

PHARMACOLOGICAL CHAPERONE SEARCH FOR ARYLSULFATASE A  
ENZYME (ARSA) AND CHARACTERIZATION OF TWO NOVEL MUTATIONS

by

Ayşe Eren

B.S., Chemical Engineering, Boğaziçi University, 2015

Submitted to the Institute for Graduate Studies in  
Science and Engineering in partial fulfillment of  
the requirements for the degree of  
Master of Science

Graduate Program in Chemical Engineering  
Boğaziçi University  
2017

## ACKNOWLEDGEMENTS

First, I would like to gratitude to my thesis supervisor, Assoc. Prof. Elif Özkırmırlı and my thesis co-supervisor, Prof. Kutlu Ülgen. If they were not, I would not be able to learn so many things and decide to pursue an academic career. Their endless support and encouragement were so valuable for me. I would like to thank them for always being there for me to complete my Master's degree.

I would also like to thank my committee members, Prof. Asuman Özkara, Prof. Nilgün Lütfiye Karalı, and Assoc. Prof. Burak Alakent for their valuable comments on this thesis. I would like to especially express my thankfulness to Prof. Asuman Özkara for providing the starting point of this project, to Prof. Nilgün Lütfiye Karalı for synthesizing the molecules that we used in our studies, and to Assoc. Prof. Burak Alakent for all his support and help whenever I asked during my undergraduate and Master's degree studies.

Life without friends would not be cherished. I cannot even imagine where I would be today if they were not there. A handful of friends who have given me a heart full of joy, that is why I would like to thank my friends for all of our good memories: Aslı, Pınar, Canan, Ece, Esra, Dicle, İdil, Esra, Umut, Begüm, Göksu, Oğuz, and my classmates at Boğaziçi University. I am also grateful to Begüm Alaybeyoğlu and Özge Bayraktar for their help, friendship, and support whenever I asked.

Finally, I would like to express my gratefulness to my mom and dad for everything. I dedicate my thesis to them.

Last, I would like to acknowledge Technological Research Council of Turkey (TUBITAK Project 115S208) for financial support.

## **ABSTRACT**

### **PHARMACOLOGICAL CHAPERONE SEARCH FOR ARYLSULFATASE A ENZYME (ARSA) AND CHARACTERIZATION OF TWO NOVEL MUTATIONS**

Arylsulfatase A (ARSA) is an important lysosomal enzyme whose deficiency due to misfolding or oligomerization loss causes a lethal neurodegenerative disease called Metachromatic Leukodystrophy Disease (MLD). This is an autosomal, recessively inherited and fatal disease causing loss of motor function and intellectual skills mostly in infants and children. Expensive treatments such as enzyme replacement therapy can be used to increase ARSA levels, but any drug candidate has not been determined yet, thus there is an urgent need for the discovery of a drug candidate for ARSA. Pharmacological chaperones (PC) have been considered to be drug candidates for neurodegenerative diseases. The aim of this study was to understand the effects of two disease causing mutations and to find a novel PC candidate for two novel ARSA mutations: W318C and E307K+T391S. Molecular dynamics (MD) simulations were performed on wild type (WT) and mutant ARSA monomers and dimers at both acidic, and neutral pH to understand the effects of the mutations on the enzyme. The analyses of the MD simulations showed that E307K+T391S might be causing a change in the association angle at the dimer interface, while W318C may be affecting folding. Docking calculations and pharmacophore model search were performed on the monomer forms using various molecule sets and natural ligands. Six of these molecules were selected for further MD simulations. The results of the apo forms and ligand bound forms were compared in terms of enzyme dynamics and structural changes on the protein. Finally, two molecules (R\_5 and ZINC90709065) were selected as putative PC candidates for W318C mutant ARSA.

## ÖZET

### ARİLSÜLFATAZ A ENZİMİ İÇİN FARMAKOLOJİK ŞAPERON ARAMA VE İKİ ÖZGÜN MUTASYONUN NİTELENDİRİLMESİ

Arilsülfataz A yanlış katlanmadan veya oligomerleşme kaybından dolayı çalışmadığı zaman nörodejeneratif ve ölümcül bir hastalık olan Metakromatik Lökodistrofi Hastalığı'na yol açan önemli bir lizozomal enzimdir. Bu hastalık otozomal resesif geçişli hastalık olup genellikle bebeklerde ve çocuklarda motor fonksiyonların ve düşünsel yetilerin kaybolmasına ve ölüme sebep olur. Enzim replasman tedavisi gibi pahalı tedavi yöntemleri enzim seviyesini yükseltmek için uygulanıyor olsa da bu hastalık için henüz bir ilaç adayı bulunamamıştır; bu nedenle bir ön ilaç adayının Arilsülfataz A enzimi için keşfedilmesine acil ihtiyaç vardır. Farmakolojik şaperonlar, nörodejeneratif hastalıkların tedavisi için değerlendirilmeye başlanmıştır. Bu çalışmanın amacı W318C ve E307K+T391S mutasyonlarının Arilsülfataz A enzimi üzerindeki etkilerini anlamak ve bu mutant enzimler için özgün farmakolojik şaperon adayları bulmaktır. Moleküler dinamik (MD) simülasyonları mutasyonların enzim üzerindeki etkilerini anlamak için, yabanıl tür ve mutant monomer ve dimer Arilsülfataz A yapıları kullanılarak hem asidik hem de nötr pH ortamında koşturulmuştur. MD simülasyonlarının analizleri göstermiştir ki E307K+T391S mutasyonu dimer yapıdaki dimeleşme bölgesindeki birleşme açısını etkileyebilirken W318C mutasyonunun enzimin katlanmasını etkileyebilmektedir. Bağlama hesaplamaları ve farmakofor model araması monomer yapı üzerinde değişik molekül setleri kullanılarak yapılmıştır. Bu moleküllerden altı tanesi ikinci aşama MD simülasyonları için seçilmiştir. Açık form ve ligand bağlı formların sonuçları enzim dinamikleri ve protein yapısı üzerindeki etkileri bakımından karşılaştırılmıştır. Sonunda, enzim dinamikleri sonuçları bu iki molekülün (R\_5 and ZINC90709065) geliştirilerek W318C mutant enzimi için varsayımsal farmakolojik şaperon adayı olabileceğini göstermiştir.

## TABLE OF CONTENTS

ACKNOWLEDGEMENTS .....	iii
ABSTRACT .....	iv
ÖZET .....	v
TABLE OF CONTENTS .....	vi
LIST OF FIGURES .....	ix
LIST OF TABLES .....	xxii
LIST OF ACRONYMS/ABBREVIATIONS .....	xxviii
1. INTRODUCTION .....	1
2. THEORETICAL BACKGROUND .....	3
2.1. Sphingolipid Metabolism .....	3
2.2. Arylsulfatase A (ARSA) .....	4
2.3. Trafficking of Arylsulfatase A .....	5
2.4. Metachromatic Leukodystrophy Disease (MLD) .....	8
2.5. Pharmacological Chaperones (PCs) .....	9
2.6. Previous Treatments for MLD .....	9
3. METHODS .....	10
3.1. Structure Similarity Search and Residue Conservation Calculations .....	10
3.2. Protein Preparation .....	10
3.3. Ligand Preparation .....	11
3.4. Receptor Grid Generation .....	12
3.5. Screening and Docking Calculations .....	12
3.6. Clustering the Ligands .....	13
3.7. Pharmacophore Modeling .....	13
3.7.1. Ligand-based Pharmacophore Modeling .....	13
3.7.2. Structure-based Pharmacophore Modeling .....	14
3.8. Analysis of Docking Results .....	15
3.9. Molecular Dynamics (MD) Simulations .....	16
3.10. Cross-Correlation Calculations .....	17
3.11. RMSD and RMSF Calculations .....	17
3.12. Secondary Structure Calculations .....	18

3.13. Binding Free Energy Calculations .....	19
3.14. Solvent Accessible Surface Area (SASA) Calculations .....	19
3.15. Average Structure Calculations .....	19
3.16. Possible Salt Bridges and Hydrogen Bond Occupancy Calculations .....	20
3.17. Calculations of Dihedral Angles .....	20
4. RESULTS AND DISCUSSION .....	21
4.1. Residue Conservation and Similarity with Other Proteins .....	21
4.2. MD Simulations on Monomeric ARSA at Neutral pH .....	22
4.2.1. RMSD and RMSF Calculations .....	22
4.2.2. Salt Bridges and Hydrogen Bond Occupancies .....	37
4.2.3. Secondary Structure Analysis .....	39
4.2.4. State of NAG+NDG Sugar Moiety at the Glycosylation Site .....	41
4.2.5. The Distance between Glu/Lys307 and Lys395 .....	42
4.2.6. Dihedral Angle Changes in the E307K+T391S Mutant .....	43
4.3. MD Simulations on Monomeric ARSA at Acidic pH .....	46
4.3.1. RMSD Calculations for the Overall Structure and the Active Site .....	46
4.3.2. RMSD Calculations of the Specific Regions in Detail .....	49
4.3.3. RMSF Calculations .....	76
4.3.4. State of NAG+NDG Sugar Moiety at the Glycosylation Site .....	81
4.3.5. Salt Bridges and Hydrogen Bonding Occupancies .....	83
4.3.6. Secondary Structure Analysis .....	87
4.3.7. Solvent Accessible Surface Area (SASA) Calculations .....	89
4.4. Pharmacophore Modeling .....	91
4.4.1. Ligand-Based Pharmacophore Modeling .....	91
4.4.2. Structure-Based Pharmacophore Modeling .....	91
4.5. Docking Results at Neutral pH .....	92
4.5.1. Docking of Natural Ligands .....	93
4.5.2. Docking of Known Bioactive Molecules of ARSA .....	96
4.5.3. Docking Calculations of aSMase Inhibitors .....	99
4.5.4. IF Docking of Natural Substrates .....	101
4.5.5. Docking Calculations of Other Molecules .....	103
4.5.6. Docking of de novo Synthesized Molecules .....	106
4.6. Docking Results at Acidic pH .....	107

4.6.1. Docking of Natural Ligands .....	107
4.6.2. Docking of Known Bioactive Molecules of ARSA .....	111
4.6.3. Docking of aSMase Inhibitors .....	113
4.6.4. Docking Cerebroside 3-Sulfate at Acidic pH .....	121
4.7. What Happens When a Ligand is Docked to Mutant Structure? .....	131
4.7.1. MD Simulations on Apo and Holo Forms of W318C, and the Selected de novo Synthesized Molecule (D28) Complexes .....	132
4.7.2. MD Simulations on R_5 Bound WT and W318C Mutant Monomers ....	137
4.7.3. MD Simulation on ZINC90709065 compound Bound W318C Mutant Monomer .....	154
4.8. MD Simulations on Dimer ARSA Forms at Neutral pH .....	163
4.8.1. RMSD and RMSF Calculations – Overall Structure .....	167
4.8.2. RMSD Calculations – Key Structural Elements .....	173
4.8.3. Salt Bridges and Hydrogen Bond Occupancy Calculations .....	178
4.8.4. Monomer-Monomer Clashes and SASA Calculations .....	182
4.9. MD Simulations on Dimer ARSA Forms at Acidic pH .....	184
4.9.1. Visual Comparison of the Simulation Snapshots .....	184
4.9.2. Structural Changes and Average Structure .....	187
4.9.3. RMSD and RMSF Calculations – Overall Structure .....	189
4.9.4. RMSD Calculations – Key Structural Elements .....	191
4.9.5. Salt Bridges and Hydrogen Bond Occupancy Calculations .....	200
4.10. Cross Correlation Calculations on Monomer Forms at Neutral pH .....	203
5. CONCLUSIONS AND RECOMMENDATIONS .....	208
5.1. Conclusions .....	208
5.2. Recommendations .....	209
REFERENCES .....	211
APPENDIX A: ANALYSIS OF MOLECULAR DYNAMICS SIMULATIONS .....	225
APPENDIX B: DOCKING CALCULATIONS AND PHYSICOCHEMICAL PROPERTIES OF THE MOLECULES .....	277
APPENDIX C: R CODES FOR CROSS CORRELATION CALCULATIONS .....	301
APPENDIX D: CHEMICAL STRUCTURES OF DE NOVO SYNTHESIZED MOLECULES .....	303

## LIST OF FIGURES

Figure 2.1.	A part of the sphingolipid metabolism. ....	3
Figure 2.2.	Late infantile type MLD causing mutations (red), juvenile or adult types of MLD causing mutations (yellow) and epitope residues (green) on ARSA. ....	7
Figure 4.1.	The backbone RMSD as a function of time for the simulations on monomeric ARSA at neutral pH. ....	23
Figure 4.2.	Last frames from the MD simulations of the structures of WT, W318C and E307K+T391S mutant forms of ARSA were aligned. ....	24
Figure 4.3.	Apo WT ARSA structure and its important regions. ....	26
Figure 4.4.	Residues with RMSF values higher than 2.0 Å are shown as spheres on the WT ARSA structure. ....	28
Figure 4.5.	Residue based RMSF profiles in the simulations on monomeric ARSA. ....	29
Figure 4.6.	Sharp changes in the RMSD profile of the dimerization1 loop of W318C mutant form. ....	32
Figure 4.7.	Snapshots from the selected time steps (0, 22, 30, 50 ns) of the simulation on E307K+T391S mutant form. ....	33
Figure 4.8.	RMSD profiles of the dimerization3 regions of all ARSA mutant forms. ....	34
Figure 4.9.	Dimerization3 region of E307K+T391S the mutant is shown in red for first frame and 44 ns. ....	34
Figure 4.10.	RMSD profiles of the octamerization helices of all ARSA monomers. ..	35

Figure 4.11.	The biological assembly of ARSA (PDB code: 1AUK). Val470 ends on the octamerization helices are indicated with red circles. ....	36
Figure 4.12.	Secondary structure maps of the WT, W318C mutant and E307K+T391S mutant. ....	40
Figure 4.13.	Snapshots at 0 ns (left) and 50 ns (right) of T274M mutant monomer. ....	42
Figure 4.14.	The distance between Glu/Lys307, Lys395 on WT, E307K+T391S and W318C mutant forms as a function of simulation time. ....	42
Figure 4.15.	The psi angle changes of the residues of E307K+T391S monomer (vertical axis) as a function of simulation time. ....	44
Figure 4.16.	Backbone RMSD profiles of the four different ARSA forms for the simulations at acidic pH. ....	47
Figure 4.17.	RMSD profiles of the active site residues for simulations at acidic pH. ..	48
Figure 4.18.	RMSD profiles of the dimerization1 regions of all eight ARSA forms. ...	49
Figure 4.19.	Dimerization1 RMSD profiles of the nonhydrated forms. ....	50
Figure 4.20.	RMSD profiles of the dimerization1 region of nonhydrated forms of WT, W318C and E307K+T391S mutant forms of ARSA. ....	51
Figure 4.21.	Dimerization1 loop (red) of W318C-nonhydrated mutant type of ARSA at 0 ns, 32 ns and 45 ns of the simulation. ....	51
Figure 4.22.	The conformational change of Pro42 (yellow) on the dimerization1 region (red) of W318C-nonhydrated mutant form. ....	52

Figure 4.23.	RMSD profiles of the dimerization1 region of diol forms of WT, W318C and E307K+T391S mutant forms of ARSA. ....	52
Figure 4.24.	Dimerization1 region of E307K+T391S-diol form at 0 ns, 7 ns, and 28 ns of the simulation. ....	53
Figure 4.25.	The conformational change of Thr45 (yellow) on the dimerization1 region (red) of E307K+T391S-diol mutant form. ....	54
Figure 4.26.	Thr45 (circled) on the dimerization1 loop (red) of WT-diol. ....	54
Figure 4.27.	RMSD profiles of the dimerization2 regions from all simulations on monomers at acidic pH. ....	55
Figure 4.28.	RMSD profiles of the dimerization2 regions of two WT forms of ARSA. ....	56
Figure 4.29.	Dimerization2 regions of both WT ARSA forms (upper panel: diol and lower panel: nonhydrated) at 0 ns, 25 ns, 36 ns and 50 ns. ....	57
Figure 4.30.	RMSD profiles of the dimerization2 regions of four diol forms. ....	58
Figure 4.31.	RMSD profiles of the dimerization2 regions of four nonhydrated forms. ....	58
Figure 4.32.	Met289 on the dimerization3 region of W318C-nonhydrated form of ARSA at 26 ns and 40 ns. ....	59
Figure 4.33.	RMSD profiles of the dimerization3 regions of the all forms of ARSA. ....	60
Figure 4.34.	Dimerization3 region of E307K+T391S-diol form of ARSA. ....	61
Figure 4.35.	Dimerization3 region of WT-diol form of ARSA. ....	61
Figure 4.36.	Dimerization3 region of WT-diol form at 35 ns. ....	62

Figure 4.37.	RMSD profiles of the dimerization <sup>3</sup> regions of WT structures: two at acidic pH and one at neutral pH. ....	63
Figure 4.38.	RMSD profiles of the octamerization helices of all forms of ARSA. ....	64
Figure 4.39.	Structural changes on the octamerization helix of E307K+T391S-nonhydrated mutant form of ARSA. ....	65
Figure 4.40.	Leu466 and Lys307 (yellow) are indicated with black arrows at 0 ns, 9 ns, 14 ns, 20 ns, 32 ns, 39 ns of the simulation on E307K+T391S-nonhydrated form. ....	65
Figure 4.41.	Octamerization helix of WT-nonhydrated form at 0 ns, 4 ns and 26 ns. .	66
Figure 4.42.	Leu466 (right yellow) and Asp467 (left yellow) on the octamerization helix (red) of WT-nonhydrated form at 0 ns, 18 ns and 26 ns. ....	67
Figure 4.43.	RMSD profiles of the octamerization helices of all forms of ARSA with FGL residues in diol state. ....	68
Figure 4.44.	RMSD profiles of the octamerization helices of all forms of ARSA with FGL residues in nonhydrated state. ....	69
Figure 4.45.	Asp335-Arg370 region RMSD profiles of all simulations at acidic pH. ..	70
Figure 4.46.	Thr352 (right yellow residue) and Asn350 (left yellow residue) at 14th ns and 45th ns of the simulation on W318C-nonhydrated mutant. ....	71
Figure 4.47.	RMSD profiles of the Asp335-Arg370 region for the two states of E307K+T391S mutant. ....	71
Figure 4.48.	Thr365 (yellow) is shown on the Asp335-Arg370 region of E307K+T391S-diol mutant. ....	72

