wood, Jonathan B. Cooper, Shahid Mahmood Chohan, Naeem Rashid, and Muhummad Akhtar. 2017. “Structure and Function of the Thermostable L-Asparaginase from *Thermococcus Kodakarensis*.” *Acta Crystallographica Section D: Structural Biology* 73 (11): 889–95.
- Hollingsworth, Scott A., and Ron O. Dror. 2018. “Molecular Dynamics Simulation for All.” *Neuron*. Cell Press 99(6), 1129–1143.
- Hospital, Adam, Josep Ramon Goñi, Modesto Orozco, and Josep L. Gelpí. 2015. “Molecular Dynamics Simulations: Advances and Applications.” *Advances and Applications in Bioinformatics and Chemistry*. Dove Medical Press Ltd 8 Pages 37–47.
- Hozoorbakhsh, Fereshteh, Mozhgan Ghiasian, Fereshte Ghandehari, Zarrindokht Emami-Karvani, and Maryam Khademi Dehkordi. 2023. “An Immunoinformatic Approach Employing Molecular Docking and Molecular Dynamics Simulation for Evaluation of L-Asparaginase Produced by *Bacillus Velezensis*.” *Journal of Biomolecular Structure and Dynamics* 41 (18): 9057–71.
- Jeyaram, R. A., and C. Anu Radha. 2022. “Investigation on the Binding Properties of N1 Neuraminidase of H5N1 Influenza Virus in Complex with Fluorinated Sialic Acid Analog Compounds—a Study by Molecular Docking and Molecular Dynamics Simulations.” *Brazilian Journal of Physics* 52 (1).
- Jia, Ruiying, Xiao Wan, Xu Geng, Deming Xue, Zhenxing Xie, and Chaoran Chen. 2021. “Microbial L-Asparaginase for Application in Acrylamide Mitigation from Food: Current Research Status and Future Perspectives.” *Microorganisms*. MDPI AG 3;9(8):1659.
- Jiao, Linshu, Huibing Chi, Bingjie Xia, Zhaoxin Lu, Xiaomei Bie, Haizhen Zhao, Fengxia Lu, and Meirong Chen. 2022. “Thermostability Improvement of L-Asparaginase from *Acinetobacter Soli* via Consensus-Designed Cysteine Residue Substitution.” *Molecules* 27 (19).

- Kamisaki, Y, H Wada, T Yagura, A Matsushima, and Y Inada. 1981. "Reduction in Immunogenicity and Clearance Rate of Escherichia Coli L-Asparaginase by Modification with Monomethoxypolyethylene Glycol." *Journal of Pharmacology and Experimental Therapeutics* 216 (2): 410.
- Karataş, D., A. Tekin, F. Bahadori, and M. S. Çelik. 2017. "Interaction of Curcumin in a Drug Delivery System Including a Composite with Poly(Lactic-: Co -Glycolic Acid) and Montmorillonite: A Density Functional Theory and Molecular Dynamics Study." *Journal of Materials Chemistry B* 5 (40): 8070–82.
- Kumer, Ajoy, Unesco Chakma, and Sarkar Mohammad Abe Kawsar. 2021. "The Inhibitory Effect of Some Natural Bioactive D-Glucopyranoside Derivatives against SARS-CoV-2 Main Protease (Mpro) and Spike Protease (Spro)." *ASM Science Journal* 16: 1–18.
- Lemkul, Justin. 2019. "From Proteins to Perturbed Hamiltonians: A Suite of Tutorials for the GROMACS-2018 Molecular Simulation Package [Article v1.0]." *Living Journal of Computational Molecular Science* 1 (1).
- Londhe, Ashwini Mac Hhindra, Changdev Gorakshnath Gadhe, Sang Min Lim, and Ae Nim Pae. 2019. "Investigation of Molecular Details of Keap1-Nrf2 Inhibitors Using Molecular Dynamics and Umbrella Sampling Techniques." *Molecules* 24 (22).
- Lubkowski, Jacek, Juan Vanegas, Wai Kin Chan, Philip L. Lorenzi, John N. Weinstein, Sergei Sukharev, David Fushman, Susan Rempe, Andriy Anishkin, and Alexander Wlodawer. 2020. "Mechanism of Catalysis by l -Asparaginase." *Biochemistry* 59 (20): 1927–45.
- Ma, Youguang, Chunying Zhu, Peisheng Ma, and K. T. Yu. 2005. "Studies on the Diffusion Coefficients of Amino Acids in Aqueous Solutions." *Journal of Chemical and Engineering Data* 50 (4): 1192–96.
- Maggi, Maristella, Steven D. Mittelman, Jean Hugues Parmentier, Giorgio Colombo, Massimiliano Meli, Jeannette Marie Whitmire, D. Scott Merrell, Julian Whitelegge, and Claudia Scotti. 2017. "A Protease-Resistant Escherichia Coli Asparaginase with Outstanding Stability and Enhanced Anti-Leukaemic Activity in Vitro." *Scientific Reports* 7 (1).
- Meher, Biswa Ranjan, and Yixuan Wang. 2015. "Exploring the Drug Resistance of V32I and M46L Mutant HIV-1 Protease to Inhibitor TMC114: Flap Dynamics and Binding Mechanism." *Journal of Molecular Graphics and Modelling* 56: 60–73.