Figure 4.49.	Residue based RMSF values at acidic pH. ....	77
Figure 4.50.	Experimental and calculated B-factors of the carbon alpha atoms. ....	80
Figure 4.51.	Comparison of the RMSF results of both forms of WT ARSA. ....	81
Figure 4.52.	NDG+NAG sugar moiety (sticks at 0 ns, 20 ns, 31 ns, 50 ns of the simulation on E307K+T391S-diol ARSA at neutral pH. ....	82
Figure 4.53.	NDG+NAG sugar moiety (shown with sticks representation) at 0 ns, 22 ns, 36 ns, 50 ns of the simulation on W318C mutant ARSA at neutral pH. ....	83
Figure 4.54.	Asp61, Glu307, Arg311, and Arg390 on the original crystal structure of ARSA (PDB code: 1AUK). ....	87
Figure 4.55.	The secondary structure maps of E307K+T391S-nonhydrated (top) and W318C-nonhydrated (bottom) forms. ....	89
Figure 4.56.	SASA profiles as a function of simulation time at acidic pH. ....	90
Figure 4.57.	Active site of the p-nitrocatechol sulfate bound ARSA crystal structure (1E2S in blue) and SGC docked WT ARSA (1AUK in pink) structure. ....	102
Figure 4.58.	Interaction map between SGC and ARSA. ....	103
Figure 4.59.	Binding free energy results for the de novo synthesized molecules after the docking calculations on W318C mutant. ....	106
Figure 4.60.	The interaction map between rigidly docked C3S (GScore: -7.83 kcal/mol) and the WT-nonhydrated form of ARSA. ....	122

Figure 4.61.	The interaction map between rigidly docked C3S (Gscore: -9.79 kcal/mol) and the E307K+T391S-nonhydrated form of ARSA. ....	123
Figure 4.62.	The interaction map between the first pose from IFD results of WT-diol form in Table 4.32. ....	125
Figure 4.63.	The interaction map between the last pose from IFD results of WT-diol form in Table 25 is shown (8th pose). ....	125
Figure 4.64.	D28 docked W318C monomers at 0 ns and 22 ns of simulation at neutral pH. ....	132
Figure 4.65.	Snapshots at 36 ns (left) and 50 ns (right) of the first simulation on IF docked D28 + W318C monomer complex at neutral pH. ....	133
Figure 4.66.	Snapshots at 0 ns (left) and 26 ns (right) of the second simulation on IF docked D28 + W318C monomer complex at neutral pH. ....	133
Figure 4.67.	Snapshots at 46 (left) and 50 ns (right) of the second simulation on IF docked D28 + W318C monomer complex at neutral pH. ....	134
Figure 4.68.	Snapshots at 0 ns (left) and 11 ns (right) of the second simulation on rigidly docked D28 + W318C monomer complex at neutral pH. ....	134
Figure 4.69.	Snapshots at 43 ns (left) and 50 ns (right) of the second simulation on rigidly docked D28 + W318C monomer complex at neutral pH. ....	135
Figure 4.70.	RMSD profiles of dimerization1 region for WT, apo W318C mutant and ligand bound W318C mutant monomers at neutral pH. ....	135
Figure 4.71.	RMSD profiles of dimerization2 for WT, apo W318C mutant, and ligand bound W318C mutant monomers at neutral pH. ....	136

Figure 4.72.	RMSD profiles of the backbones for the apo and ligand docked WT and W318C mutant monomers of ARSA. ....	138
Figure 4.73.	RMSD profiles of the active site carbon alpha atoms at neutral pH. ....	139
Figure 4.74.	Residue based RMSF values for WT, W318C mutant, E307K+T391S mutant, WT + ligand, and W318C +ligand simulations. ....	140
Figure 4.75.	Residue based RMSF values for apo and ligand bound W318C mutant monomers. ....	141
Figure 4.76.	RMSD profiles of dimerization1 loops for apo and ligand bound ARSA monomers at neutral pH. ....	142
Figure 4.77.	Pro42 on the dimerization1 loop (red) of apo W318C mutant form at 0 ns, 31 ns, and 50 ns. ....	143
Figure 4.78.	Pro42 (sticks) on the dimerization1 loop (red) of ligand bound W318C mutant form at 0 ns, 31 ns, and 50 ns. ....	143
Figure 4.79.	Thr45 at the 8th ns of the simulation on apo WT form of ARSA at 0ns, 8 ns and 50 ns. ....	144
Figure 4.80.	The conformations of the dimerization1 region (red) of ligand bound WT ARSA form at neutral pH at 0ns, 8ns and 50ns. ....	145
Figure 4.81.	Dimerization2 region RMSD profiles of ARSA at neutral pH. ....	146
Figure 4.82.	Met289 (shown with stick representation) in apo W318C monomer (top panel) and ligand bound W318C mutant form (bottom panel). ....	147
Figure 4.83.	RMSD profiles of the dimerization3 region at neutral pH. ....	148

Figure 4.84.	RMSD profiles of the octamerization helices at neutral pH. ....	149
Figure 4.85.	Snapshots from at 0 ns (left) and 50 ns (right) for apo form of W318C mutant monomer at neutral pH. ....	151
Figure 4.86.	Snapshots at 0 ns (left) and 50 ns (right) for R_5 bound form of W318C mutant monomer at neutral pH. ....	152
Figure 4.87.	SASA profiles of the monomeric ARSA forms at neutral pH. ....	152
Figure 4.88.	Heatmap of the secondary structure analysis for R_5 bound W318C monomer. ....	153
Figure 4.89.	ZINC90709065 (orange) in the active site cavity of W318C monomer (blue) at neutral pH. ....	154
Figure 4.90.	Interaction map between ZINC90709065 and W318C monomer. ....	155
Figure 4.91.	RMSD profiles of the backbones for WT and three forms of W318C mutant monomers at neutral pH. ....	156
Figure 4.92.	RMSD profiles of the active site residues for WT and three forms of W318C mutant monomers at neutral pH. ....	157
Figure 4.93.	ZINC90709065 compound at 0 ns, 29 ns and 50 ns for the simulation on W318C monomer. ....	157
Figure 4.94.	RMSD profiles of dimerization1 regions for WT and W318C mutant monomers at neutral pH. ....	158
Figure 4.95.	Thr45 (sticks) and the dimerization1 region (red) at 0 ns, 25 ns and 50 ns of the simulation on ZINC90709065 bound W318C mutant monomer at neutral pH. ....	159

Figure 4.96. RMSD profiles of dimerization <sup>2</sup> regions for WT and three forms of W318C mutant monomers at neutral pH. ....	159
Figure 4.97. Dimerization <sup>3</sup> RMSD profiles of WT and three forms of W318C mutant monomers at neutral pH. ....	160
Figure 4.98. Octamerization helix RMSD profiles of WT and three forms of W318C mutant monomers at neutral pH. ....	161
Figure 4.99. Last frame coordinates of E307K+T391S mutant (blue), WT (green) and W318C mutant (orange) dimers. ....	164
Figure 4.100. Simulation snapshots at 0 ns (blue) and 50 ns (brown) for E307K+T391S and W318C mutant dimers. ....	165
Figure 4.101. Two octamerization helices of the dimers when the average dimer structures after alignment on the backbone structure. ....	167
Figure 4.102. Dimer backbone RMSD profiles of three forms of ARSA dimers. ....	168
Figure 4.103. RMSD profiles of the chains of the dimer forms of ARSA. ....	169
Figure 4.104. The RMSD profiles of the chains of the dimer forms of ARSA. ....	170
Figure 4.105. Residue based RMSF values of all atoms of the chains in dimers at neutral pH. ....	170
Figure 4.106. Residue based RMSF values of all dimer forms of ARSA. ....	172
Figure 4.107. The red region (Thr60-Phe220) shows high RMSF values for chain <sup>2</sup> of WT and chain <sup>1</sup> of W318C mutant dimers. ....	172

Figure 4.108. RMSD profiles of the dimerization1 regions of the chains in the dimer forms of ARSA at neutral pH. ....	174
Figure 4.109. Dimerization1 regions (red) of both chains of WT ARSA dimer. ....	174
Figure 4.110. Ser44 (stick) at 0 ns, 8 ns, 34 ns, and 50 ns, in the simulation on E307K+T391S mutant dimer. ....	175
Figure 4.111. Aligned structures for the monomer forms including WT, E307K+T391S, T274M, and P377L forms (upper panel) and specific parts of ARSA dimers aligned both based on one of the octamerization helices and the backbone (below). ....	177
Figure 4.112. WT monomers in orange and green as aligned on the WT dimer (blue). ....	182
Figure 4.113. Dimer forms of ARSA aligned based on the backbone (upper panel). Lower panel: the same forms as aligned on one of the octamerization helices. ....	185
Figure 4.114. Octamerization helices ( $\alpha$ I, horizontal ones) and longest helices ( $\alpha$ D and $\alpha$ E) of the averagere structures after backbone alignment. ....	188
Figure 4.115. The RMSD profiles of the whole dimer backbones at acidic pH. ....	189
Figure 4.116. RMSD profiles of each chain in the dimer forms at acidic pH. ....	190
Figure 4.117. RMSF profiles of the residues on each chain in dimer forms. ....	190
Figure 4.118. RMSD profiles of the dimerization1 regions of the chains in the dimers at acidic pH. ....	191
Figure 4.119. Dimerization1 region RMSD profiles of WT dimer chains at acidic pH. ....	192

Figure 4.120. RMSD profiles of the dimerization <sup>2</sup> regions of the chains in the dimers at acidic pH. ....	193
Figure 4.121. RMSD profiles of the dimerization <sup>3</sup> regions of the chains at acidic pH. ....	193
Figure 4.122. RMSD profiles of the octamerization helices of the chains at acidic pH. ....	194
Figure 4.123. His41 (yellow), Ser43 (light blue above yellow residue), Ser295 (green), and Pro435 (light blue on the dark blue chain) in both chains of C300F dimer, at 0 ns, 21 ns, and 50 ns. ....	195
Figure 4.124. The distance between the residues His41 and Ser295 in one of the chains of C300F dimer. ....	196
Figure 4.125. Met289 (in circle) at 25 ns, and 50 ns in C300F dimer. ....	196
Figure 4.126. Met289 (circled) in the other chain at 25 ns, and 50 ns in C300F dimer. ....	197
Figure 4.127. Lys307 (left) and Pro435 (right) in the same chain of E307K+T391S dimer at 8 ns and 50 ns. ....	198
Figure 4.128. Lys3307 (left) and Pro435 (right) at 0 ns and 50 ns in the simulation on E307K+T391S dimer. ....	198
Figure 4.129. Glu240 and Lys433 (yellow on the right and left) and their interaction at 42 ns (left panel) and at 50 ns (right panel). ....	199
Figure 4.130. Dimerization <sup>1</sup> region of W318C dimer at 25 ns and 50 ns. ....	199
Figure 4.131. Cross correlation maps of the four structures. ....	203
Figure 4.132. Difference between the cross correlation matrix of E307K+T391S mutant and that of WT. ....	204

Figure 4.133. Difference between the cross correlation matrix of W318C mutant and that of WT. ....	205
Figure 4.134. Difference between the cross correlation matrix of R_5 bound WT and that of WT. ....	205
Figure 4.135. The regions that were circled in the difference matrix of W318C mutant monomer at neutral pH ....	206
Figure 4.136. The regions that were circled in the difference matrix of E307K+T391S mutant monomer at neutral pH. ....	206
Figure 4.137. The regions that were circled in the difference matrix of R_5 bound WT monomer at neutral pH. ....	207
Figure A.1. Residue based RMSF values of WT, W318C mutant, E307K+T391S mutant forms, and the experimental values from PBD. The results were scaled to B factor, with the experimental B factor values. ....	225
Figure A.2. Comparison of the RMSF results of both forms of W318C mutant ARSA at acidic pH. ....	232
Figure A.3. Comparison of the RMSF results of both forms of E307K+T391S mutant ARSA at acidic pH. ....	233
Figure A.4. Comparison of the RMSF results of both forms of E307K mutant ARSA at acidic pH. ....	233
Figure D.1. The chemical structure of G7. ....	303

Figure D.2.	The chemical structure of D2. ....	304
Figure D.3.	The chemical structure of D11-tm. ....	304
Figure D.4.	The chemical structure of D28. ....	305
Figure D.5.	The chemical structure of D16. ....	305
Figure D.6.	The chemical structure of I7. ....	305
Figure D.7.	The chemical structure of I10. ....	306
Figure D.8.	The chemical structure of G2. ....	306

## LIST OF TABLES

Table 2.1.	Some of the mutations have been identified on ARSA and their corresponding phenotypes. Late infantile type of MLD is the most severe type of MLD. ....	7
Table 4.1.	Free energy calculation results (kcal/mol) from FoldX Suite. ....	22
Table 4.2.	Average and standard deviation values of the RMSD profiles of the backbone, dimerization loops, and octamerization helix. ....	25
Table 4.3.	RMSD results after the protein backbone alignment of the average structures of all forms of ARSA monomers at neutral pH. ....	27
Table 4.4.	Structurally important ARSA residues. ....	27
Table 4.5.	Average RMSF values of the residues for the overall protein structure and specific regions with respect to the initial coordinates. ....	30
Table 4.6.	Possible salt bridges in the corresponding simulations. ....	38
Table 4.7.	The residues with significant psi angle changes on the monomer form of E307K+T391S mutant when compared to WT ARSA. ....	45
Table 4.8.	Average RMSD values with their corresponding standard deviation values at acidic pH. ....	73
Table 4.9.	Average RMSF values and their corresponding standard deviations all residues and the specific regions of ARSA. ....	77
Table 4.10.	Residues, with RMSF values higher than 2.0 Å. ....	79

Table 4.11. Average distances (whole trajectory) between possible salt bridge pairs. .	84
Table 4.12. Structure-based pharmacophore features by E-pharm of two forms of ARSA monomers at neutral pH. ....	92
Table 4.13. IFD and rigid docking results of the ARSA substrates in the order of XP score (kcal/mol). ....	93
Table 4.14. The distances between the sulfate and nitrogen atoms of SGC molecules and Mg <sup>2+</sup> in the active site cavities of the WT and mutant ARSA forms and their interactions. ....	95
Table 4.15. Rigid docking results of the known bioactive molecules of ARSA from ChEMBL database in the order of XP score (kcal/mol). ....	96
Table 4.16. Rigid docking results of the known aSMase inhibitors at acidic pH in the order of XP score (kcal/mol). ....	99
Table 4.17. The rigid docking results of Verubecestat in neutral pH are given for three forms of ARSA in the increasing order of GScore values. ....	104
Table 4.18. The rigid docking results of GW4869 in neutral pH are given for three forms of ARSA in the increasing order of GScore values. ....	105
Table 4.19. Rigid docking results of the substrates of ARSA at acidic pH as XP GScores (kcal/mol). ....	107
Table 4.20. The interactions between docked ARSA substrates and eight forms of ARSA at acidic pH. ....	108
Table 4.21. Physicochemical properties of the natural substrates of ARSA for the top resulting poses of the IF docking calculations on WT forms at acidic pH.	109

Table 4.22.	XP docking results of the known bioactive molecules of ARSA in increasing order of binding free energy (Gscore in kcal/mol) for both forms of all monomers at acidic pH. ....	111
Table 4.23.	The rigid docking results of the known bioactive molecules of ARSA at acidic pH in the increasing order of Gscore values (kcal/mol) for WT-diol form. ....	112
Table 4.24.	Top different ten compounds from aSMase inhibitors dataset in increasing order of binding free energy values (GScore in kcal/mol). ....	113
Table 4.25.	The physicochemical properties of the top scoring molecules from the rigid docking calculations of aSMase inhibitors on WT-diol form. ....	115
Table 4.26.	The physicochemical properties of the top scoring molecules from the rigid docking calculations of aSMase inhibitors on WT-nonhydrated form in increasing order of GScore values. ....	116
Table 4.27.	The physicochemical properties of the top scoring molecules from the rigid docking calculations of aSMase inhibitors on W318C-diol form in increasing order of GScore values. ....	117
Table 4.28.	The physicochemical properties of the top scoring molecules from the rigid docking calculations of aSMase inhibitors on W318C-nonhydrated form in increasing order of GScore values. ....	118
Table 4.29.	The physicochemical properties of the top scoring molecules from the rigid docking calculations of aSMase inhibitors on E307K+T391S-diol form in increasing order of GScore values. ....	119
Table 4.30.	The physicochemical properties of the top scoring molecules from the rigid docking calculations of aSMase inhibitors on E307K+T391S-nonhydrated form in increasing order of GScore values. ....	120

Table 4.31. Rigid docking results of C3S for the eight forms of ARSA at acidic pH.	121
Table 4.32. The induced fit docking results of C3S to WT-diol form are given with the calculated physicochemical properties of the all resulting poses in increasing order of GScore values. ....	124
Table 4.33. The induced fit docking results for the top ten poses of C3S to WT-nonhydrated form with their physicochemical properties in increasing order of GScore. ....	126
Table 4.34. The induced fit docking results of the top ten poses of C3S to W318C-diol form in increasing order of GScore. ....	127
Table 4.35. The induced fit docking results of the top ten poses of C3S to W318C-nonhydrated form with their physicochemical properties in increasing order of GScore. ....	128
Table 4.36. The induced fit docking results of the top ten poses of C3S to E307K+T391S-nonhydrated form with their physicochemical properties in increasing order of GScore. ....	129
Table 4.37. IF docking results of top ten poses of C3S to E307K+T391S-nonhydrated form with their calculated physicochemical properties in increasing order of GScore. ....	130
Table 4.38. The possible salt bridges in the simulations on the corresponding forms of ARSA forms. ....	150
Table 4.39. Distances between the oxygen-nitrogen pairs of the basic and acidic residues for the last 20 ns. ....	162
Table 4.40. RMSD values between the simulation average structures of ARSA dimers at neutral pH. ....	166

Table 4.41.	The possible salt bridges in the ARSA chains in dimer forms. ....	179
Table 4.42.	The average SASA values of the monomers and dimers with their corresponding standard deviations in Å <sup>2</sup> . ....	183
Table 4.43.	RMSD values between last simulation snapshots of dimer ARSA forms at acidic pH. ....	185
Table 4.44.	RMSD values between the simulation average structures of ARSA dimers at acidic pH. ....	187
Table 4.45.	The possible salt bridges in the corresponding dimers at acidic pH. ....	202
Table A.1.	Hydrogen bond occupancies of the selected residues are given in percentage for all monomer forms of ARSA at neutral pH. ....	226
Table A.2.	All possible salt bridges for ARSA monomers at neutral pH. ....	231
Table A.3.	Hydrogen bond occupancies between the selected residues are given for the simulations at acidic pH. ....	234
Table A.4.	The hydrogen bond occupancy percentages of the selected residues in the chains of dimer forms of ARSA at neutral pH. ....	241
Table A.5.	All possible salt bridges of the monomers forms of WT and W318C mutant monomers at neutral pH. ....	251
Table A.6.	All hydrogen bond occupancies of the dimer forms at acidic pH. ....	252
Table A.7.	Possible salt bridge forming residues for the dimers at acidic pH. ....	269
Table B.1.	Converted IC50 and pIC50 values of the known bioactive molecules of ARSA are listed in the decreasing order of their PIC50 values. ....	277

Table D.1. The SMILES codes of the de novo synthesized molecules that were used in docking calculations. ....	303
--	-----

**LIST OF ACRONYMS/ABBREVIATIONS**

#nonHatm	Number of Non-Hydrogen Atoms
%HOA	Percent Human Oral Absorption
2D	Two Dimensional
3D	Three Dimensional
Ala	Alanine
Arg	Arginine
ARSA	Arylsulfatase A
aSMase	Acid Sphingomyelinase
Arg	Arginine
Asn	Asparagine
Asp	Aspartic Acid
Avg	Average
BEI	Binding Efficiency Index
C1	Chain 1
C2	Chain 2
C3S	Cerebroside 3-Sulfate
COM	Center of Mass
Cys	Cysteine
DeltaG	Gibbs Free Energy
Diol	Arylsulfatase A with the Hydrated Form of FGL Residue
ER	Endoplasmic Reticulum
FGL	Formyl Glycine Residue
FISA	Hydrophilic Component of Surface Area
FOSA	Hydrophobic Component of Surface Area
Frag	Fragment
Glide LE	Calculated Glide Ligand Efficiency
Gln	Glutamine
Glu	Glutamic Acid
Gly	Glycine

Hbond	Hydrogen-Bond
His	Histidine
HOA	Human Oral Absorption
IC <sub>50</sub>	Half Maximal Inhibitory Concentration
IF	Induced Fit
IFD	Induced Fit Docking
Ile	Isoleucine
Kcal	Kilocalorie
Ki	Calculated Dissociation Constant
Leu	Leucine
Lig	Ligand
Lipo	Lipophilic Term
LLE	Lipophilic Efficiency Index
LSD	Lysosomal Storage Disease
Lys	Lysine
M6P	Mannose-6-Phosphate
Mg <sup>2+</sup>	Magnesium Cation
Met	Methionine
MLD	Metachromatic Leukodystrophy Disease
MW	Molecular Weight
NMR	Nuclear Magnetic Resonance
NAG	N-Acetyl-D-Glucosamine
NDG	2-(Acetylamino)-2-Deoxy-A-D-Glucopyranose
Nonhyd	Arylsulfatase A with the Non-Hydrated Form of FGL Residue
Ns	Nanosecond
p(Ki)	Logarithm of the Dissociation Constant
PC	Pharmacological Chaperone
PDB	Protein Data Bank
PDK1	Phosphoinositide-Dependent Kinase-1
pH	Potential of Hydrogen
Phe	Phenylalanine
PISA	II Component of Surface Area
PLS	Partial Least Squares

Pro	Proline
ps	Picosecond
PSA	Polar Surface Area
PSB	Possible Salt Bridges
QPlogBB	Blood-Brain Barrier Coefficient
QPlogPo/w	Predicted Octanol/Water Partition Coefficient
QSAR	Quantitative Structure Activity Relationship
RMSD	Root-Mean-Square Deviation
RMSF	Root Mean Square Fluctuations
RO5	Rule of Five
SASA	Total Solvent Accessible Surface Area
Schröd	Schrödinger
Std	Standard
Std. Dev.	Standard Deviation
SEI	Surface Efficiency Index
Ser	Serine
SP	Standard Precision
SPC	Water Molecule (Simple Point Charge)
Thr	Threonine
Trp	Tryptophan
Tyr	Tyrosine
Val	Valine
vdW	van Der Waals
VSW	Virtual Screening Workflow
WT	Wild Type
WT2	Second Simulation on Wild Type Dimer at Acidic pH
XP	Extra Precision

## 1. INTRODUCTION

Arylsulfatase A (ARSA) is a homo-octameric lysosomal hydrolase that catalyzes the break down reaction of the sulfate ester bonds. The main substrate of ARSA is cerebroside 3-sulfate. After the dimers form the octameric structure, the enzyme achieves full activity [1,2]. ARSA is an enzyme with several prerequisites for activity: folding, dimerization, segregation, octamerization, and hydration of FGL residue in the lysosome [3,4]. Any fault in these steps cause the inactivity of the enzyme, and the deficiency of ARSA emerges as Metachromatic Leukodystrophy Disease (MLD) which is a type of Lysosomal Storage Disease (LSD) [5–13].

MLD is an autosomal recessive inherited disease and occurs due to sulfatide accumulation and rapid demyelination of the neurons. It causes progressive destruction of the white matter leading to damage on intellectual functions, motor skills, and finally death [11-16]. Three types of MLD were determined: late infantile, juvenile and adult in the decreasing order of the severity. Although it is a rare disease it is frequently encountered in closed communities like Ashkenazi Jews [10,17]. A drug towards MLD rather than gene or enzyme replacement therapies has been missing since the discovery of the disease [18], [19], and discovery of a drug that recovers ARSA function is an area of interest [18–24].

Pharmacological chaperones (PC) are small molecules that act on targeted proteins. Unlike molecular chaperones, PC molecules are very selective [25-33]. Generally, the active sites of the proteins are targeted for PC binding but in some cases other regions like oligomerization or allosteric regions can also be targeted for PC binding [27]. They stabilize misfolded or partially folded mutant proteins by binding to the selected site. They help mutant proteins fold correctly or ensure the integrity of the oligomerized chains of the proteins, thus proteins can gain their functions and be segregated [25–33].

In this work, we aimed to find novel compounds that can be evaluated further as candidate pharmacological chaperones for two novel MLD causing Arylsulfatase A mutants: W318C and E307K+T391S [8,34]. We hypothesized that these mutations misfolding leading

to degradation in the ER or a hindrance in the oligomerization leading to inactivity. Binding a stabilizing agent can make the enzyme pass the ER control mechanism or enhance oligomerization. For this purpose, we performed docking and MD simulations to analyze the dynamics of apo WT, mutant, ligand bound WT and mutant structures in neutral and acidic environment. The results were compared and evaluated in the light of the trafficking process of lysosomal enzymes and functioning of ARSA.

This thesis is organized as follows: In the background part, information about the Sphingolipid metabolism, Arylsulfatase A enzyme (ARSA), trafficking of lysosomal enzymes as well as ARSA, Metachromatic Leukodystrophy Disease (MLD), Pharmacological Chaperones (PC), and previous treatments towards MLD is given. The steps that were followed in the docking, molecular dynamics simulations, and the analysis of the simulations are described in the Methods section. Results and Discussion section explains all the results from the computational work and the interpretation of the data in light of enzyme dynamics, docking calculations, and the current drug design efforts. The aim of the study and the last comments on overall results along with a few possible suggested improvements are summarized in the Conclusion and Recommendations section. Additional data which is not shown in the Results and Discussion part are given in the Appendix section.

## 2. THEORETICAL BACKGROUND

### 2.1. Sphingolipid Metabolism

Sphingolipids are a class of lipids with a backbone chain of sphingoid. They generally consist of 14 to 18 amino-alcohol chain and various additional side groups. The modifications of the side groups and the length of the acyl chains give rise to different type of molecules that serve in the signaling process or form the building blocks of essential biological structures such as myelin sheath. The central molecule is ceramide as shown in Figure 2.1. The molecules are generally modified by the enzymes to transform to each other but also there is an additional de novo ceramide synthesis pathway from palmitoyl CoA and serine [35–37].

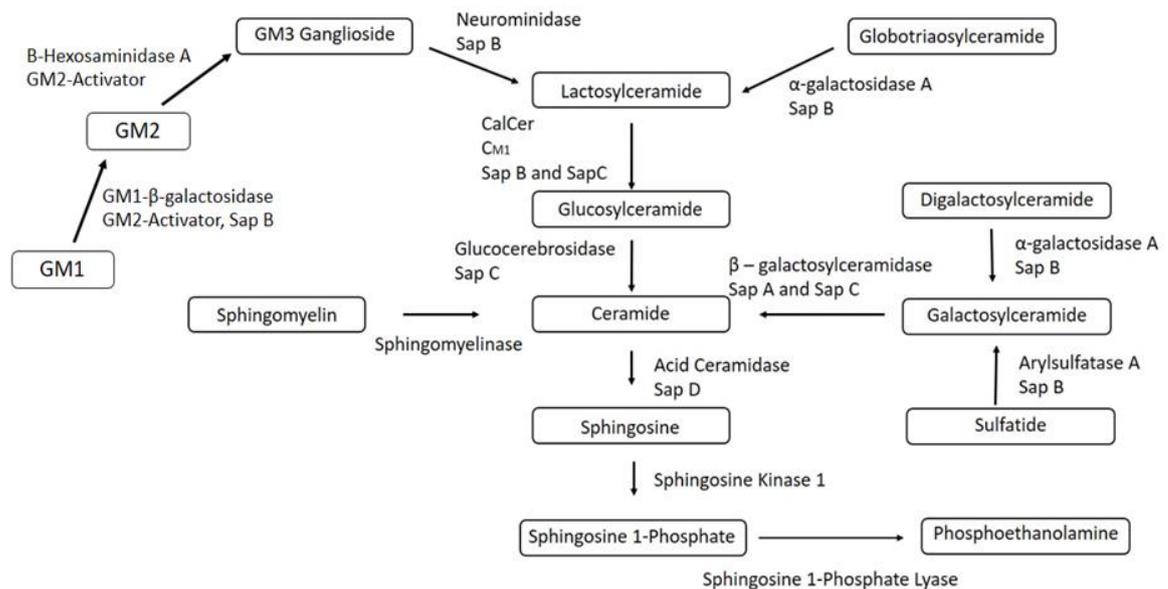


Figure 2.1. A part of the sphingolipid metabolism is shown.

In this metabolism, various enzymes take role in converting the sphingoid based compounds into one and other. These conversions are crucial for the survival and functioning of the mammals. Each compound in the metabolism has multiple unique roles from cell apoptosis to building the plasma membrane. Each step in this metabolism is vital such that

in terms of their deficiencies cause multiple lethal or severe diseases like cancer, Gaucher, MLD, Tay-Sachs, Niemann-Pick, and Fabry [40–42].

In Figure 2.1, Arylsulfatase A is shown at the right bottom part of the pathway. It has a role to convert cerebroside 3-sulfate into cerebroside and sulfate by breaking down the sulfate ester bond. Cerebroside has two main important roles. One, it forms the myelin sheath and the other is that it is one of the precursor molecules to form ceramide. Myelin sheath provides insulation for the neurons, as well as it forms the medium around the neurons which accelerates the impulse conduction. This is the points where the importance of ARSA comes to light. In the deficiency of this enzyme, a lysosomal storage disease (LSD) named as Metachromatic Leukodystrophy Disease (MLD) occurs [12,20,43]. This disease is a lethal neurodegenerative disease and it is described in detail under Title 2.4.

## 2.2. Arylsulfatase A (ARSA)

Arylsulfatase A (ARSA) is a lysosomal enzyme which functions as a homo-octamer. It comes from a highly-conserved sulfatase family and catalyzes the desulphurization reaction of cerebroside-3-sulfate. In case of deficiency of ARSA, Metachromatic Leukodystrophy Disease (MLD) occurs due to accumulation of sulfatides in the cells and demyelination of the neurons [3,8,13,15,36,44–47]. Three types of MLD have been identified namely; late-infantile, juvenile and adult as in decreasing order in severity. The incident of MLD is estimated as 1:40,000 [13,48].

The product, cerebroside is one of the main constituents of the myelin sheath around neurons. The wild type enzyme structure was determined in 1997 with a  $Mg^{2+}$  ion in its catalytic cavity along with NAG and NDG oligosaccharides [1,49]. Thereafter,  $Mg^{2+}$  ion was determined to be  $Ca^{2+}$  in human ARSA [50]. ARSA monomers have a hat-like shape and they form dimers in neutral pH whereas, formation of octamers is suggested to take place in acidic milieu depending on the protonation state of Glu424. The dimer-octamer equilibrium is controlled by the switch mechanism of Glu424 due to pH change. At neutral pH, Glu424 deprotonates and forms an intramolecular hydrogen bond between Gln460 which favors the dimeric form of ARSA. On the other hand, in acidic milieu (when the pH is below 6), Glu424 is protonated and forms an intermolecular hydrogen bond between

Phe398 which sustains the homo-octameric structure [4,51,52]. Arylsulfatase A has six cysteine knots (156-172, 161-168, 300-414, 489-502, 490-500, 493-499) and two free cysteines (38 and 294). These cysteine knots are important for structural integrity of the enzyme that they should always be included in the crystal structure. The most important residue is formyl glycine (FGL) 69 at the bottom of the active site pocket. It was found to be highly conserved and catalytically active residue among sulfatase family [1,53–57]. FGL69 becomes hydrated in the lysosome to perform a nucleophilic attack on the sulfate group of cerebroside 3-sulfate so that the sulfate group is detached from the substrate [45].

ARSA shows optimal activity when it forms a homo-octameric structure in acidic milieu while in neutral milieu it forms a homo-dimer. Dimerization is mediated by four loops in the monomer-monomer interface and certain hydrogen bonds between the residues His328-Pro42, Ser43-Ser432, and Tyr439-Thr408. These loops are loop 32-44, loop 286-300, loop 406-414, and loop 430-439 and the first two of them contain free cysteines which help to mediate the dimerization. Octamerization is mainly due to the hydrophobic interactions between the  $\alpha$ I helices (Pro450-Val470) as well as the hydrogen donation of Lys457 to Glu382 and Asp467 [1,50].

There are 10 different crystal structures related to and of Arylsulfatase A that were determined with X-Ray crystallography method in Protein Databank (PDB). The apo form wild type (WT) crystal structure that has been determined for the WT form (PDB ID: 1AUK) in 1998 [45,50,57,58]. Other structures are mutant and covalently bound inhibitor including structures. Seven of the crystal structures belong to ARSA with mutations on the catalytic residue FGL69 and  $\text{Ca}^{2+}$  ion in the active site cavity whereas, three of them belong to formyl glycine generating enzymes [1,50]. In this work, we used the apo form of WT structure without inhibitor and  $\text{Mg}^{2+}$  containing, 1AUK.

### **2.3. Trafficking of Arylsulfatase A**

The trafficking starts with the cleavage of the 18-residue signal peptide in ER during synthesis to translocate the enzyme into the lumen of ER [20,59,60]. The second step is the glycosylation of the enzyme in ER, which is read as a signal for correctly folded protein to leave ER. There are three glycosylation sites on ARSA which are Asn158, Asn184 and

Asn350. Phosphorylation is suggested to be sequential on these sites [21] and after glycosylation, due to modifications on the N-linked oligosaccharides the enzyme is carried to Golgi for phosphorylation where transmembrane glycoproteins (M6P) are bound to the enzyme on these modified oligosaccharide side chains [13, 21-23]. Each glycosylation site is glycosylated on ARSA but, the phosphorylation *in vivo* was found to be restricted on the oligosaccharides on Asn158 and Asn350. In addition to these, the salt bridge between Asp335-Arg370 was found to be important to shield and stabilize Asn350. In the previous simulations, glycan at the position Asn350 was determined to interact with Asp335-Arg370 salt bridge and nearby sites. The stability of Asp335-Arg370 loop and the presence of the salt bridge is significant for the enzyme trafficking because it protects the glycan bound Asn350 from the proteolytic attacks in lysosomes [24].

Recognition by phosphotransferase enzyme is critical for the translocation of ARSA. On the enzyme, there are certain sites/epitopes that were determined to have an important role in this step as well as the lysine rich surface [22]. Phosphorylation step initiates the transfer of the enzyme into a pre-lysosomal compartment which carries the enzyme into lysosome while providing an acidic environment for the detachment of the receptors and octamerization [20]. Misfolding or recognition site change due to missense mutations are the main reasons for this process to be interrupted leading to lysosomal storage diseases. The epitopes on ARSA were determined as Arg19, Met85-Gly86, Val93, Arg97, Ala165-Asp169, Met202-His206, Asp216, Ala239-Glu240, Ala256, Thr260, Ala287-Leu459, and Leu462-Ala507 [21]. Retention of the enzyme in ER was correlated with the disturbance in the formation of the epitopes forming late in synthesis [13]. These late forming epitopes are Ala165-Asp169, Met202-His206, Asp216, Ala239-Glu240, Ala287-Leu459 and, Leu462-Ala507 [13, 22]. Some of the missense mutations were listed along with their phenotypes in Table 2.1.

Table 2.1. Some of the mutations have been identified on ARSA and their corresponding phenotypes are listed. Late infantile type of MLD is the most severe type of MLD.

Mutations	MLD Type
C300F, E253K, G309S, S44P, L81P, W318C, E307K, A212P, C156X, P284S, E312D, A314T, S95N, S250Y, D335V, R370W, P377L, G245R, T274M, S295Y, G309S, P136L, S96L, S96F, G86D	Late Infantile
N99G, P425T, P426L, T408I, T286P, P135L, G301R, E312D, D255H, Y201C, E382K, R84Q, I179S, R370Q, T409I, P426L	Juvenile or Adult

The locations of these mutations and the epitope regions are shown on the 3D structure of ARSA in Figure 2.2.

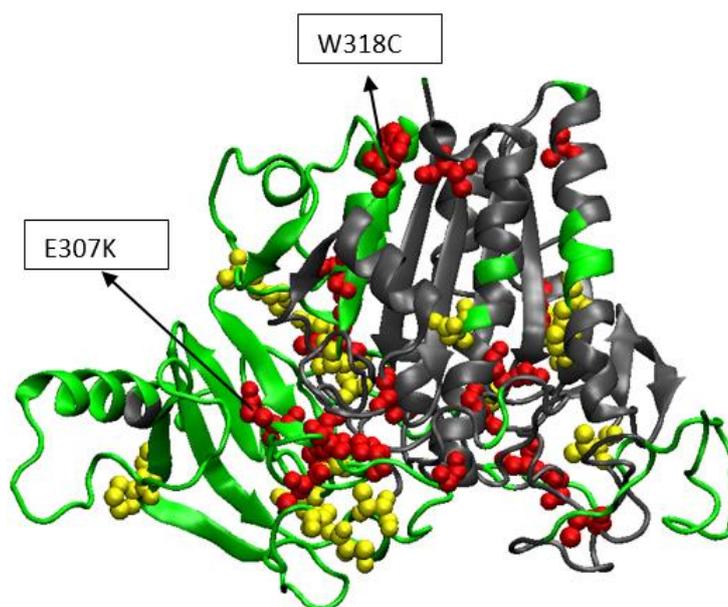


Figure 2.2. Late infantile type MLD causing mutations (red), juvenile or adult types of MLD causing mutations (yellow) and epitope residues (green) are shown on ARSA.

Large number of mutations were determined to cause MLD but few of them were explained with their inactivity mechanism in subcellular or molecular level. Some mutations have indirect effects on the oligomerization of ARSA. In other words, the expected effects of mutations may not be seen and other side effects of the mutations on enzyme activity,

oligomerization or trafficking can occur. For instance, as mentioned above C300F mutation on ARSA was expected to hinder the dimerization of the enzyme since C300-C414 disulfide bridge was determined to be important for the dimer stability, but the octamerization of ARSA was hindered [7, 59–61]. The location or the function of a residue can be misleading in terms of the inactivity mechanism.