- Meng, Xuan-Yu, Hong-Xing Zhang, Mihaly Mezei, and Meng Cui. 2012. "Molecular Docking: A Powerful Approach for Structure-Based Drug Discovery." *Curr Comput Aided Drug Des.* 7(2): 146–157.
- Michalet, Xavier. 2010. "Mean Square Displacement Analysis of Single-Particle Trajectories with Localization Error: Brownian Motion in Isotropic Medium." *Physical review. E, Statistical, nonlinear, and soft matter physics* 82(4 Pt 1), 041914.
- Mohideen, Abdul Khader Sultan. 2020. "Molecular Docking Study of L-Asparaginase i from *Vibrio Campbellii* in the Treatment of Acute Lymphoblastic Leukemia (ALL)." *Eurobiotech Journal* 4 (1): 8–16.
- Nakagawa, Hiroshi, and Taro Tamada. 2021. "Hydration and Its Hydrogen Bonding State on a Protein Surface in the Crystalline State as Revealed by Molecular Dynamics Simulation." *Frontiers in chemistry*, 9, 738077.
- Orabi, Hanaa M., Esmail M. El-Fakharany, Eman S. Abdelkhalek, and Nagwa M. Sidkey. 2019. "L-Asparaginase and L-Glutaminase: Sources, Production, and Applications in Medicine and Industry." *Journal of Microbiology, Biotechnology and Food Sciences*, 2: 179–90.
- Pagadala, Nataraj S., Khajamohiddin Syed, and Jack Tuszynski. 2017. "Software for Molecular Docking: A Review." *Biophysical Reviews*. Springer Verlag 9(2), 91–102
- Pavelka, Antonin, Eva Chovancova, and Jiri Damborsky. 2009. "HotSpot Wizard: A Web Server for Identification of Hot Spots in Protein Engineering." *Nucleic Acids Research* 37 (SUPPL. 2).
- Pitera, Jed W. 2014. "Expected Distributions of Root-Mean-Square Positional Deviations in Proteins." *Journal of Physical Chemistry B* 118 (24): 6526–30.
- Ram Lamichhane, Tika, and Madhav Prasad Ghimire. 2020 "Structural Analysis of COVID-19 Main Protease and Its Interaction with the Inhibitor N3." *ChemRxiv* 1.
- Rampogu, Shailima, Gihwan Lee, Jun Sung Park, Keun Woo Lee, and Myeong Ok Kim. 2022. "Molecular Docking and Molecular Dynamics Simulations Discover Curcumin Analogue as a Plausible Dual Inhibitor for SARS-CoV-2." *International Journal of Molecular Sciences* 23 (3).
- Raval, Keval, and Tejas Ganatra. 2022. "Basics, Types and Applications of Molecular Docking: A Review." *IP International Journal of Comprehensive and Advanced Pharmacology* 7 (1): 12–16.

- Reddy, Erva Rajeswara, Rajulapati Satish Babu, Potla Durthi Chandrasai, and Pola Madhuri. 2016. "Exploration of the Binding Modes of L-Asparaginase Complexed with Its Amino Acid Substrates by Molecular Docking, Dynamics and Simulation." *3 Biotech* 6 (1).
- Salzer, Wanda L., Barbara Asselin, Jeffrey G. Supko, Meenakshi Devidas, Nicole A. Kaiser, Paul Plourde, Naomi J. Winick, et al. 2013. "Erwinia Asparaginase Achieves Therapeutic Activity after Pegaspargase Allergy: A Report from the Children's Oncology Group." *Blood* 122 (4): 507–14.
- Samanta, Susruta, and Danilo Roccatano. 2013. "Interaction of Curcumin with PEO-PPO-PEO Block Copolymers: A Molecular Dynamics Study." *Journal of Physical Chemistry B* 117 (11): 3250–57.
- Shamim, Amen, Sumra Wajid Abbasi, and Syed Sikander Azam. 2015. "Structural and Dynamical Aspects of Streptococcus Gordonii FabH through Molecular Docking and MD Simulations." *Journal of Molecular Graphics and Modelling* 60 (July): 180–96.
- Siegel, Rebecca L., Kimberly D. Miller, Nikita Sandeep Wagle, and Ahmedin Jemal. 2023. "Cancer Statistics, 2023." *CA: A Cancer Journal for Clinicians* 73 (1): 17–48.
- Singh, Dev Bukhsh. 2020. "Computer-Aided Drug Design. Computer-Aided Drug Design." *Springer Singapore*
- Srivastava, Mitul, Lovika Mittal, Anita Kumari, Ashish Kumar Agrahari, Mrityunjay Singh, Rajani Mathur, and Shailendra Asthana. 2022. "Characterizing (Un)Binding Mechanism of USP7 Inhibitors to Unravel the Cause of Enhanced Binding Potencies at Allosteric Checkpoint." *Protein Science* 31 (9).
- Sung, Hyuna, Jacques Ferlay, Rebecca L. Siegel, Mathieu Laversanne, Isabelle Soerjomataram, Ahmedin Jemal, and Freddie Bray. 2021. "Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries." *CA: A Cancer Journal for Clinicians* 71 (3): 209–49.
- Surpeta, Bartłomiej, Carlos Eduardo Sequeiros-Borja, and Jan Brezovsky. 2020. "Dynamics, a Powerful Component of Current and Future in Silico Approaches for Protein Design and Engineering." *International Journal of Molecular Sciences*. MDPI AG 21(8):2713.
- Swain, Amy L, Mariusz Jaskólski, Dominique Houssett, J K Mohana Rao, and Alexander Włodawert. 1993. "Crystal Structure of Escherichia Coli L-Asparaginase, an

- Enzyme Used in Cancer Therapy (Amidohydrolase/Leukemia/Active Site/Aspartate).” *Proc. Natl. Acad. Sci. USA* 90(4), 1474–1478.
- Tam, Justin Z., Talulla Palumbo, Julie M. Miwa, and Brian Y. Chen. 2022. “Analysis of Protein-Protein Interactions for Intermolecular Bond Prediction.” *Molecules* 27 (19).
- Tao, Xuan, Yukun Huang, Chong Wang, Fang Chen, Lingling Yang, Li Ling, Zhenming Che, and Xianggui Chen. 2020. “Recent Developments in Molecular Docking Technology Applied in Food Science: A Review.” *International Journal of Food Science and Technology*. Blackwell Publishing Ltd 55(44)
- Topham, Christopher M, and Jeremy C Smith. 2019. “Tri-Peptide Reference Structures for the Calculation of Relative Solvent Accessible Surface Area in Protein Amino Acid Residues Running Title: Tri-Peptide Reference Structures.” *Computational biology and chemistry* 54, 33–43.
- Tou, Weng Jeong, Su Sen Chang, Cheng Chun Lee, and Calvin Yu Chian Chen. 2013a. “Drug Design for Neuropathic Pain Regulation from Traditional Chinese Medicine.” *Scientific Reports* 3.
- Townsend, Philip D., Thomas L. Rodgers, Laura C. Glover, Heidi J. Korhonen, Shane A. Richards, Lucy J. Colwell, Ehmke Pohl, et al. 2015. “The Role of Protein-Ligand Contacts in Allosteric Regulation of the Escherichia Coli Catabolite Activator Protein.” *Journal of Biological Chemistry* 290 (36): 22225–32.
- Trisciuzzi, Daniela, Orazio Nicolotti, Maria A Miteva, and Bruno O Villoutreix. 2019. “Analysis of Solvent-Exposed and Buried Co-Crystallized Ligands: A Case Study to Support the Design of Novel Protein-Protein Interaction Inhibitors.” *Drug discovery today*, 24(2), 551–559.
- Vranken, Wim. 2007. “A Global Analysis of NMR Distance Constraints from the PDB.” *Journal of Biomolecular NMR* 39 (4): 303–14.
- Wu, Di, and Zhijun Wu. 2010. “Superimposition of Protein Structures with Dynamically Weighted RMSD.” *Journal of Molecular Modeling* 16 (2): 211–22.
- Yao, Min, Yoshiaki Yasutake, Hazuki Morita, and Isao Tanaka. 2005. “Structure of the Type I L-Asparaginase from the Hyperthermophilic Archaeon Pyrococcus Horikoshii at 2.16 Å Resolution.” *Acta Crystallographica Section D: Biological Crystallography* 61 (3): 294–301.
- Zalewska-Szewczyk, Beata, Agnieszka Gach, Krystyna Wyka, Jerzy Bodalski, and Wojciech Młynarski. 2009. “The Cross-Reactivity of Anti-Asparaginase Antibodies

against Different L-Asparaginase Preparations.” *Clinical and Experimental Medicine* 9 (2): 113–16.

Zhang, Wenli, Quanyu Dai, Zhaolin Huang, and Wei Xu. 2023. “Identification and Thermostability Modification of the Mesophilic L-Asparaginase from *Limosilactobacillus Secaliphilus*.” *Applied Biochemistry and Biotechnology* 1.

Zhang, Xian, Zhi Wang, Yimai Wang, Xu Li, Manchi Zhu, Hengwei Zhang, Meijuan Xu, Taowei Yang, and Zhiming Rao. 2021. “Heterologous Expression and Rational Design of L-Asparaginase from *Rhizomucor Miehei* to Improve Thermostability.” *Biology* 10 (12).

APPENDICES

APPENDIX A

```
em.mdp
; LINES STARTING WITH ';' ARE COMMENTS
title           = Minimization ; Title of run

; Parameters describing what to do, when to stop and what to save
integrator      = steep        ; Algorithm (steep = steepest descent minimization)
emtol           = 1000.0       ; Stop minimization when the maximum force < 10.0 kJ/mol
emstep         = 0.01         ; Energy step size
nsteps         = 50000        ; Maximum number of (minimization) steps to perform

; Parameters describing how to find the neighbors of each atom and how to calculate the
interactions
nstlist        = 1            ; Frequency to update the neighbor list and long range
forces
cutoff-scheme  = Verlet
ns_type        = grid         ; Method to determine neighbor list (simple, grid)
rlist          = 1.2          ; Cut-off for making neighbor list (short range forces)
coulombtype    = PME          ; Treatment of long range electrostatic interactions
rcoulomb       = 1.2          ; long range electrostatic cut-off
vdwtype        = cutoff
vdw-modifier   = force-switch
rvdw-switch    = 1.0
rvdw           = 1.2          ; long range Van der Waals cut-off
pbc            = xyz          ; Periodic Boundary Conditions
DispCorr       = no
```

Figure A.1. Parameter file context of energy minimization step "em.mdp"

```

nvt.mdp
title = Protein-ligand complex NVT equilibration
define = -DPOSRES ; position restrain the protein and ligand
; Run parameters
integrator = md ; leap-frog integrator
nsteps = 50000 ; 2 * 50000 = 100 ps
dt = 0.002 ; 2 fs
; Output control
nstenergy = 500 ; save energies every 1.0 ps
nstlog = 500 ; update log file every 1.0 ps
nstxout-compressed = 500 ; save coordinates every 1.0 ps
; Bond parameters
continuation = no ; first dynamics run
constraint_algorithm = lincs ; holonomic constraints
constraints = h-bonds ; bonds to H are constrained
lincs_iter = 1 ; accuracy of LINCS
lincs_order = 4 ; also related to accuracy
; Neighbor searching and vdW
cutoff-scheme = Verlet
ns_type = grid ; search neighboring grid cells
nstlist = 20 ; largely irrelevant with Verlet
rlist = 1.2
vdwtype = cutoff
vdw-modifier = force-switch
rvdw-switch = 1.0
rvdw = 1.2 ; short-range van der Waals cutoff (in nm)
; Electrostatics
coulombtype = PME ; Particle Mesh Ewald for long-range electrostatics
rcoulomb = 1.2 ; short-range electrostatic cutoff (in nm)
pme_order = 4 ; cubic interpolation
fourierspacing = 0.16 ; grid spacing for FFT
; Temperature coupling
tcoupl = V-rescale ; modified Berendsen thermostat
tc-grps = Protein_LIG Water_and_ions ; two coupling groups - more
accurate
tau_t = 0.1 0.1 ; time constant, in ps
ref_t = 300 300 ; reference temperature, one for
each group, in K
; Pressure coupling
pcoupl = no ; no pressure coupling in NVT
; Periodic boundary conditions
pbc = xyz ; 3-D PBC
; Dispersion correction is not used for proteins with the C36 additive FF
DispCorr = no
; Velocity generation
gen_vel = yes ; assign velocities from Maxwell distribution
gen_temp = 300 ; temperature for Maxwell distribution
gen_seed = -1 ; generate a random seed

```

Figure A.2. Parameter file context of NVT ensemble step "nvt.mdp" that belongs first equilibration phase.

```

npt.mdp
title = Protein-ligand complex NPT equilibration
define = -DPOSRES ; position restrain the protein and ligand
; Run parameters
integrator = md ; leap-frog integrator
nsteps = 50000 ; 2 * 50000 = 100 ps
dt = 0.002 ; 2 fs
; Output control
nstenergy = 500 ; save energies every 1.0 ps
nstlog = 500 ; update log file every 1.0 ps
nstxout-compressed = 500 ; save coordinates every 1.0 ps
; Bond parameters
continuation = yes ; continuing from NVT
constraint_algorithm = lincs ; holonomic constraints
constraints = h-bonds ; bonds to H are constrained
lincs_iter = 1 ; accuracy of LINCS
lincs_order = 4 ; also related to accuracy
; Neighbor searching and vdW
cutoff-scheme = Verlet
ns_type = grid ; search neighboring grid cells
nstlist = 20 ; largely irrelevant with Verlet
rlist = 1.2
vdwtype = cutoff
vdw-modifier = force-switch
rvdw-switch = 1.0
rvdw = 1.2 ; short-range van der Waals cutoff (in nm)
; Electrostatics
coulombtype = PME ; Particle Mesh Ewald for long-range electrostatics
rcoulomb = 1.2
pme_order = 4 ; cubic interpolation
fourierspacing = 0.16 ; grid spacing for FFT
; Temperature coupling
tcoupl = V-rescale ; modified Berendsen thermostat
tc-grps = Protein_LIG Water_and_ions ; two coupling groups - more accurate
tau_t = 0.1 0.1 ; time constant, in ps
ref_t = 300 300 ; reference temperature, one for each group, in K
; Pressure coupling
pcoupl = Berendsen ; pressure coupling is on for NPT
pcoupltype = isotropic ; uniform scaling of box vectors
tau_p = 2.0 ; time constant, in ps
ref_p = 1.0 ; reference pressure, in bar
compressibility = 4.5e-5 ; isothermal compressibility of water, bar^-1
refcoord_scaling = com
; Periodic boundary conditions
pbc = xyz ; 3-D PBC
; Dispersion correction is not used for proteins with the C36 additive FF
DispCorr = no
; Velocity generation
gen_vel = no ; velocity generation off after NVT