#### **2.4. Metachromatic Leukodystrophy Disease (MLD)**

Metachromatic Leukodystrophy Disease (MLD) occurs in the case of deficiency of ARSA. The inactivity of ARSA due to mutations causes the accumulation of sulphatides in the lysosomes. This affects the myelin sheath formation because the product, cerebroside, of the hydrolysis reaction is one of the main constituents of the myelin sheath around neurons. It is a neurodegenerative disease that results from progressive destruction of white matter causing damage on intellectual functions and motor skills due to demyelination. Three types of MLD are known in the increasing order its severity: Late infantile, juvenile and adult. MLD is generally a fatal disease and is a subtype of Lysosomal Storage Diseases (LSD) [12,13,20,23,43,64,65].

More than 150 mutations on ARSA gene which are known to cause MLD. These can be deletions or frame shifts or missense mutations [8-17]. Mutations have various effects on the enzyme function, structure and stability. Most of the missense mutations were identified to cause instability on the enzyme structure and cause ARSA being degraded in the lysosome or pre-lysosomal compartment [8,18] whereas a few were identified to cause retention of the enzyme in ER before secretion in the trafficking process [13]. Two of the most common mutations on ARSA are T391S polymorphism and N350S leading to the loss of N-glycosylation. The frequency of other mutations changes with changing gene pools [19].

When mapped on the 3D structure, we see that severe type of MLD causing mutations are generally located on the top portion of the enzyme or around the active site cavity at the bottom center of the protein. Three of both severe type and juvenile or adult type of MLD causing mutations are located on the dimerization site of ARSA (bottom left part in Figure 2.2). In addition to these two mutations, one missense and one frameshift on 464th and 467th amino acids respectively, are severe type of MLD causing mutations (not shown in

Figure 2.2). They are located on the octamerization helix (Pro450 – Val470) thus they may cause a hindrance on octamerization leading to inactivity. On the other hand, although it is not located on the octamerization helix, P426L mutation is known to prevent ARSA from forming homo-octamers due to its location close to Glu424 which acts as a switch residue for octamerization [18].

## **2.5. Pharmacological Chaperones (PCs)**

As an emerging treatment for the diseases that are found to result from mutations causing misfolded proteins, pharmacological chaperones are the protein-specific small molecules that stabilize the misfolded proteins and help them to pass the ER control mechanism. Generally, the binding in ER to the active site of the enzymes are aimed, but in the case of neurodegenerative diseases that are emerged from the inactivity of corresponding enzymes like lysosomal storage diseases or Alzheimer, trafficking and oligomerization steps of the enzymes are also considered for selecting the binding sites according to the reasons of the inactivity. The binding of the small molecules makes the misfolded enzymes or the ones that cannot oligomerize more stable in terms of having a minimum in the energy funnel. They act like glue and enhance the formation of favorable interactions within the proteins [29,33,34].

## **2.6. Previous Treatments for MLD**

There have been administered various treatments like hematopoietic stem cell transplantation, bone marrow transplantation, intrathecal enzyme replacement therapy, and gene therapies for MLD patients. Small molecules have been considered as enzyme enhancers for a certain time, alas there is not any chaperone candidate yet. Molecular chaperones have been considered as a good alternative to other treatments. A subset of molecular chaperones is pharmacological chaperones which work selectively on certain enzymes to enhance their folding in ER but one of the most challenging issue in developing drugs from pharmacological chaperones for any lysosomal storage disease (LSD) is crossing the blood brain barriers (BBB) [64,65]. Because of this difficulty, a novel compound has not been proposed as a drug candidate yet [18,19,22].

### 3. METHODS

The structure of wild type form of ARSA was retrieved from Protein Data Bank (PDB) (PDBID: 1AUK) [1]. It is the wild type (WT) apo form of ARSA. The structure for the mutants is not available therefore the mutant residues were changed by using “Mutate Residue” tool of Maestro [68]. Firstly, the free energy changes were calculated by using the FoldX program [69]. Then the protein forms were prepared by using Protein Preparation Wizard of Schrödinger [70]. Ligands were generated and prepared before docking [65,66]. Minimized ligands were docked into the active site [73]. For ligand preparation, protein preparation, docking, molecular dynamics (MD) simulation and analysis steps, 2D-Sketcher, Protein Preparation Wizard, Prime, Epik, Ligprep, GLIDE, Desmond, Simulation Event analysis and, Qikprop programs in Maestro of Schrodinger were used. Salt bridges, and hydrogen bonds were identified and distance measurements were performed using VMD as well as the solvent accessible surface calculations and cross-correlation calculations [68,69].

#### 3.1. Structure Similarity Search and Residue Conservation Calculations

Protein Databases were scanned to find other proteins that have similar secondary structures to ARSA. For this purpose, PDBeFOLD, Dali and iSARST servers were used. Conservation calculations were performed by ConSurf server and all the amino acids on ARSA were scored according to their evolutionary conservation profiles. The FASTA sequence of the crystal structure (PDB: 1AUK) of ARSA was uploaded into the servers for the calculations [1].

#### 3.2. Protein Preparation

The structures were prepared using Protein Preparation Wizard of Schrödinger’s Maestro Suite [72]. Coordinates of the missing residues which form the loop before  $\alpha$ helix-I (Gly444, Val445, Ala446, Gly447) and the missing atoms of the residues Arg19, Glu131, Arg241 and Gln454 were predicted using the same utility. The structure was optimized and then minimized before building FGL69 residue. The structure was optimized by sampling water orientations and minimizing hydrogens of altered species. A restrained minimization

step was performed by using OPLS2005 force field [70]. Additionally, waters beyond 5 Å of HET groups were deleted. After relaxing the structure, the side chain of FGL69 was built with the Build tool of Maestro [72]. After constructing the missing atoms and residues the structure was optimized again and finally restrained minimization was performed as the last step in protein preparation. This preparation step was performed for all simulations with the same options. Mutants were prepared by “Mutate Residue” option in Maestro [68]. After the residues were changed, optimization and minimization steps were performed one more time.

MD simulations in the acidic milieu were performed for two states of FGL residue: diol and nonhydrated. In the diol form beta carbon of the FGL residue has two -OH groups whereas, there are one H atom and one O atom with double bonds in its nonhydrated form. Eight different forms of ARSA were prepared using the same options indicated above only with the difference of the state of FGL in the building step. We also included the mutant form of ARSA only with the mutation of E307K in our MD simulations in acidic milieu [72].

Dimer structures were prepared by taking the biological assembly from PDB and deleting the other three dimers in the entry. The monomers in the dimer structures were prepared separately as described above by selecting the same options and following the same steps. After preparing each monomer, both monomers were selected and put into a water box for the system preparation as described under the MD Simulations title of Methods

### **3.3. Ligand Preparation**

Ligands were generated by 2D-Sketcher menu of Maestro [72]. After creating 3D entries using “Convert To 3D” option of the 2D Workspace tab, 3D entries were generated in the workspace. LigPrep was used to generate various conformations of the ligands with correct chirality and low energy. Ligand preparation was performed at neutral pH with the options of retained specific chiralities and no included metal binding sites [71].

### 3.4. Receptor Grid Generation

Grid for ARSA active site was generated by selecting the centroid of the six residues out of the active site residues, Asp29, Asp30, His125, His229, Asp281, and Asn282. Only the active site residues if listed by the program, and the residues which are 6 Å distant to the active site residues could rotate. Docking calculations that were performed on mutant structures were done by selecting the coordinates from the last frames of the simulations. They were taken and the grid was generated on them [73].

### 3.5. Screening and Docking Calculations

Virtual Screening Workflow (VSW) is a method for screening databases with large number of molecules, after the target and database preparation are done. VSW includes three stages namely; High Throughput Virtual Screening (HTVS), Standard Precision Docking (SP Docking), and Extra Precision Docking (XP Docking). The top-ranking results of XP Docking step of VSW were taken and used for E-Pharmacophore modeling which is described in Section 3.7. The settings were adjusted for HTVS as keeping 12,000 of best compound, performing post-docking minimization, docking flexibly, and retaining all states. The settings were for SP Docking stage is as keeping 10% of best compounds, docking flexibly, and retaining all good scoring states. For the last stage, XP Docking, the settings are as keeping 10% of the best compounds, docking flexibly, and retaining only best scoring states [76].

Direct docking calculations were performed by GLIDE in Maestro. The coordinates from the last snapshot of the MD simulations of WT and W318C mutant were taken as the input structures for the docking. After generating the grid Extra Precision (XP) docking and Induced Fit Docking (IFD) protocols were performed by GLIDE [71,75,76]. Ligands with more than 500 atoms or 100 rotatable bonds were excluded. Nitrogen inversions and ring conformations were sampled and Epik state penalties to docking scores were added. No constraints were used. Post-docking minimization was performed and strain correction terms were applied [73].

### 3.6. Clustering the Ligands

There are the known 653 bioactive molecules of ARSA which were retrieved from ChEMBL database . They were taken into Canvas tool for creating the clusters with common chemical properties [79]. For this purpose, 2D-fingerprints of all dataset were created, then from these fingerprints clusters were developed. Hierarchical Clustering protocol was performed in Canvas on 653 molecules. The similarity metric was set to Tanimato similarity for this protocol [80].

### 3.7. Pharmacophore Modeling

In drug design, pharmacophore modeling is a crucial step for discovering the main features of molecules that can be drug candidates. “Pharmacophore” means that a part or a substructure of a molecule that is responsible for a biological or pharmacological interaction. These parts define the features that is necessary for optimal interactions between the ligand and the receptor. There are two methods under this title that have been developed namely, ligand-based pharmacophore modeling and structure-based pharmacophore modeling [81]. In the pursuit of finding a pharmacological chaperone for our mutant ARSA forms we applied both methods to develop a common model for candidate chaperones. Docking calculations for both methods were performed on WT and W318C mutant ARSA monomers at neutral pH because the binding of the chaperone candidate was aimed to be in ER [73].

#### 3.7.1. Ligand-based Pharmacophore Modeling

Ligand-baes pharmacophore model of ARSA was developed by using the known 653 bioactive molecules which were retrieved from ChEMBL database [74,75]. SMILES codes of the molecules were used in 2D Sketcher tool in Maestro to generate the 2D structures of the molecules [72] and prepared as previously described in Section 3.2. After preparation, 650 of them were ready to be used in QSAR model development. The IC<sub>50</sub> values of the molecules were not reported, but since their potencies were reported and generally the potency is considered to be equivalent of IC<sub>50</sub>/EC<sub>50</sub>/K<sub>i</sub> of the compounds [76–78]. Reported potency values from the database were used in place of IC<sub>50</sub> values, and converted to corresponding pIC<sub>50</sub> values by using the formula ( $pIC_{50} = -\log[IC_{50}]$ ). The activity

threshold was set according to the pIC<sub>50</sub> values of the molecules such that it divides the active and inactive ligands distinctively. For instance, if a molecule is active, its pIC<sub>50</sub> value should be greater than a certain value like 5 and if a molecule is inactive, then its value should be less than 3. The remaining molecules that are in between these threshold values were set as moderately active. To develop a training and a test set, 75% of all molecules were selected into the training set [87]. The pharmacophore hypothesis and QSAR model were developed by using the training set. Among six pharmacophore features (Hydrogen bond acceptor-A, hydrogen bond donor-D, hydrophobic group-H, negatively charged group-N, positively charged group-P, and aromatic ring-R), site points were selected according to their survival scores. In this work, the highest survival score having three sites were selected for structure-based pharmacophore modeling. For developing a hypothesis Q<sub>2</sub> and R<sub>2</sub> results of the models should be higher than the commonly accepted threshold values 0.6 and 0.8 respectively. The developed hypothesis was used in generating 3D-QSAR model according to the training set and validated test set. Partial Least Square (PLS) method is used in 3D-QSAR modeling [88]. This procedure was repeated for four times changing the test and training sets and the best one was selected for QSAR modeling for ARSA.

### 3.7.2. Structure-based Pharmacophore Modeling

This method uses the binding site and how the pre-docked ligands were docked to the binding site. In this method, we do not need to know the potency values of the molecules but their binding poses and the binding site. For ARSA, the VSW results of four sets of molecule databases were used namely; OTAVA library, ChEMBL database, Schrodinger's fragment library which is composed of 667 unique small fragments, and a ZINC dataset with 4 million compounds. In addition to the VSW results of the databases, selected clusters from the known bioactive molecules of ARSA which were retrieved from ChEMBL database, were used in E-Pharmacophore script in Maestro program [72]. After clustering, ten clusters were found with molecule numbers greater than 20. The biggest cluster with 38 molecules was selected for docking with XP protocol, then its descriptor files were used in E-Pharmacophore modeling. In this work, pharmacophore hypotheses were generated from seven pharmacophore sites by selecting the top ranking three of them.

### 3.8. Analysis of Docking Results

Qikprop was used to calculate the physicochemical properties of the ligands such as predicted blood brain barrier coefficient (QPlogBB), predicted octanol-water partition coefficient (QPlogPo/w), Glide Ligand Efficiency (LE), and solvent exposed surface area (SASA). These properties have predefined threshold values given as:

$$\begin{aligned} \text{QPlogBB:} & \quad -3.0 - 1.2 \\ \text{QPlogPo/w:} & \quad -2.0 - 6.5 \\ \text{SASA:} & \quad 300 - 1000 \end{aligned}$$

Glide LE is calculated from the Glide Score (GScore) of the docked compounds as given below. It is given by the program, GLIDE [73].

$$\text{Glide LE} = \frac{\text{GlideScore (GScore)}}{\text{Number of Heavy Atoms}} \quad (3.1)$$

Other physicochemical properties namely; free energy of binding ( $\Delta G$ ), dissociation constant ( $K_i$ ), negative logarithm of dissociation constant ( $p(K_i)$ ), ligand lipophilicity efficiency (LLE), binding efficiency index (BEI), and surface-binding efficiency index (SEI) were calculated using the equations :

$$\Delta G = (\text{number of nonH atoms}) \times (\text{Glide LE}) \quad (3.2)$$

$$K_i = e^{\frac{\Delta G}{RT}} \quad (3.3)$$

$$p(K_i) = -\log_{10}(K_i) \quad (3.4)$$

$$BEI = \frac{p(K_i) \times 1000}{MW} \quad (3.5)$$

$$SEI = \frac{p(K_i) \times 100}{PSA} \quad (3.6)$$

$$LLE = p(K_i) - \log P \quad (3.7)$$

The values of these properties help to define the drug-likeness of a compound. LE is the normalization of the affinities and it is given by GLIDE after docking. Dissociation constant is preferred to be as small as possible for strong binding. Accordingly, BEI and SEI values will be higher which also means stronger binding. Ligand lipophilicity efficiency (LLE) combines in vitro potency and lipophilicity. A positive value is preferred for drug-likeness [83,84,85]. logP is denoted with QPlogPo/w in Maestro and has the range -2.0-6.5 [26,27].

The binding free energy of the ligands are the main output of docking calculations. It is calculated as Glide Gscore (Gscore), which is a weighted sum of several properties, by the equation as shown below:

$$\begin{aligned} \text{Glide Gscore} = & 0.05 \times \text{vdW} + 0.15 \times \text{Coul} + \text{Lipo} + \text{Hbond} + \text{Metal} \quad (3.8) \\ & + \text{Rewards} + \text{RotB} + \text{Site} \end{aligned}$$

where “vdW” is the van der Waals energy, “Coul” is the Coulomb energy, “Lipo” is the lipophilic term derived from hydrophobic grid potential, “Hbond” is for hydrogen-bonding, “Metal” is metal-binding term considering only the interactions with anionic or highly polar acceptor atoms, “Rewards” stands for all other terms that are not mentioned, “RotB” is the penalty for freezing rotatable bonds, and “Site” is the polar interactions in the active site [73].

### 3.9. Molecular Dynamics (MD) Simulations

The systems were built with the System Builder menu in Desmond program of Schrödinger. Predefined SPC waters were used to build an orthorhombic box with sizes of 10 Å. The systems were neutralized by adding Na<sup>+</sup> and Cl<sup>-</sup> ions. The number of ions were calculated automatically by the program for each system. 0.15 M of salt was also added into the systems. Molecular dynamics simulations were performed by Desmond with OPLS2005 force field. Simulation time was set to 50 ns. Ensemble class was chosen as NPT and the temperature was set to 310 K for each simulation. Both thermostat and barostat methods were chosen as Langevin. A relaxation protocol of 2.0 ps was applied and isotropic coupling style was chosen in the barostat tab. For Coulombic interactions, cutoff was chosen for short range method and the radius of the cutoff was set to 9.0 Å. On the other hand, smooth particle

mesh Ewald was chosen for the long-range method and its Ewald tolerance was set to 1e-09. In all simulations, systems were relaxed before MD simulations and no positional restraints was applied. All the preparation and simulation steps were performed by using PROPKA for neutral and acidic pH [70,90,91].

### 3.10. Cross-Correlation Calculations

Trajectories of the MD simulations were used for cross-correlation calculations. Below equation was used to calculate the cross-correlation between the residues [92,93].

$$C_{ij} = \frac{\langle \Delta \vec{r}_i \cdot \Delta \vec{r}_j \rangle}{\sqrt{\langle \Delta r_i^2 \rangle \langle \Delta r_j^2 \rangle}} \quad (3.9)$$

R program and a tcl script for VMD were used to perform the calculation. The results were compared and found to be exact. Accordingly the tcl script was used again to connect the alpha carbon atoms of the correlated residues with a higher/lower score than 0.5/-0.5. Because of the computational cost due to the large trajectory files, trajectory files were reduced from having 10419 frames to 2084 frames by taking stride as five when saving the trajectories. The R script and the tcl script which were used was provided in the Appendix section [75], [96].

### 3.11. RMSD and RMSF Calculations

Root Mean Square Deviation (RMSD) and Root Mean Square Fluctuation (RMSF) calculations of the coordinates from all simulations were calculated using Simulation Event Analysis (SEA) and Simulation Interactions Diagram (SID) tools of Schrödinger's Desmond in Maestro GUI. These tools took the information of the coordinates of all atoms for each time step and both tools use the same equations to calculate the RMSD and RMSF values from the coordinates, given as below [92]:

$$RMSD_x = \sqrt{\frac{1}{N} \sum_{i=1}^N (r'_i(t_x) - r'_i(t_{ref}))^2} \quad (3.10)$$

$$RMSF_i = \sqrt{\frac{1}{T} \sum_{t=1}^T \langle (r'_i(t) - r'_i(t_{ref}))^2 \rangle} \quad (3.11)$$

Where N denotes the number of the atoms in the selection,  $t_{ref}$  is the reference time (the first frame of the simulation in most of the cases), and  $r'$  is the position (coordinates) of the selected atom in the current frame,  $x$ , where it is recorded at time  $t_x$ . The  $r'$  is obtained after superimposing the coordinates on the reference frame. In the equation for RMSF calculation, T denotes the number of the frames (time steps or time) in the simulation, while  $r_i$  is the position of the  $i$ th atom selection (e.g. carbon alpha of  $i$ th residue), and  $r'$  is the position of the  $i$ th atom after superposition on the reference position.