```

Figure A.3. Parameter file context of NPT ensemble step "npt.mdp" that belongs second equilibration phase.

```

md.mdp
title = Protein-ligand complex MD simulation
; Run parameters
integrator = md ; leap-frog integrator
nsteps = 25000000 ; 2 * 25000000 =50000 ps (50 ns)
dt = 0.002 ; 2 fs
; Output control
nstenergy = 5000 ; save energies every 10.0 ps
nstlog = 5000 ; update log file every 10.0 ps
nstxout-compressed = 5000 ; save coordinates every 10.0 ps
; Bond parameters
continuation = yes ; continuing from NPT
constraint_algorithm = lincs ; holonomic constraints
constraints = h-bonds ; bonds to H are constrained
lincs_iter = 1 ; accuracy of LINCS
lincs_order = 4 ; also related to accuracy
; Neighbor searching and vdW
cutoff-scheme = Verlet
ns_type = grid ; search neighboring grid cells
nstlist = 20 ; largely irrelevant with Verlet
rlist = 1.2
vdwtype = cutoff
vdw-modifier = force-switch
rvdw-switch = 1.0
rvdw = 1.2 ; short-range van der Waals cutoff (in nm)
; Electrostatics
coulombtype = PME ; Particle Mesh Ewald for long-range electrostatics
rcoulomb = 1.2
pme_order = 4 ; cubic interpolation
fourierspacing = 0.16 ; grid spacing for FFT
; Temperature coupling
tcoupl = V-rescale ; modified Berendsen thermostat
tc-grps = Protein_LIG Water_and_ions ; two coupling groups - more accurate
tau_t = 0.1 0.1 ; time constant, in ps
ref_t = 300 300 ; reference temperature, one for each group, in K
; Pressure coupling
pcoupl = Parrinello-Rahman ; pressure coupling is on for NPT
pcoupltype = isotropic ; uniform scaling of box vectors
tau_p = 2.0 ; time constant, in ps
ref_p = 1.0 ; reference pressure, in bar
compressibility = 4.5e-5 ; isothermal compressibility of water, bar^-1
; Periodic boundary conditions
pbc = xyz ; 3-D PBC
; Dispersion correction is not used for proteins with the C36 additive FF
DispCorr = no
; Velocity generation
gen_vel = no ; continuing from NPT equilibration

```

Figure A.4. Parameter file context of NPT ensemble step "md.mdp" that are required to start the MD run.

```

ie.mdp
title = Protein-ligand complex MD simulation
; Run parameters
integrator = md ; leap-frog integrator
nsteps = 5000000 ; 2 * 5000000 = 10000 ps (10 ns)
dt = 0.002 ; 2 fs
; Output control
nstenergy = 5000 ; save energies every 10.0 ps
nstlog = 5000 ; update log file every 10.0 ps
nstxout-compressed = 5000 ; save coordinates every 10.0 ps
energygrps = Protein LIG
; Bond parameters
continuation = yes ; continuing from NPT
constraint_algorithm = lincs ; holonomic constraints
constraints = h-bonds ; bonds to H are constrained
lincs_iter = 1 ; accuracy of LINCS
lincs_order = 4 ; also related to accuracy
; Neighbor searching and vdW
cutoff-scheme = Verlet
ns_type = grid ; search neighboring grid cells
nstlist = 20 ; largely irrelevant with Verlet
rlist = 1.2
vdwtype = cutoff
vdw-modifier = force-switch
rvdw-switch = 1.0
rvdw = 1.2 ; short-range van der Waals cutoff (in nm)
; Electrostatics
coulombtype = PME ; Particle Mesh Ewald for long-range electrostatics
rcoulomb = 1.2
pme_order = 4 ; cubic interpolation
fourierspacing = 0.16 ; grid spacing for FFT
; Temperature coupling
tcoupl = V-rescale ; modified Berendsen thermostat
tc-grps = Protein_LIG Water_and_ions ; two coupling groups - more accurate
tau_t = 0.1 0.1 ; time constant, in ps
ref_t = 300 300 ; reference temperature, one for each group, in K
; Pressure coupling
pcoupl = Parrinello-Rahman ; pressure coupling is on for NPT
pcoupltype = isotropic ; uniform scaling of box vectors
tau_p = 2.0 ; time constant, in ps
ref_p = 1.0 ; reference pressure, in bar
compressibility = 4.5e-5 ; isothermal compressibility of water, bar^-1
; Periodic boundary conditions
pbc = xyz ; 3-D PBC
; Dispersion correction is not used for proteins with the C36 additive FF
DispCorr = no
; Velocity generation
gen_vel = no ; continuing from NPT equilibration

```

Figure A.5. Parameter file context of “ie.mdp” that provide to analysis of protein-ligand interaction energies as Coul-SR and LJ-SR energies after MD run.

```

; Command line:
;   gmx_mpi pdb2gmx -f TKA_relax.pdb -o TKA_processed.gro -water spce -ignh
; Force field was read from the standard GROMACS share directory.
;

; Include forcefield parameters
#include "amber99sb.ff/forcefield.itp"

; Include chain topologies
#include "topol_Protein_chain_A.itp"
#include "topol_Protein_chain_B.itp"

; Include water topology
#include "amber99sb.ff/spce.itp"

#ifdef POSRES_WATER
; Position restraint for each water oxygen
[ position_restraints ]
; i funct      fcx      fcy      fcz
... 1 1      1000      1000      1000
#endif

; Include topology for ions
#include "amber99sb.ff/ions.itp"

[ system ]
; Name
16-MAR-23   XXXX in water

[ molecules ]
; Compound      #mols
Protein_chain_A    1
Protein_chain_B    1
SOL                37459
NA                 10