ARSA has nine different regions that are considered as specific for the enzyme activity namely: active site, four dimerization sites, octamerization site and three glycosylation sites. In this work, we used six different sites which are active site, three dimerization regions, octamerization helix and one of the glycosylation sites. Dimerization region was divided into three regions namely, dimerization1, 2 and 3 including the residues Asp29-Thr46, Thr286-Cys300, and Ala404-Gly447 respectively.

B-factor values of the residues were calculated by using Simulation Event Analysis tool in Maestro of Schrodinger [92]. The experimental B-factor values were taken from the PDB file of ARSA and plotted along with the calculated B-factor values in Excel.

### 3.12. Secondary Structure Calculations

Simulation Interaction Diagram (SID) in Maestro suite of Schrodinger was used for secondary structure analyses. The P-SSE tab in SID gives the secondary structure state of the protein as a function of residue number and time.

### 3.13. Binding Free Energy Calculations

MM-GBSA (Molecular Mechanics - Generalized Born Solvent Accessible Surface Area) is the method that was used to calculate the binding free energy terms and the total binding free energy of the ligand throughout the MD simulation on a ligand bound protein. Binding free energy is calculated as shown below [97]:

$$\text{MMGBSA } dG \text{ Bind} = \text{Complex} - \text{Receptor} - \text{Ligand} \quad (3.12)$$

There are different types of energy terms that Prime script in Maestro GUI calculates and gives as results like Coulomb, Covalent, vdW, Lipophilic energy, Generalized Born electrostatic solvation energy, Total energy (Prime energy), Hydrogen-bonding energy, Pi-pi packing energy, and Self-contact correction. These help us to understand the behaviors of the ligands during the simulation, and find out whether the binding of the corresponding ligand is intended [93].

### 3.14. Solvent Accessible Surface Area (SASA) Calculations

The SASA calculations were performed by using a tcl script for VMD. SASA of the residues for each frame of the simulations were calculated in a loop from the first frame to the last frame of the simulations by using “measure sasa 1.4” command. 1.4 is the probe radius in Å. The individual SASA values (in angstrom squared) is saved and then plotted to see the trend of the protein SASA [73,96–98].

### 3.15. Average Structure Calculations

The average structure of the ARSA forms were calculated by using a tcl script in VMD’s Tk Console. For this purpose, we selected the time frames of the trajectories of each simulation the time after the corresponding backbone RMSD profile reach a plateau. We loaded the coordinate file and the selected time frames from the trajectories of the simulations and then we used the tcl script. It takes the coordinate from each time frame and then takes the average of those. Finally it generates a PDB file of that average form. We used

those average forms in our RMSD calculations to see the overall structural differences between the WT and mutant forms after their backbone reach equilibrium.

### **3.16. Possible Salt Bridges and Hydrogen Bond Occupancy Calculations**

The possible salt bridges and the percent occupancies of the hydrogen bonds between the residue pairs as well as between the ligands and the residues were calculated by using the Salt Bridges and Hydrogen Bonds tools in VMD. Possible salt bridges were determined for the all frames of the simulations with the options of selection as the protein, oxygen-nitrogen cut-off distance as 3.2 Å while selecting the side-chain Center of Mass (COM) distance cut-off as none. For the hydrogen bond occupancy calculations, donor-acceptor distance was selected as 3.2 Å while the angle cut-off was selected as 20°. The hydrogen bonding calculations were performed for all frames of the simulations and the detailed information was calculated for residue pairs in terms of percentages.

### **3.17. Calculations of Dihedral Angles**

Dihedral angles (phi and psi) of the residues were calculated by using both VMD program's timeline tool and Simulation Interaction Diagram (SID) tool of Maestro GUI. The trajectories were taken and all the angles were calculated for each time frames in both programs [75,92].

## 4. RESULTS AND DISCUSSION

Firstly, the conservation scores and the energy change calculations on the protein structure due to the mutations of the residues and protein similarity search for ARSA were performed. ARSA is a unique protein such that the structural similarity search did not give any other protein than the Arylsulfatase A proteins in other organisms. The results were discussed in detail in Section 4.1.

MD simulations were performed on monomer and dimer forms of WT, T274M, C300F, P377L W318C mutants, E307K single mutant, ligand and substrate bound W318C mutant, and E307K+T391S mutant types of ARSA at neutral and acidic pH (FGL in its nonhydrated state) for 50 ns. Later, docking calculations were performed at both pH values based on the aim of the ligand binding. Lastly, the results of the apo form simulations were discussed first, then the results of the docking calculations, and lastly the the results of the MD simulations on the ligand bound forms were discussed.

### 4.1. Residue Conservation and Similarity with Other Proteins

Arylsulfatase A is a unique protein such that the structural similarity search did not give any other protein than the Arylsulfatase A proteins in other organisms different than human unless the similarity percentage falls to 13 %. In addition to this, the conservation scores of the residues showed that Glu307 has the highest conservation score, which is 9, while Trp318 has an intermediate score, 5. Thr391 has the lowest score, 1 which is expected because the polymorphism mutation T391S is common among the mutations on ARSA. We also performed free energy calculations on the monomer, dimer and octamer structures of the mutant ARSA forms and summarized our results as in Table 4.1.

Table 4.1. Free energy calculation results (kcal/mol) from FoldX Suite.

	<b>Mutations</b>	<b>Total Energy</b>	<b>Van der Wals</b>	<b>Solvation Hydrophobic</b>	<b>VdW Clashes</b>
<b>Monomer</b>	T391S	0.356	0.375	0.696	-0.234
	T391S + E307K	13.1	-0.173	-0.381	3.61
	W318C	1.02	1.62	2.68	-0.07
<b>Dimer</b>	T391S	3.99	25.04	35.3	-0.161
	T391S + E307K	22.89	23.69	33.4	3.473
	W318C	4.43	17.9	25.28	0.002
<b>Octamer</b>	T391S	30.87	100.5	141.1	1.47
	T391S + E307K	92.74	95.7	132.2	18.4
	W318C	30.89	72.83	103.1	0.67

Free energy calculations show that our mutations have a significant effect on the free energy of the protein in all forms, thus to better understand the molecular effects of the mutations on the dynamics and the structure of ARSA we analyzed our MD simulations as follows.

## 4.2. MD Simulations on Monomeric ARSA at Neutral pH

MD simulations were performed on monomer forms of WT, T274M, P377L W318C, and E307K+T391S mutant types of ARSA at neutral pH (FGL in its nonhydrated state) for 50 ns. The analysis were done as described previously in Methods and discussed in detail as following.

### 4.2.1. RMSD and RMSF Calculations

Three forms of ARSA (WT, W318C mutant, and E307K+T391S mutant) were prepared and put in solvent boxes with specified size. All structures include the nonhydrated form of FGL due to the neutral environment. We aimed understand the effects of the two novel MLD causing mutations on the dynamics and the structure of the enzyme, thus we performed MD simulations on WT, E307K+T319S mutant, and W318C mutant ARSA structures for 50 ns.

Root mean square deviation (RMSD) and Root Mean Square Fluctuation (RMSF) calculations were done by taking the first frame structures of the corresponding simulations as the reference structures. These calculations were done to determine the structural and dynamic instabilities or deviations of the overall structures and specific regions on the proteins during the simulation. Backbone RMSD calculations show (Figure 4.1) the deviation of ARSA backbone coordinates from the initial coordinates throughout the simulations.

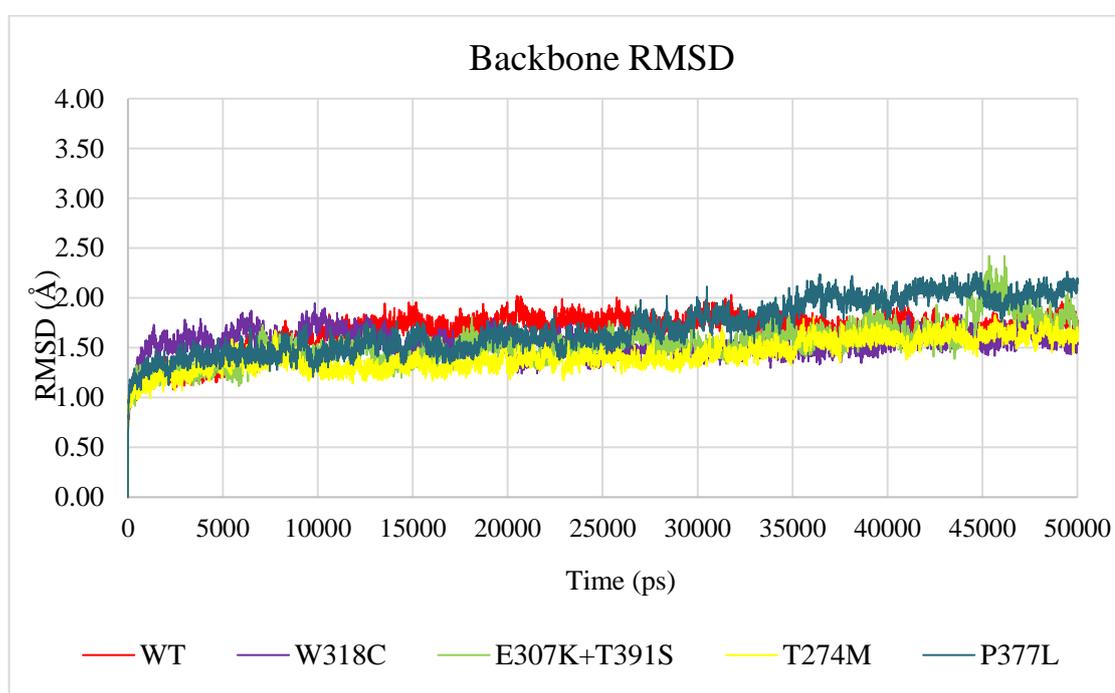


Figure 4.1. The backbone RMSD as a function of time for the simulations on monomeric ARSA at neutral pH.

RMSD profile remains stable around 1.65 Å without any sudden increase or decrease for simulations on WT ARSA. RMSD profile of W318C stays lower at around 1.5 Å and then starts to increase after 36 ns whereas, the RMSD profile of E307K+T391S shows a lower trend but it continually increases during 50 ns. The sudden increase in the RMSD profile of E307K+T391S mutant around 45 ns can be linked to the movement of the loop (Leu441-Ala448) located on the one end of helix I. The divalent cation in the active site maintained its position in the cavity. The active site configuration and dynamics were not affected by the mutations such that RMSD of the active site coordinates to the initial

structure was less than 0.2 Å for all three simulations. The last snapshots of the simulations are superimposed in Figure 4.2.

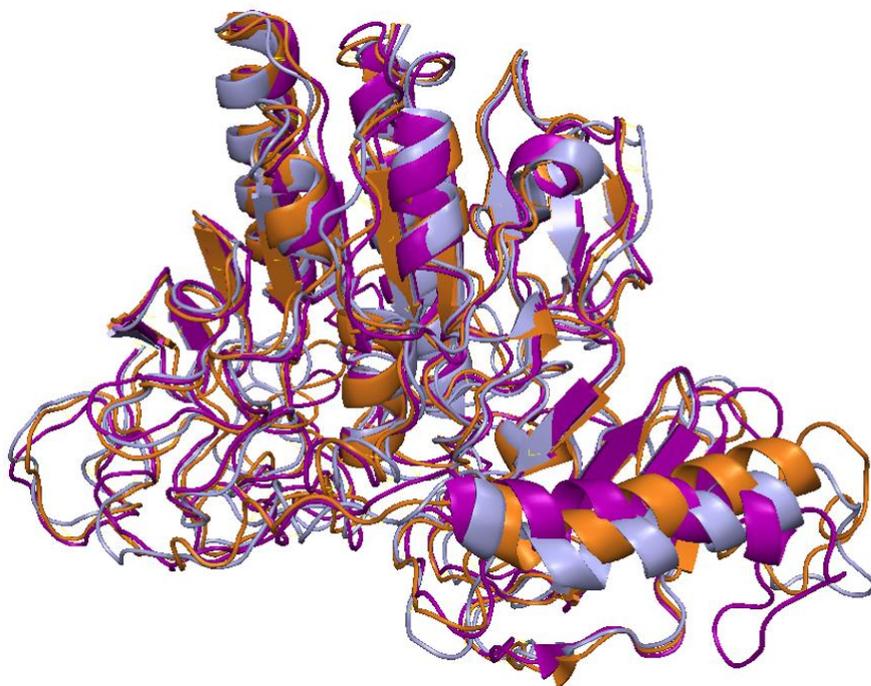


Figure 4.2. Last frames from the MD simulations of the structures of WT, W318C and E307K+T391S mutant forms of ARSA were aligned.

In Figure 4.2, the color codes are as follows: WT is blue, W318C mutant is orange, and E307K+T391S mutant is purple. The octamerization helices are on the right and the left end of the helices differ in conformation. W318C mutant preserves the helical secondary structure of the left end of its helix whereas, the different alignment of the octamerization helix of E307K+T391S mutant can be clearly seen.

Structural alignment showed that conformational changes in E307K+T391S simulation occurred on the entire and especially the Val470 end of the octamerization helix. The secondary structure maps that are shown on the following sections also validated that a more severe change in the secondary structure of this region occurs in the E307K+T391S mutant simulation. This shows that especially E307K mutation may cause a hindrance in the octamerization process of ARSA which leads to severe type of MLD.

**4.2.1.1. RMSF and Average RMSD Calculations – Overall Structure.** To understand the regional effects of the mutations, RMSD calculations for the other important sites on ARSA were done as well the average RMSD values for the whole protein and its specific regions were calculated and summarized in Table 4.2. The important sites of ARSA is shown in Figure 4.3. The results showed that significant differences between the RMSD profiles of dimerization1 and dimerization2 loops present whereas, active site RMSD profiles showed no significant change. As previously reported [102], the stability of the glycosylation site Asn350 and the surrounding region, Asp335-Arg370, is very crucial for the enzyme trafficking. The calculations on the structure along with the RMSF calculations for Asn350 showed that mutations or the presence of a ligand in the active site do not affect the rigidity of Asp335-Arg370 loop. The average distances of the  $\alpha$ -carbons of the mutations E307K and W318C to the  $\alpha$ -carbon of Asn350 are 32.6 Å and 28.6 Å respectively thus, no direct effect of the mutations on glycosylation region is expected. Results of the RMSF calculations as in Figure A.1 showed that mutations caused Val445-Gly447 region to have high RMSF values. This region was missing in the crystal structure and built later in protein preparation step. Its conformational change was linked to the jump in the backbone RMSD trend of E307K+T391S mutant. B-factors were calculated and plotted in Figure A.1 with the experimental B-factor values to validate our findings with the experimental data.

Table 4.2. Average and standard deviation values of the RMSD profiles of the backbone, dimerization loops, and octamerization helix.

<b>Backbone RMSD (Å)</b>				<b>Dimerization1 RMSD (Å)</b>			
	WT	W318C	E307K+T391S		WT	W318C	E307K+T391S
<b>Avg</b>	1.66	1.54	1.54	<b>Avg</b>	1.79	2.06	2.85
<b>±Std</b>	0.18	0.12	0.20	<b>±Std</b>	0.32	0.73	0.36
<b>Active Site RMSD (Å)</b>				<b>Dimerization2 RMSD (Å)</b>			
	WT	W318C	E307K+T391S		WT	W318C	E307K+T391S
<b>Avg</b>	1.44	1.27	1.23	<b>Avg</b>	1.37	2.00	1.57
<b>±Std</b>	0.20	0.17	0.14	<b>±Std</b>	0.27	0.36	0.20
<b>Octamerization Helix RMSD (Å)</b>				<b>Dimerization3 RMSD (Å)</b>			
	WT	W318C	E307K+T391S		WT	W318C	E307K+T391S
<b>Avg</b>	2.10	2.60	2.86	<b>Avg</b>	2.10	2.60	2.86
<b>±Std</b>	0.22	0.40	0.36	<b>±Std</b>	0.22	0.40	0.36

In Table 4.2, average and standard deviation values of the RMSD profiles of the backbone, dimerization1, dimerization2, dimerization3 loops, and octamerization helix are listed. The results indicate that the significant changes occur on the dimerization1 region (Asp29-Thr46) of E307K+T391S mutant form. Octamerization helix region of E307K+T391S mutant has also a higher average RMSD value than the WT and W318C mutant forms. Dimerization2 region of W318C has the highest average RMSD value meaning that W318C mutation may be preventing ARSA monomers from forming the dimer. These results suggest that especially the addition of E307K mutant to the T391S polymorphism may cause conformational changes leading a hindrance in the dimerization and consequently the octamerization whereas, W318C mutation may cause the retention of the enzyme in ER or a hindrance in the dimerization or octamerization of ARSA. But the trends and fluctuations of the RMSD profiles of the specific regions should be analyzed in detail before coming to a conclusion. The residues that belong to the important sites of ARSA as in Figure 4.3 is listed in Table 4.4.

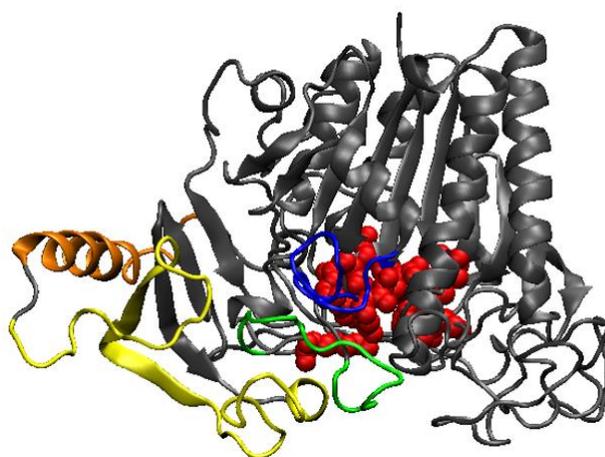


Figure 4.3. Apo WT ARSA structure and its important regions.

In Figure 4.3, Active site is in red, while dimerization1, 2 and 3 regions are in blue, green, and yellow respectively. Octamerization helix is orange. The coordinates of the residues that belong to the important regions of ARSA were used in the RMSD calculations.

In addition to the results shown above, the average structure of each form was calculated and generated after their backbone RMSD profiles reach a plateau. These average structures for the monomers were aligned based on the protein backbone and the RMSD values were calculated for the mutant forms, based on the average protein structure of WT ARSA as shown in Table 4.3. The backbone alignment and after RMSD calculation were performed based on WT structure, thus WT has 0.00 RMSD.

Table 4.3. RMSD results after the protein backbone alignment of the average structures of all forms of ARSA monomers at neutral pH.

<b>Monomer Type</b>	<b>RMSD (Å)</b>
W318C mutant	1.33
E307K+T391S mutant	1.39
P377L mutant	1.19
T274M mutant	1.17

As seen from Table 4.3 the overall protein structures do not show significant differences after reaching equilibrium, but the dimerization regions of the mutant monomer forms show structural changes. Due to these results, we calculated the dimer RMSD values based on their average structures in the same manner by hypothesizing that the mutations may be affecting the dimer structure. The results are discussed in Sections 4.8 and 4.9. We selected a big region for dimerization3 region for the sake of computational convenience instead of dividing the region as determined [1], we simply take all the residues at the dimer interface to understand the effects of the each residue dynamics on the region.

Table 4.4. Structurally important ARSA residues.

<b>Sites</b>	<b>Residues</b>
Active Site	Asp29, Asp30, FGL69, Asp281, Asn282, His125, His229, Lys123, Lys302, Ser150
Dimerization1	Asp29-Thr46
Dimerization2	Thr286-Cys300
Dimerization3	Ala404-Gly447
Octamerization Helix	Pro450-Val470
Glycosylation	Asn158, Asn184, and Asn350

The RMSF calculations are needed as well as the RMSD calculations to determine any fluctuations of the residues throughout the simulation for gaining a better understanding on the possible effects of their conformational changes on the overall stability, and the movements of the specific regions of the enzyme. ARSA works as an oligomer and goes through post translational modifications thus, movements of the residues on these specific regions are crucial to determine possible obstacles in its folding on the dynamics, oligomerization and trafficking. For this reason, we performed RMSF calculations and discussed the results in the light of our MD trajectories and the information of the trafficking and oligomerization process of ARSA.

RMSF calculations on overall trajectory showed that the most mobile residues for all forms are the ones that were missing in the crystal structure and built later (Gly444-Gly447). In Figure 4.4 the red spheres are located on the dimerization3 loop and very close to the octamerization helix. A disruption in this region may cause a hindrance on dimerization or octamerization. To be more accurate, we also calculated the RMSF values of the carbon alpha atoms of each residue as well as their average values for better comparison.

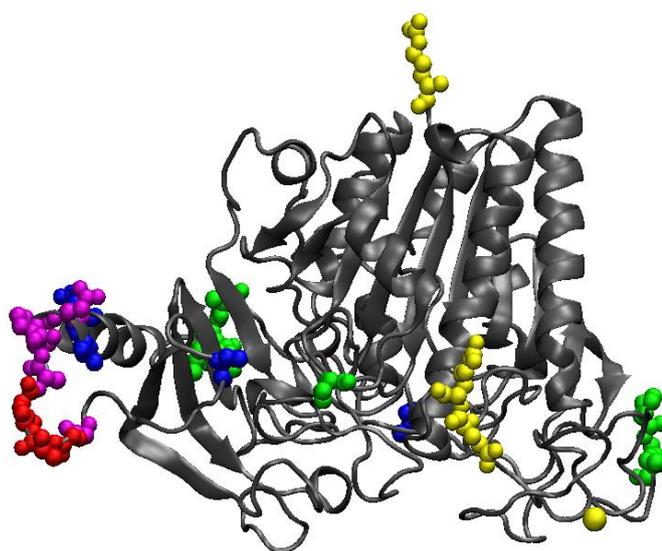


Figure 4.4. Residues with RMSF values higher than 2.0 Å are shown as spheres on the WT ARSA structure.

Red spheres in Figure 4.4 are the residues with RMSF values common in all three ARSA forms (WT and two mutants). Purple spheres are common in two mutant forms (W318C and E307K+T391S). Yellow, green and blue spheres are the residues with RMSF values higher than 2.0 Å in WT, W318C mutant and E307K+T391S mutant respectively.

We selected a threshold value of 2.0 Å to determine the residues that fluctuate more than average. 2.0 Å is much higher than the average RMSF value of the carbon alpha atoms of the whole protein but we also considered that loop regions especially when they are exposed to solvent are expected to move and change their conformations more. We also wanted to be sure to select the residues change their conformations significantly that may affect the stability and oligomerization states of ARSA.

Average RMSF values of the carbon alpha atoms of the residues forming the backbone, core beta sheets, and the important regions of ARSA are given as well as the list of the residues with RMSF values higher than 2.0 Å in Table 4.5.

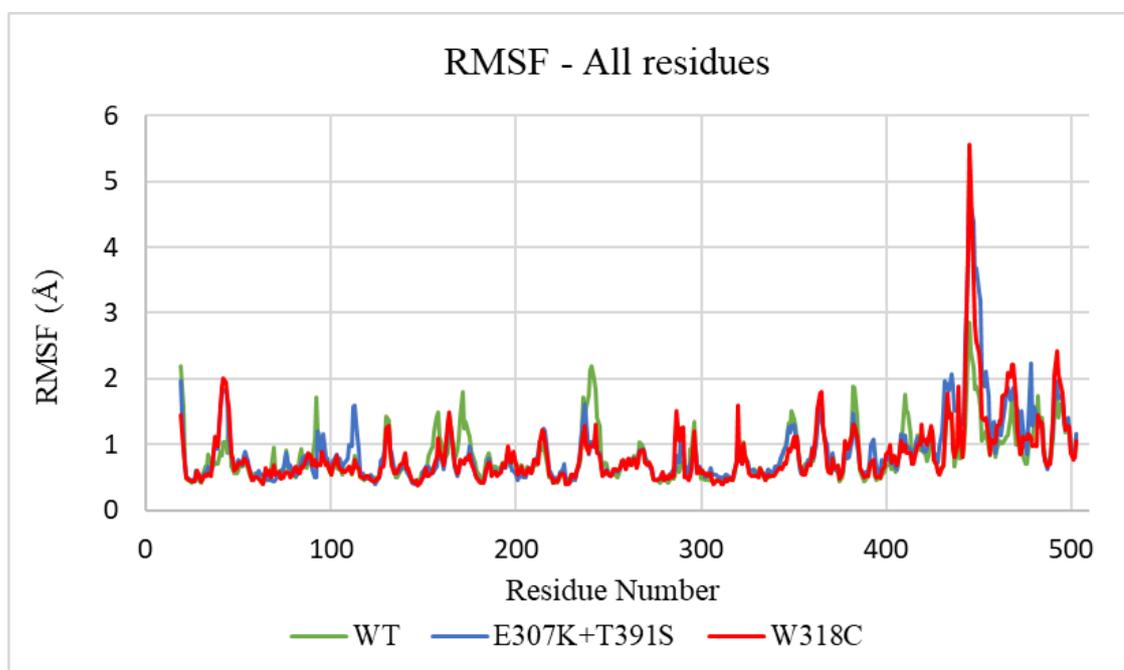


Figure 4.5. Residue based RMSF profiles in the simulations on monomeric ARSA.

The highest RMSF values belong to Val445 residue of mutant forms. It is 5.21 Å for E307K+T391S mutant's Val445 and 5.55 Å for W318C mutant's Val445. On the other hand, the RMSF result of WT Val445 is 2.85 Å. In addition, average RMSF values are given in Table 4.5 for overall structures and certain regions of ARSA forms, along with the residues with RMSF values higher than 2.0 Å. The values were calculated for the whole trajectory. The first frame coordinates of the proteins were taken as references.

Table 4.5. Average RMSF values of the residues for the overall protein structure and specific regions with respect to the initial coordinates.

<b>WT</b>							
	overall	core	dim1	dim2	dim3	octamerization	active site
<b>Avg.</b>	0.82	0.47	0.77	0.71	1.12	1.15	0.52
<b>Residues: &gt; 2.0 Å</b>	19, 240, 241, 444-447						
<b>W318C</b>							
	overall	core	dim1	dim2	dim3	octamerization	active site
<b>Avg.</b>	0.84	0.48	1.05	0.85	1.33	1.59	0.50
<b>Residues: &gt; 2.0 Å</b>	42, 444-451, 466-469, 491-493						
<b>E307K+T391S</b>							
	overall	core	dim1	dim2	dim3	octamerization	active site
<b>Avg.</b>	0.89	0.50	1.01	0.75	1.50	1.71	0.51
<b>Residues: &gt; 2.0 Å</b>	435, 443-452, 454, 478						

Results in Table 4.5 indicate that the residues in the dimerization1 region of the mutant forms become more mobile. The core structure of ARSA (six beta sheets at the center) does not change its rigidity due to mutations. Dimerization3 region residues of E307K+T391S mutant form also become more restless. This region is very long that it consists of 44 residues nevertheless, the average RMSF of the residues increases almost 0.4 Å in the simulation on E307K+T391S mutant form. When we looked at the percent increase of the average RMSF values taking the WT form as reference, we saw that even though the increases of the results of overall protein, active site, core region, and the dimerization2 region are negligible, the percent increases in the average RMSF results for the dimerization1 region are 36 % and 31 % for W318C and E307K+T391S mutants respectively. It becomes 34 % for the dimerization2 region of E397K+T391S mutant form while it is 19 % for the W318C mutant

form. The percent increase becomes more significant for the octamerization helix of E307K+T391S mutant form as 49 % and it is again 38 % for the octamerization helix of W318C mutant. These results agree with the RMSD profiles of the mutant forms of ARSA.

To understand which of the residues are more mobile and fluctuate more than expected we plotted the RMSF results of the carbon alpha atoms of the residues in the B-factor scale along with the experimental B-factors of the residues of ARSA. Results for three simulations are given in Figure A.1. The RMSF calculations were performed for the whole trajectories.

The highest difference between the calculated and experimental RMSF results belong to the residue Val445. It was missing in the crystal structure of ARSA and built in the protein preparation process. The Gly443-Gly447 region of the protein is very mobile. These residues form a loop part in the dimerization3 region which is connected to the octamerization helix at the Pro450 end. Since this loop was built later its initial coordinates are not known and these deviations can be expected but the difference of the differences from the experimental data is still too large. WT form shows a much smaller difference compared to the mutant forms. The flexibility of this region should be inspected further and the results from the simulations in the acidic environment should also be considered.

4.2.1.2. RMSF and Average RMSD Calculations – Specific Regions. ARSA should be glycosylated for lysosomal trafficking and should form dimeric and then the oligomeric structure to become fully active thus, the stability and dynamics of these regions are crucial. RMSD calculations were performed to understand the effects of the mutations on these regions and revealed that most of the changes were observed on the dimerization1, dimerization3 and octamerization regions.

The dimer contacts are maintained by the four loops in the monomer-monomer interface and certain hydrogen bonds between the residues His328-Pro42, Ser43-Ser432, and Tyr439-Thr408. These loops are loop 32-44, loop 286-300, loop 406-414, and loop 430-439. To inspect the structural changes in monomer-monomer interface we divided this region into three main loops and calculated the RMSD profiles of these loops. They were named as Dimerization1, 2 and 3 in the order of the loops: Asp29-Thr46, Thr286-Cys300, and Ala404-

Gly447. The RMSD profiles of these regions are given in Figure 4.6-Figure 4.8 and Figure 4.10. The changes on the trends of RMSD profiles for the dimerization regions are depicted for the mutant structures.

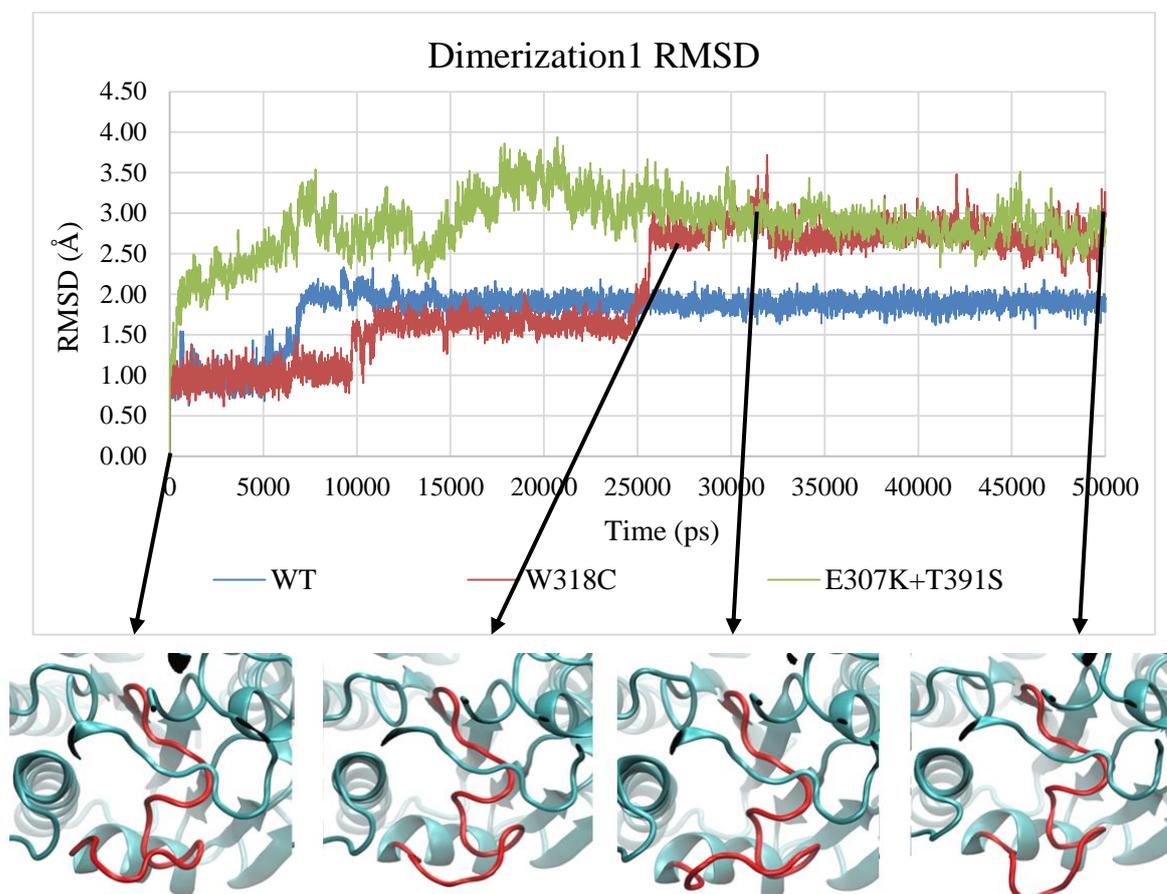


Figure 4.6. Sharp changes in the RMSD profile of the dimerization1 loop of W318C mutant form.

The sharp changes in Figure 4.6 is linked to its movements throughout the simulation. Conformational changes are shown corresponding to the selected time frames from the simulation (0, 26, 31, 50 ns). The same procedure was done for our other mutant E307K+T391S as follows in Figure 4.7.

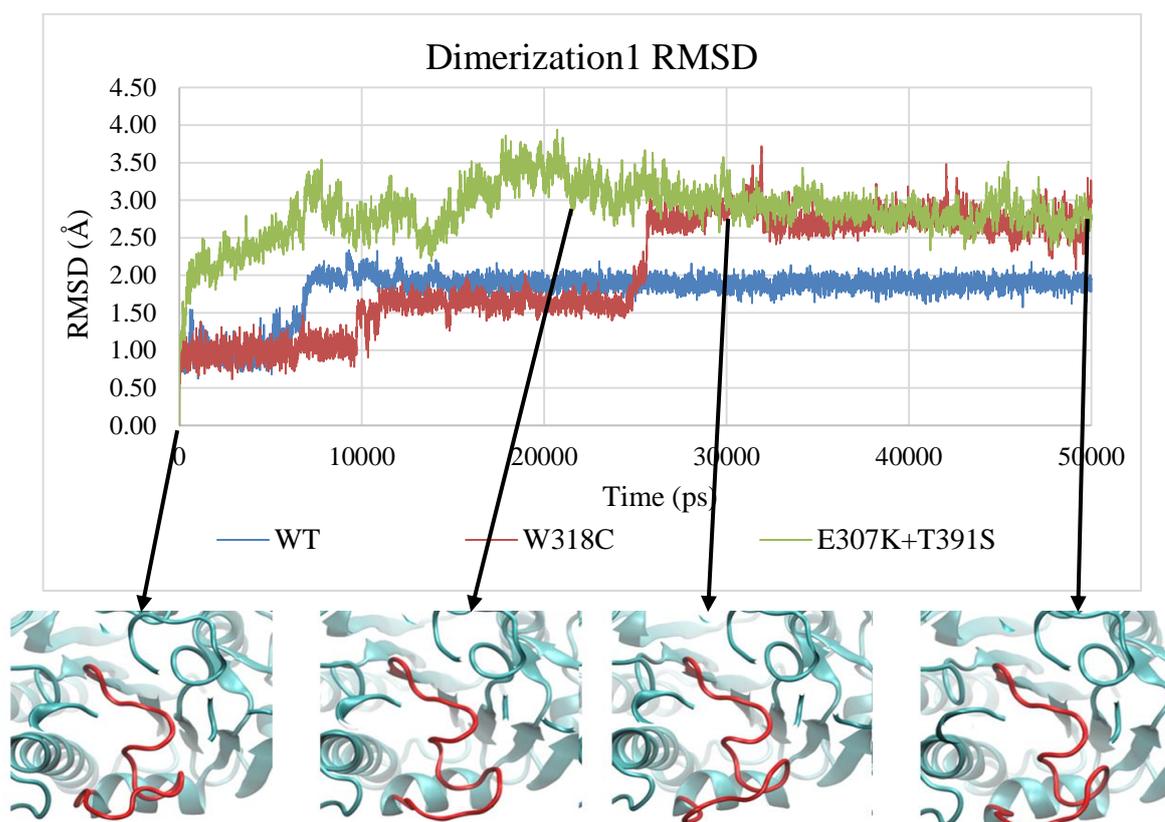


Figure 4.7. Snapshots from the selected time steps (0, 22, 30, 50 ns) of the simulation on E307K+T391S mutant form.

In Figure 4.6 and Figure 4.7, the changes on the dimerization1 regions of the mutant forms are shown. These deviations from the first frame coordinates show that this region cannot reach an equilibrium throughout the simulations. Significant conformational changes at the monomer-monomer interface may disrupt the electrostatic and hydrophobic interactions between the monomers and prevent ARSA from forming dimer structure. Dimerization2 loop RMSD profiles are similar for all three forms of ARSA. They follow a flat trend around 1.5 Å as in equilibrium whereas, a difference was observed in the RMSD profiles of the dimerization3 region of E307K+T391S mutant form. The RMSD profiles of dimerization3 region are given in Figure 4.8.