```

Figure A.6. An example topology file content of protein in water simulation

```

; Command line:
;   gmx_mpi pdb2gmx -f wildkoda.pdb -o wildkoa_processed.gro -ter -ignh
; Force field was read from current directory or a relative path - path added.
;

; Include forcefield parameters
#include "./charmm36-jul2022.ff/forcefield.itp"

; Include ligand topology
#include "LIG.itp"

; Include chain topologies
#include "topol_Protein_chain_A.itp"
#include "topol_Protein_chain_B.itp"

; Ligand position restraints
#ifdef POSRES
#include "posre_LIG.itp"
#endif

; Include water topology
#include "./charmm36-jul2022.ff/tip3p.itp"

#ifdef POSRES_WATER
; Position restraint for each water oxygen
[ position_restraints ]
; i funct      fcx      fcy      fcz
| 1 1         1000     1000     1000
#endif

; Include topology for ions
#include "./charmm36-jul2022.ff/ions.itp"

[ system ]
; Name
Protein in water

[ molecules ]
; Compound      #mols
Protein_chain_A  1
Protein_chain_B  1
LIG              1
SOL              64555
NA               9

```

Figure A.7. An example topology file content of protein and ligand complex MD simulations.

APPENDIX B

Protein in Water MD Simulation Protocol

```
# "pdb2gmx" module creates the topology of the molecule, the position constraint file
and the processed structure file as topol.top file that is seen an example in figure A.6.
$ gmx pdb2gmx -f TKA_relax.pdb -o TKA_processed.gro -water spce -ighn
# Selections:
   CHARMM36 ForceField
   TIP3P solvent
   Terminus type NH3+ / COO-
# Use editconf to define the box
$ gmx editconf -f TKA_processed.gro -o TKA_newbox.gro -c -d 2.0 -bt cubic
# Get the necessary parameters for the configuration of the simulation with the "grompp"
module and compile it as ".tpr" in the process file.
$ gmx grompp -f ions.mdp -c TKA_solv.gro -p topol.top -o ions.tpr
# Use "gmx genion" to add ions for system in order to obtain neutral charge.
$ gmx genion -s ions.tpr -o TKA_solv_ions.gro -p topol.top -pname NA -nname CL -
neutral
# Use "gmx grompp" command with em.mdp file and minimize energy of the system.
$ gmx grompp -f em.mdp -c solv_ions.gro -p topol.top -o em.tpr
$ gmx mdrun -v -deffnm em
# Use "gmx grompp" command with nvt.mdp file and balance volume and temperature
of the system.
$ gmx grompp -f nvt.mdp -c em.gro -r em.gro -p topol.top -n index.ndx -o nvt.tpr
$ gmx mdrun -deffnm nvt
# Use "gmx grompp" command with npt.mdp file and balance pressure and temperature
of the system.
$ gmx grompp -f npt.mdp -c nvt.gro -r nvt.gro -t nvt.cpt -p topol.top -o npt.tpr
$ gmx mdrun -deffnm npt
# Use "gmx grompp" command with md.mdp file and balance pressure and temperature
$ gmx grompp -f md.mdp -c npt.gro -t npt.cpt -p topol.top -o md.tpr
$ gmx mdrun -deffnm md
```

APPENDIX C

Protein-Ligand MD Simulation Protocol

#Use docked complex from molecular docking studies.

#Prepare protein and ligand separately. Bond orders "@<TRIPOS>BOND" are arranged differently in each file with texteditor.

```
$ perl sort_mol2_bonds.pl LIG.mol2 LIG.mol2
```

The LIG.mol2 file was then subjected to a CHARMM36 force field using the SWISSPARAM web tool.

#Create "LIG.gro" with editconf command

```
$ gmx editconf -f LIG.pdb -o LIG.gro
```

#The ".gro" file is created from the "TkA_relax.pdb" file, as in the protein in water simulation.

```
$ gmx pdb2gmx -f TkA_relax.pdb -o TkA_processed.gro -water spce -igh
```

#Selections:

- CHARMM36 ForceField
- TIP3P solvent
- Terminus type NH3+ / COO-

Create complex.gro with texteditor from TkA_processed.gro and LIG.gro

Use editconf to define the box

```
gmx editconf -f TkA_processed.gro -o TkA_newbox.gro -c -d 2.0 -bt cubic
```

Get the necessary parameters for the configuration of the simulation with the "grompp" module and compile it as ".tpr" in the process file.

```
gmx grompp -f ions.mdp -c TkA_solv.gro -p topol.top -o ions.tpr
```

Use "gmx genion" to add ions for system in order to obtain neutral charge.

```
gmx genion -s ions.tpr -o TkA_solv_ions.gro -p topol.top -pname NA -nname CL -neutral
```

Use "gmx grompp" command with em.mdp file and minimize energy of the system.

```
$ gmx grompp -f em.mdp -c solv_ions.gro -p topol.top -o em.tpr
```

```
$ gmx mdrun -v -deffnm em
```

Use "gmx grompp" command with nvt.mdp file and balance volume and temperature of the system.

```
$ gmx grompp -f nvt.mdp -c em.gro -r em.gro -p topol.top -n index.ndx -o nvt.tpr
```

```
$ gmx mdrun -deffnm nvt
```

```
# Use “gmx grompp” command with npt.mdp file and balance pressure and temperature  
of the system.
```

```
$ gmx grompp -f npt.mdp -c nvt.gro -r nvt.gro -t nvt.cpt -p topol.top -o npt.tpr
```

```
$ gmx mdrun -deffnm npt
```

```
# Use “gmx grompp” command with md.mdp file and balance pressure and temperature  
of the system.
```

```
$ gmx grompp -f md.mdp -c npt.gro -t npt.cpt -p topol.top -o md.tpr
```

```
$ gmx mdrun -deffnm md
```

APPENDIX D

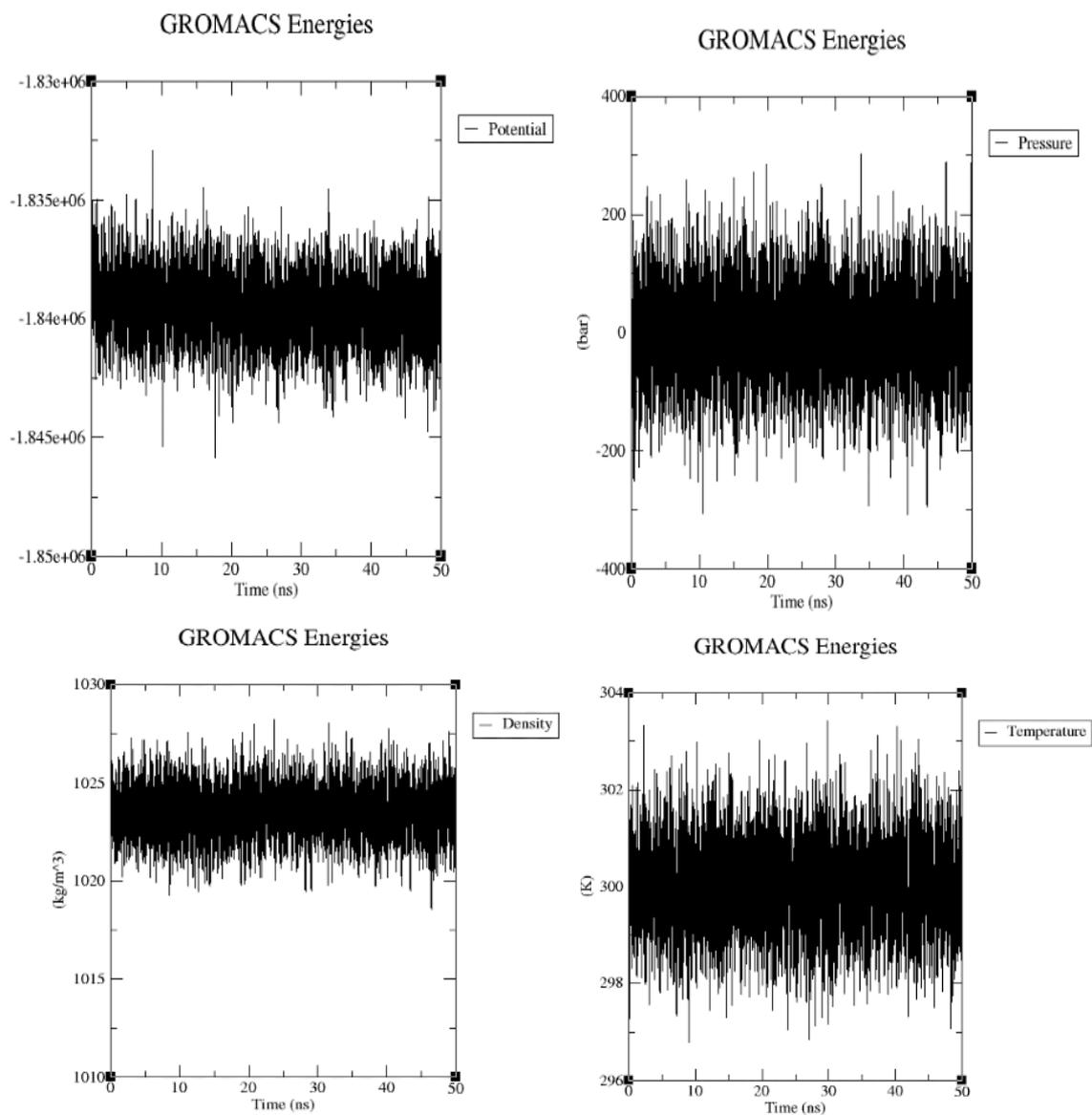


Figure D.1. Apo WT Tka in water MD simulation with initial parameters plots.