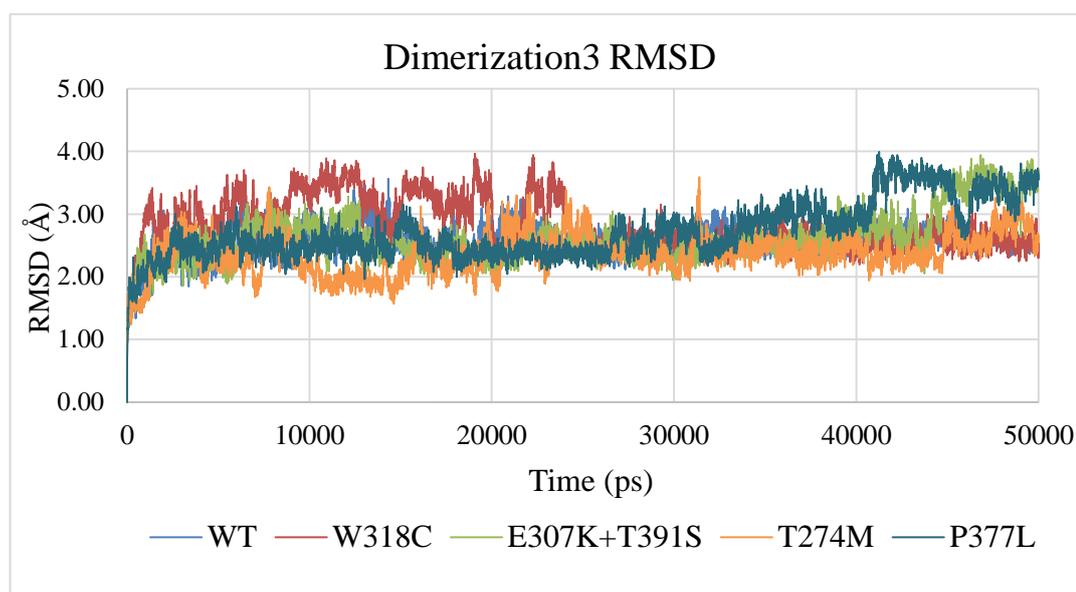


Figure 4.8. RMSD profiles of the dimerization3 regions of all ARSA mutant forms.

The jump in the RMSD profile of E307K+T391S mutant form around 45th ns can be explained with the movement of Val445 on this region. This residue also has the highest RMSF value in E307K+T391S simulation. The conformational change on the secondary structure of dimerization3 loop is also shown in Figure 4.9. The conformational changes of Val445 causes an unwinding movement on this region at 44 ns. This conformational change at the interface of two monomers may be an obstacle in the dimerization of ARSA.

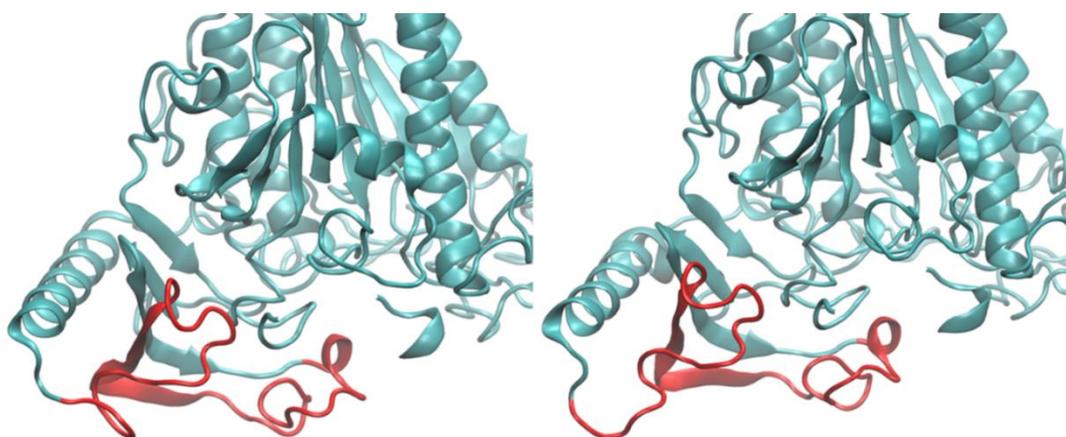


Figure 4.9. Dimerization3 region of E307K+T391S the mutant is shown in red for first frame and 44 ns.

Octamerization helix stability is crucial as well as the dimerization regions because the activity of ARSA depends on its oligomerization state. Octamerization is mainly mediated by the hydrophobic attractions between the two alpha helices at the back of the protein (Figure 1). To understand the effects of the mutations on the octamerization, RMSD profiles of the octamerization helices were calculated. The results are given in Figure 4.10.

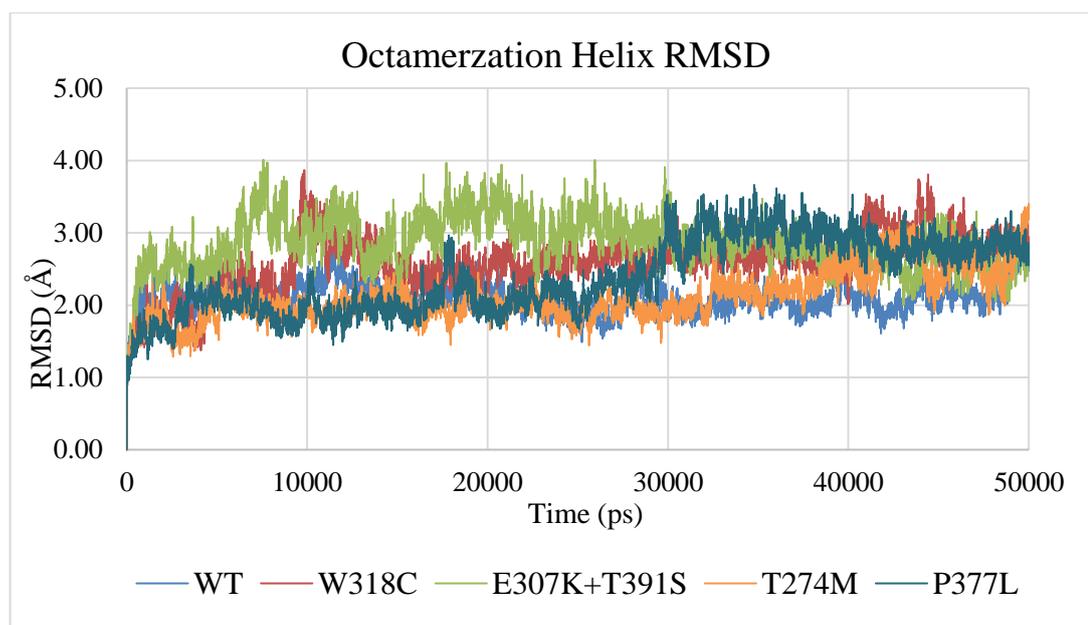


Figure 4.10. RMSD profiles of the octamerization helices of all ARSA monomers.

The difference between the RMSD profiles of the octamerization helices of WT and E307K+T391S mutant form is significant (around 1.2 Å) between 8 ns and 32 ns. W318C has a gradually increasing RMSD profile although this increase is slight, and jumps in the RMSD trend of its octamerization helix. From the profiles of the mutant forms, the octamerization helix of W318C starts to change its secondary structure over time while the helix of E307K+T391S mutant changes its conformation in the middle of the simulation. The structural change on the octamerization helix of W318C mutant form is significant only on the Val470 end of the helix and it occurs due to Asp467 which is a negatively charged, polar residue, turning its side chain towards the solvent. Also, Leu466 on the octamerization helix of W318C mutant form turns its side chain towards the bottom of the helix and becomes partially solvent accessible. Leu466 on the octamerization helix of E307K+T391S mutant changes its conformation more significantly and turns its side chain towards the top of the

helix and becomes partially solvent accessible. These turning movements of Leu466 were not observed in the simulation on WT form, but WT RMSD profile shows a steep increase at the end of the simulation for the octamerization helix. This can be linked to the movement of Asp467 going away from Ala464 due to the lost hydrogen bond between Gln456 and Ala469. The lack of hydrogen bond between these residues causes an unwinding movement at the Val470 end of the octamerization helix. This change is not a major change on the secondary structure. When the secondary structure of each octamerization helix of the mutant forms is observed, the unwinding behavior of Val470 end of the helix becomes significant along with the conformational changes of Leu466.

The general trend is that Val470 end of the octamerization helix is flexible in the simulations on monomer forms of ARSA in neutral pH, the results should be compared with the ones in acidic pH for better understanding. For clarity, the ARSA octamer (biological assembly from PDB) is shown in Figure 4.11 and the Val470 end of the octamerization helix is shown as circled.

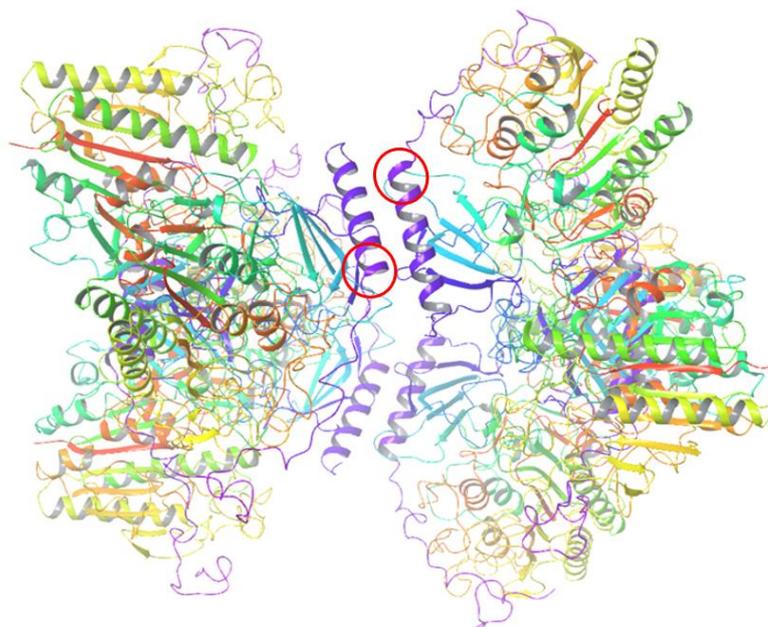


Figure 4.11. The biological assembly of ARSA (PDB code: 1AUK). Val470 ends on the octamerization helices are indicated with red circles.

The most mobile part of the octamerization helices in the simulations is shown in Figure 4.11 in red circles. This region seems to be in a complete contact with the other octamerization helix of the other monomer of the octamer thus, it is probably important for the durability of the octameric structure. On the other hand, even in the simulation on WT ARSA form, this region was still observed to fluctuate for a short time. This fluctuation is not a major structural change and we hypothesized that it may occur if the monomers do not form the octameric structure. More significant changes should be considered as hindrance on the octamerization. For this purpose, we also performed secondary structure analysis to determine significant changes and their duration on the ARSA protein.

In order to understand the effects of these changes on the specific regions and the oligomerization process of the enzyme, we analyzed the trajectories and calculated the hydrogen bond occupancies between the residues on the specific regions of ARSA.

#### **4.2.2. Salt Bridges and Hydrogen Bond Occupancies**

Possible salt bridge formation and hydrogen bond occupancies between certain residues were calculated in addition to the overall structural analysis. Possible salt bridges were calculated for the entire enzyme in each simulation and the additional salt bridges from each simulation were listed in Table 4.6. Average distances between the residues were calculated for the whole simulation time in Table 4.6. The residues for hydrogen bond occupancy calculations were selected according to their locations, functions, being mutant, their fluctuations based of RMSF calculations and visual inspection. For instance, Arg19 was selected due to being the early forming epitope and the first residue after signal peptide cleavage hence, solvent exposure and hydrogen bonding of Arg19 may be crucial for protein folding. The hydrogen bond occupancies for Arg19 and similar selected residues were listed in Table A.1 as well as all the possible salt bridges are summarized under the corresponding simulation titles in Table A.2.

Table 4.6. Possible salt bridges in the corresponding simulations.

Residue Pairs	WT	W318C	E307K+T391S	T274M	P377L
ARG19-ASP216	-	-	-	7.00 Å	7.24 Å*
ARG19-GLU272	-	-	-	11.02 Å	-
ARG19-GLU273	-	-	3.25 Å	4.20 Å	-
ARG299-GLU437	-	-	-	-	6.23 Å**
ARG97-GLU131	-	6.18 Å	-	-	9.25 Å
ASP152-ARG288	-	-	6.83 Å	-	8.55 Å
ASP29-HIS229	-	-	4.80 Å	-	-
ASP61-HIS328	-	-	-	6.24 Å	-
ASP61-LYS393	-	-	-	-	12.56 Å
GLU131-ARG496	9.17 Å	-	-	-	-
GLU424-LYS463	-	-	-	8.13 Å <sup>†</sup>	6.12 Å***
GLU272-HIS321	-	-	-	8.66 Å	-
GLU307-ARG311	4.23 Å	4.27 Å	-	4.25 Å	4.24 Å
GLU307-ARG390	-	-	-	6.03 Å	-
GLU307-LYS395	3.63 Å	3.63 Å	-	3.62 Å	3.68 Å
GLU382-ARG384	-	-	-	9.07 Å	-
LYS433-GLU451	-	-	-	-	12.20 Å

\*: Last 10 ns the average distance is approximately 4.2 Å

\*\* : Last 25 ns the average distance is approximately 5.0 Å

\*\*\*: After 18 ns, the average distance is approximately 4.0 Å

<sup>†</sup>: Last 5 ns the average distance is approximately 3.0 Å

In the simulation on E307K+T391S mutant form, there are two missing salt bridges between Glu307 and both Arg311 and Lys395. This may cause an instability in this region which is also close to the late forming epitope region. In addition, A possible salt bridge between Asp152 and His229 is present in both simulations of the mutant forms. This may cause a decrease in the degree of freedom of around His229. Due to additional hydrogen bonds or salt bridges, misfolding may occur leading protein to stay in other local minima in its energy funnel. By this way, misfolded mutant enzymes may aggregate and degraded before leaving ER or they may accumulate in other places within the cell. We did not consider additional salt bridges with an average distance more than 5 Å as important, and to verify these assumptions, we calculated the hydrogen bond occupancies between these residues and summarized our results in Table A.1.

The results in Table 4.6 show that additional salt bridge between Asp152 and His229 is not actually present in the simulation because these residues do not form hydrogen bonds throughout the simulations of both mutant forms (W318C and E307K+T391S). In WT ARSA and W318C mutant form simulations Glu307 forms hydrogen bonds mainly with Lys395 and Glu437 in agreement with the salt bridges. Hydrogen bond occupancies of these pairs are higher than 70% for both simulations. On the other hand, these hydrogen bonds are lost or their occupancies are decreased in E307K+T391S simulation as expected. This loss may cause an instability in the structure of this region. Glu307 is located at the center of the late forming epitope region, and it is close to the dimerization site. This instability may prevent ARSA from folding correctly or from forming dimers causing inactivity. The same is not valid for W318C mutant, Cys318 forms hydrogen bond with a nearby water molecule mostly and does not form hydrogen bond with Glu272 but in WT and E307K+T391S mutant simulations Trp318 also forms hydrogen bond with a water molecule for a significant amount of time. Surprisingly, in the simulation on E307K+T391S mutant form Trp318 forms less hydrogen bonds with neighboring water molecules or residues than the other two simulations. In addition to these, hydrogen bond occupancies of Lys433 change with the mutations. It is in the middle of the late forming epitope region and close to the dimerization site. The total hydrogen bond occupancy of Lys433 in WT simulation is 79% and in W318C simulation is 98% while it falls to 32% in the simulation on E307K+T391S mutant. When we inspect the trajectories, we saw that Lys433 turns its sidechain towards to solvent in the simulation on E307K+T391S mutant but it bows its head and forms hydrogen bond with Asp434 in the simulation on W318C mutant.

#### **4.2.3. Secondary Structure Analysis**

Changes in the RMSD profiles of the three forms can be explained by the movements of these specific regions. These movements can cause a change in the secondary structure change like unwinding an alpha helix to a coil. Secondary structure changes can disturb the oligomerization states of the proteins disturbing the hydrophobic balance between the monomers. For a better understanding on the inactivity of ARSA due to mutations we performed secondary structure analysis using the Simulation Interactions Diagram (SID) of Schrödinger and Timeline tool in VMD. The results showed that the sudden changes like jumps and constant increases in RMSD profiles can be linked to the secondary structure

changes. An example of a result from the Timeline tool of VMD is shown for three ARSA monomers at neutral pH in Figure 4.12.

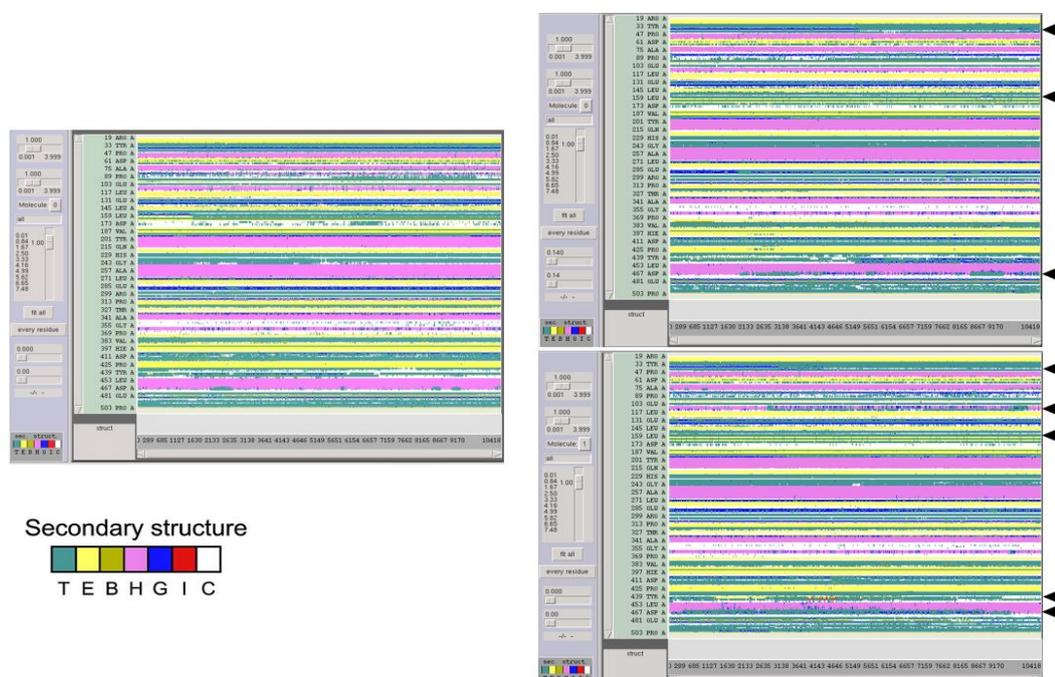


Figure 4.12. Secondary structure maps of the WT, W318C mutant and E307K+T391S mutant.