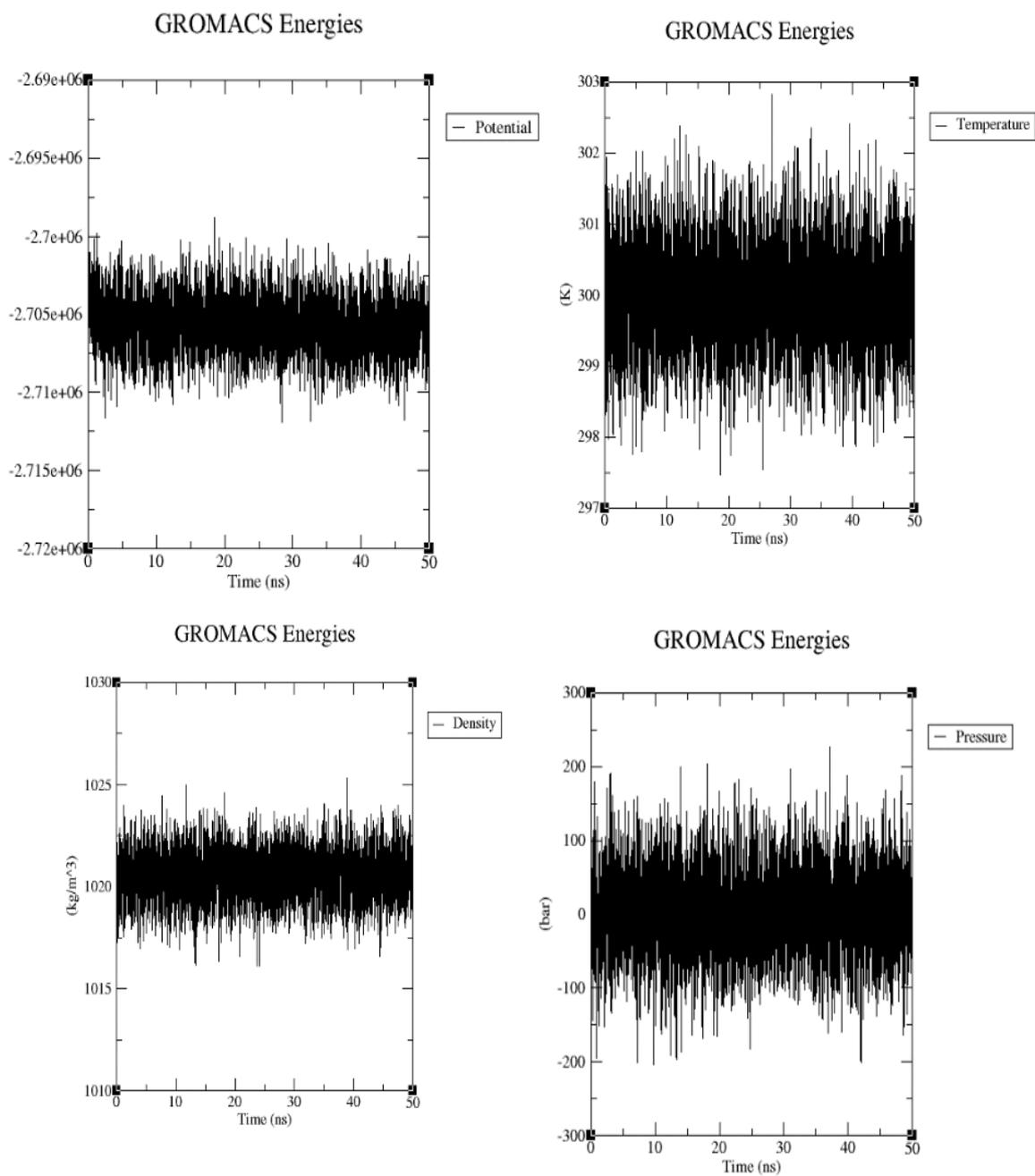


Figure D.2. WT TkA with ligand, MD simulation with initial parameter plots.

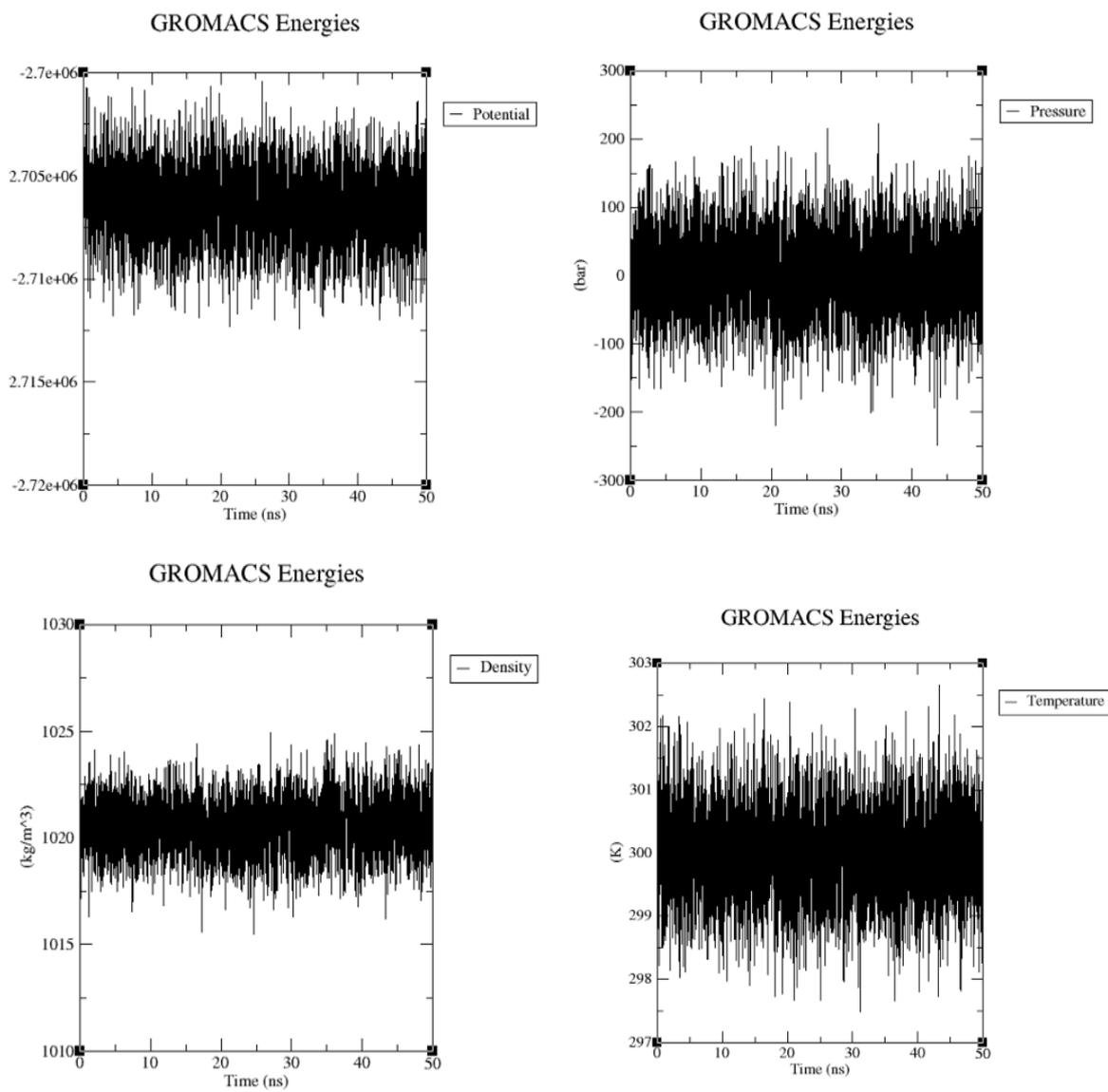


Figure D.3. T55E TkA and L-asparagine ligand initial parameter plots.

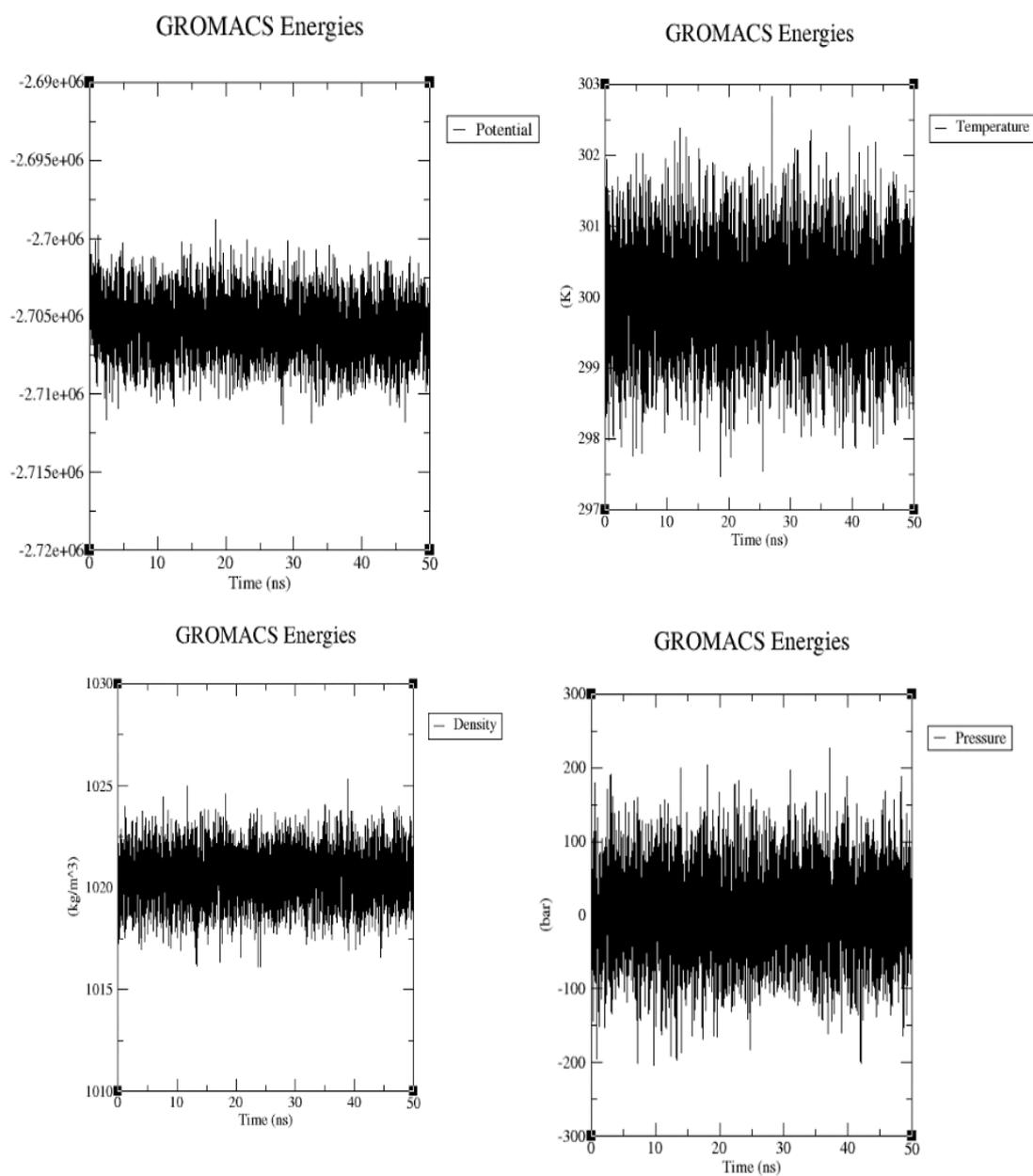


Figure D.4. D86S TkA and L-asparagine ligand initial parameter plots.

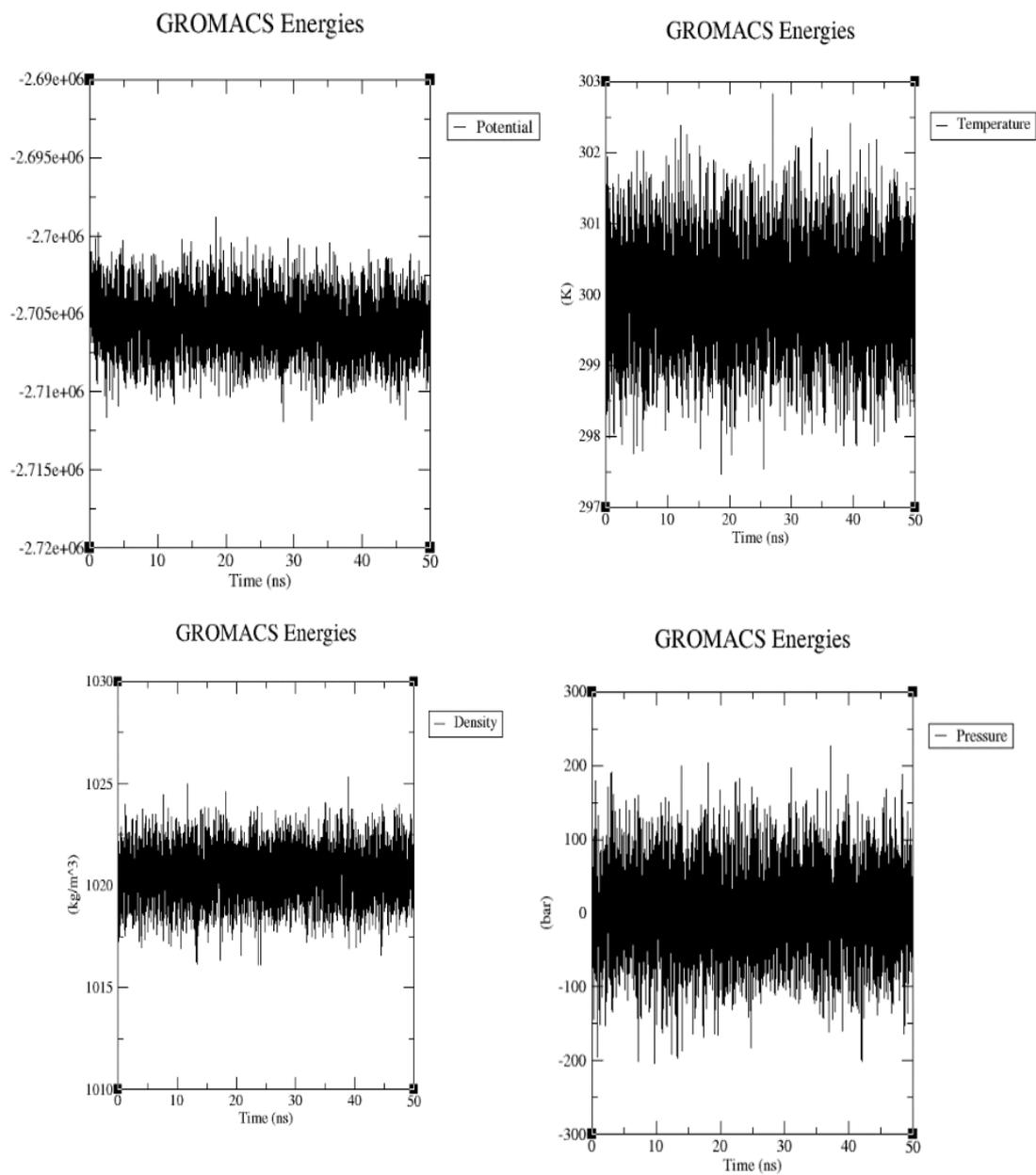


Figure D.5. T11E TkA and L-asparagine ligand initial parameter plots.