The regions that change their colors during the simulations on mutant forms were indicated with arrows in Figure 4.12. The top map belongs to the W318C mutant form, whereas the bottom map belongs to the E307K+T391S mutant form. The color codes are as follows: Green is Turn, Yellow is Extended configuration, Sand is Isolated bridge, Pink is Alpha helix, Blue is 3-10 Helix, Red is Pi-helix, and White is Coil.

Secondary structure maps show that mutations may influence the secondary structure of specific regions. These regions on the E307K+T391S mutant form are Tyr33-Thr45 loop, Glu103-Gly115, Asn158-Gly171, Glu437-Val445 regions, and Val470 end of the octamerization helix. On W318C, changes in the secondary structure occur on the regions: Tyr33-Thr45 loop, Asn158-Gly171 region and Val470 end of the octamerization helix. When the trajectory of the simulation on the E307K+T391S mutant was analyzed, the dimerization1 region was observed to change its conformation. This region includes Tyr33-

Thr45 loop, and this difference is linked to the conformation change of Phe41. The sidechain of this hydrophobic residue first forms a hydrogen bond with Ser44 and remains its position towards the center of the molecule. Then this hydrogen bond breaks and Phe41 becomes exposed to the solvent. This behavior happens in the dimerization region and can cause a repulsion between the monomers. The unwinding movement of the Glu103-Gly115 region on the alpha helix on the top of E307K+T391S mutant form of ARSA can be linked to the conformation changes of Arg114. Throughout the simulation, it constantly changes its conformation. This region is close to the Arg19, the first forming epitope, and may cause an instability during protein folding. When the trajectory of the simulation on W318C was analyzed, three of the same regions as on E307K+T391S mutant were determined to change their secondary structures. Phe41 turns its side chain towards the solvent also in the simulation on W318C. On the other hand, this movement of Phe41 is absent in WT simulation. This suggests that the exposure of the hydrophobic side chain of Phe41 in the dimerization region may be causing a hindrance in the dimerization of ARSA after its synthesis.

#### **4.2.4. State of NAG+NDG Sugar Moiety at the Glycosylation Site**

In its crystal structure, ARSA has a sugar entity (NDG+NAG) which is covalently bonded to Asn184, one of the three glycosylation sites [49,14,100,101]. When we observed all the simulations on all monomer forms at neutral pH, we saw that sugar moieties in all simulations, except on W318C mutant monomer, behave almost as the same. As an example the first and the last frame snapshots from the simulation on T274M mutant monomer form is shown in Figure 4.13.

In all simulations, NAG+NDG group bends (shown with sticks representation), and then turns its backbone almost 180 degrees around Asn184 as shown in Figure 4.13. The different conformation that of the sugar moiety in the simulation on W318C mutant monomer at neutral pH is shown in Section 4.3.4.

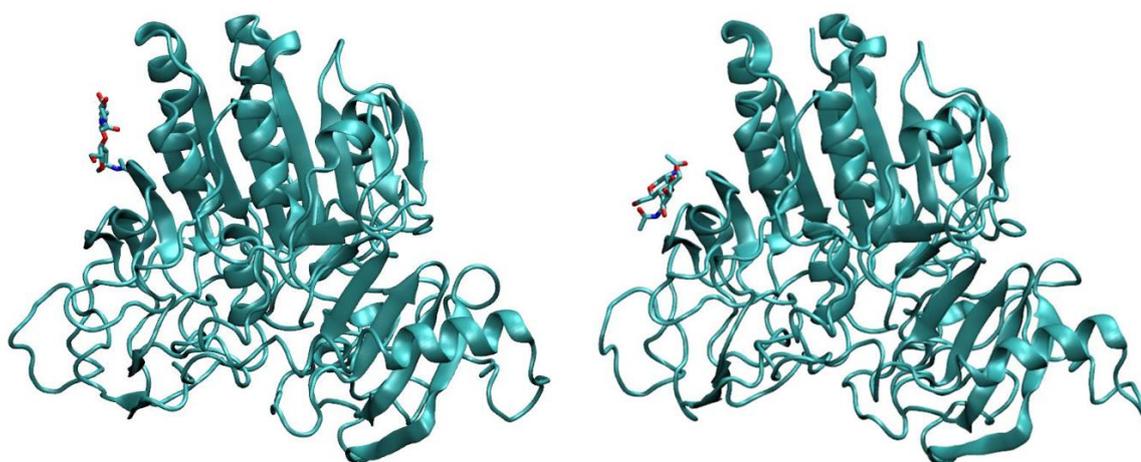


Figure 4.13. Snapshots at 0 ns (left) and 50 ns (right) of T274M mutant monomer.

#### 4.2.5. The Distance between OE2/O atom of Glu/Lys307 and NZ atom of Lys395

In the simulation of the mutant form the distance between the mutant residues show significantly different trends from the WT and W318C monomers throughout the simulation as shown in Figure 4.14.

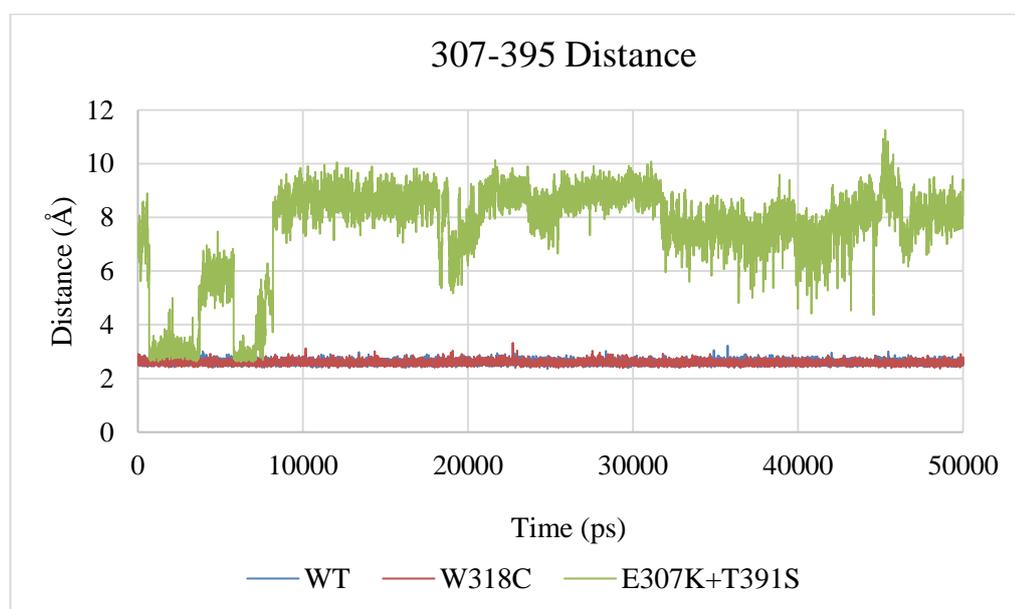


Figure 4.14. The distance between the OE2/O atom of Glu/Lys307, and NZ atom of Lys395 on WT, E307K+T391S and W318C mutant forms as a function of simulation time.

This difference occurs because of the electrostatic changes in this region due to mutation of E307K. Water molecules come in between and form a shield between the residues, Lys307 and Lys395. These residues are near the dimerization region and in the late forming epitope region as previously stated. The significant difference between these residues may imply an instability near the dimerization region of E307K+T391S. The instability on this region can also cause a change on the association angle between the monomers as in the case of C300F mutant ARSA form and consequently affect the octamerization of the mutant in the lysosome. In addition to these, this situation may also be affecting the folding. These are all the scenarios that can lead to MLD, thus dimer dynamics should be examined carefully and experimental examination is needed.

#### **4.2.6. Dihedral Angle Changes in the E307K+T391S Mutant**

Dihedral angles (psi and phi) were calculated for all simulations through whole trajectories. When we examined the results of E307K+T391S monomer we saw that Thr45 changes its phi angle after 15 ns different than the WT monomer. In addition, in the simulation on E307K+T391S mutant monomer the region between Gln215 and His229 show more stable phi trends than the WT. The residues Met120-Pro130 which forms the loop structures at the back of the protein (around octamerization helix) change their psi angles severely as shown in Table 4.17. In addition, the residues in the region between Tyr33 and Pro47 change their phi angles notably in the simulation on E307K+T391S mutant monomer. We mapped the secondary structure changes throughout the simulations for the three forms. The results show that there are notable the secondary structures of the specific regions. These regions on the E307K+T391S mutant form are Tyr33-Thr45 loop on the dimerization1 region, Glu103-Gly115, Asn158-Gly171, Glu437-Val445 on the dimerization3 region as well as the C-terminus of the octamerization helix of E307K+T391S mutant. They are in accordance with the phi angle changes as shown in the dihedral angle maps of E307K+T391S monomer (Table 4.17). When we compare the mutant E307K+T391S with severely misfolding mutant T274M the most significant change is observed on the changes of the psi angles of the residues around Tyr439.

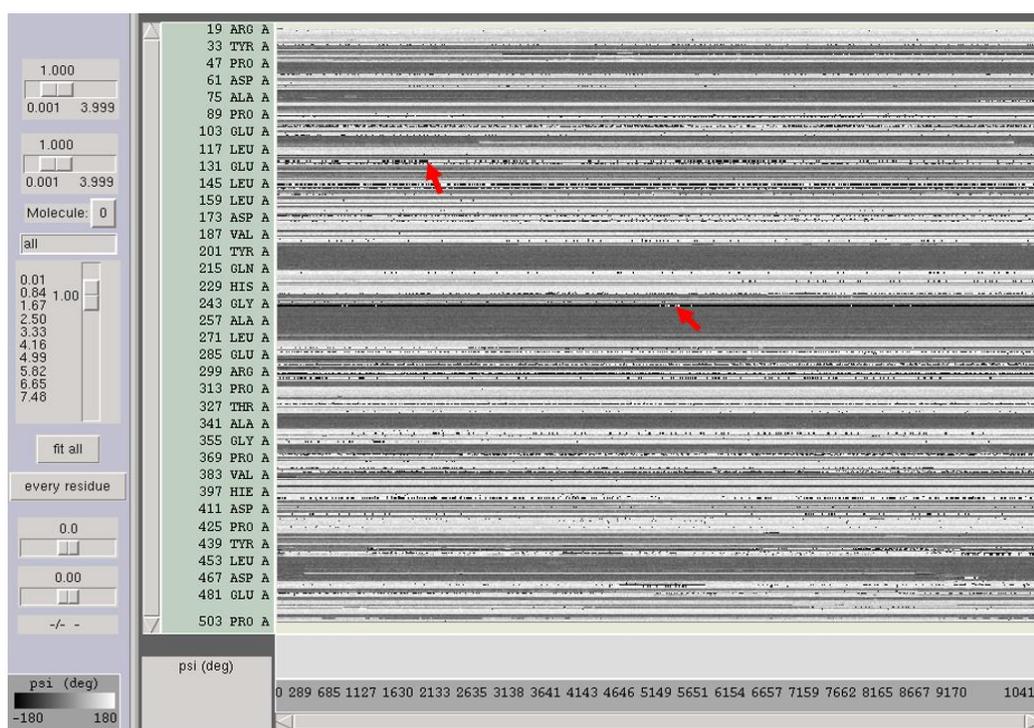


Figure 4.15. The psi angle changes of the residues of E307K+T391S monomer (vertical axis) as a function of simulation time.

The numbering of the residues in Figure 4.15 starts with 19 and ends with 503 as in the RMSF calculations. Black color represents the -180 degrees, while white color represents +180 degrees. Red arrows indicate the severe changes in the psi angle, in other words: color changes from black to white or vice versa.

Dihedral angles in the simulations on dimer forms showed that the dynamics may be very different between the two chains of a dimer even in the simulation on WT ARSA. Psi angles, especially, show different trends for the chains of the WT dimer form. When we look at the phi angles, the residues around Pro452 of the chain of WT dimer show different trends during the whole trajectory. These changes can be linked to the hetero dynamics of the two octamerization helices of WT dimer. When we focused on the changes of the dihedral angles of E307K+T391S mutant dimer, we saw that the residues around Pro155 change their psi angles more frequently and tend to preserve the opposite (-180 degrees) conformation when compared to those of WT. This region corresponds to the opposite site of the dimerization region, and it is close to the C-terminus of the protein. In addition, this region includes an

active site residue, Ser150 and this situation may not be important at neutral pH but it may be crucial for the activity when it also happens in the lysosome. Another significant difference was observed for the residues in the region between Met289 and Ala316 that their psi angles change more from -180 degrees to +180 degrees than the 307K+T391S dimer. This region is the inner loop between the loops at the dimerization region. It also includes Glu307 residue. There are two remarkable differences between the phi angles of the WT and E307K+T391S dimer forms. One is on the residues around Pro155 as in the maps of psi angles. The other is on the residues in the region Gly154-Leu182. This region is on the opposite site of the dimerization site as shown in Figure 1 and is close to the C-terminus of the proteins. The conformational changes on this region may be affecting the stretching of the bottom part of the hat-like shaped ARSA proteins leading to angle changes between monomers.

When we examined the angle changes residue by residue in the simulations on monomer forms of WT and E307K+T391S mutant we saw that some residues change their psi angles from -180 degrees to +180 degrees. We listed the residues for E307K+T391S monomer that change their psi angles significantly different when compared to WT, in Table 4.7. We selected the residues with angles changes more than approximately 90 degrees or prone to change its psi angle more or much less frequently than those of WT.

Table 4.7. The residues with significant psi angle changes on the monomer form of E307K+T391S mutant when compared to WT ARSA.

<b>Chain name</b>	<b>Residue number</b>
A	Gly90, Gly99, Gly129, Tyr149, Gly154, Pro164, Gly170, Cys172, Gln174, Gln215, Thr279, Gly293, Cys300, Thr304, Gly364, Gly366, Pro380, Asp381, Gly402, Ser419, Gly443, Gly444, Gly447, Lys463, Val470, Gly473, Gly480

The results show that most of the Gly residues show significant changes in their psi angles between two forms of ARSA monomer. Glycine is the smallest residue and its frequent angle changes can be associated with the conformational changes around corresponding Glycine residues. When we mapped these Glycine residues on the 3D crystal

structure we saw that they are spread over the whole structure, but most of them are on the loop regions at the bottom of the hat-like shaped are except both Gly364 and Gly366.

### **4.3. MD Simulations on Monomeric ARSA at Acidic pH**

ARSA forms its octameric form in the lysosome. To form the octameric structure it first forms the dimer form. The formation of these oligomeric structures depends on the hydrogen bonds, electrostatic and hydrophobic interactions between the monomers on the regions which were listed in Table 4.4. Since the activity of the enzyme depends on the oligomerization, the stability of the oligomerization regions and the overall protein can affect the activity of the enzyme. We considered the hindrance of the oligomerization leading to inactivity of ARSA thus, to understand the effects of the mutations, we performed RMSD and RMSF calculations with the existence of possible salt bridges and hydrogen bond occupancies for the simulations on eight types of ARSA in acidic environment. These consist of the two hydration states of FGL for each WT, E307K only, E307K+T391S, and W318C mutant forms of ARSA. The hydration state of FGL was taken into consideration and the diol state of FGL is expected to be formed after the ARSA monomers enter the lysosome and form the octameric structure since FGL becomes diol during the reaction in the lysosome and this state should also be analyzed in terms of stability and dynamics [1]. The results of the calculations were discussed as below.

#### **4.3.1. RMSD Calculations for the Overall Structure and the Active Site**

The same procedure was followed for the analysis of the simulations in acidic milieu but two types of ARSA forms were evaluated for each simulation on WT, E307K+T391S mutant, E307K single mutant, and W318C mutant forms in terms of having a hydrated FGL residue. FGL becomes hydrated during the catalysis of sulfate ester bonds. To understand any effect of this difference, we used two types of FGL, diol and hydrated, in our MD simulations. RMSD, RMSF, salt bridge and hydrogen bonding calculations were performed to understand the changes on the structure and the dynamics of ARSA in the acidic environment. The overall backbone RMSD profiles are given in Figure 4.16.

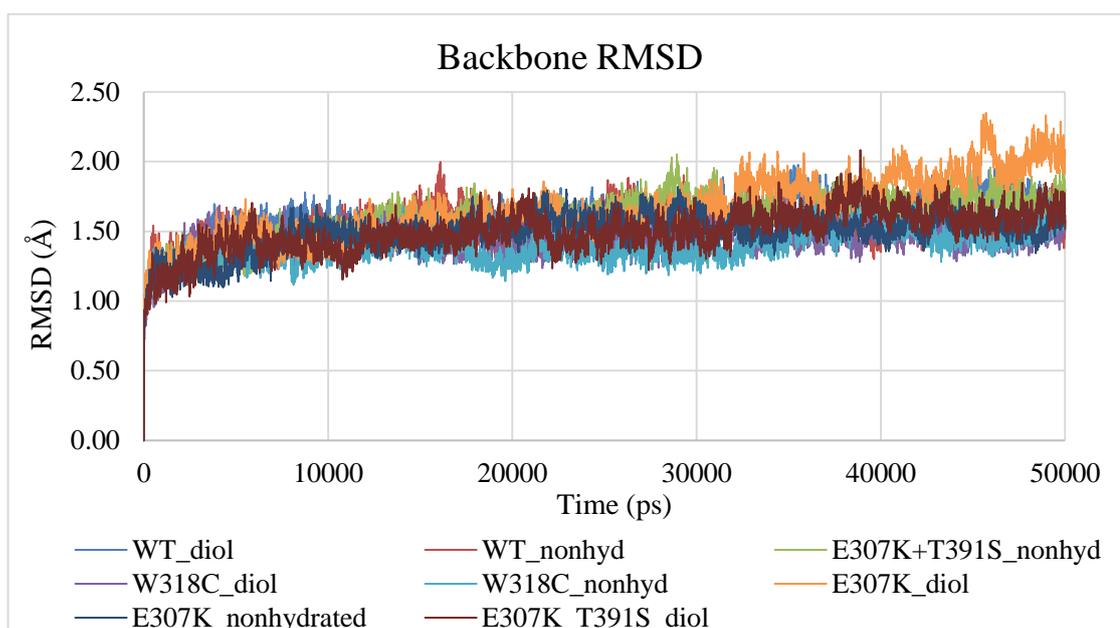


Figure 4.16. Backbone RMSD profiles of the four different ARSA forms for the simulations at acidic pH.

The backbone RMSD profiles show that no significant change on the overall structural stability is present due to mutations and hydration state of FGL. The most significant outcome is that E307K mutant form cannot reach an equilibrium throughout the 50ns simulation. Its backbone RMSD profile shows a constantly increasing trend. Otherwise, from backbone RMSD profiles, we can conclude that pH does not affect the overall stability of the protein, significantly.

To better understand the effects of the hydration state of FGL, pH and the mutations, RMSD calculations were performed for the specific regions of ARSA as well. We first examined the differences on the active site due to the state of FGL and acidic pH. Accordingly, the active site RMSD profiles of all simulations on eight different forms of ARSA are shown in Figure 4.17.

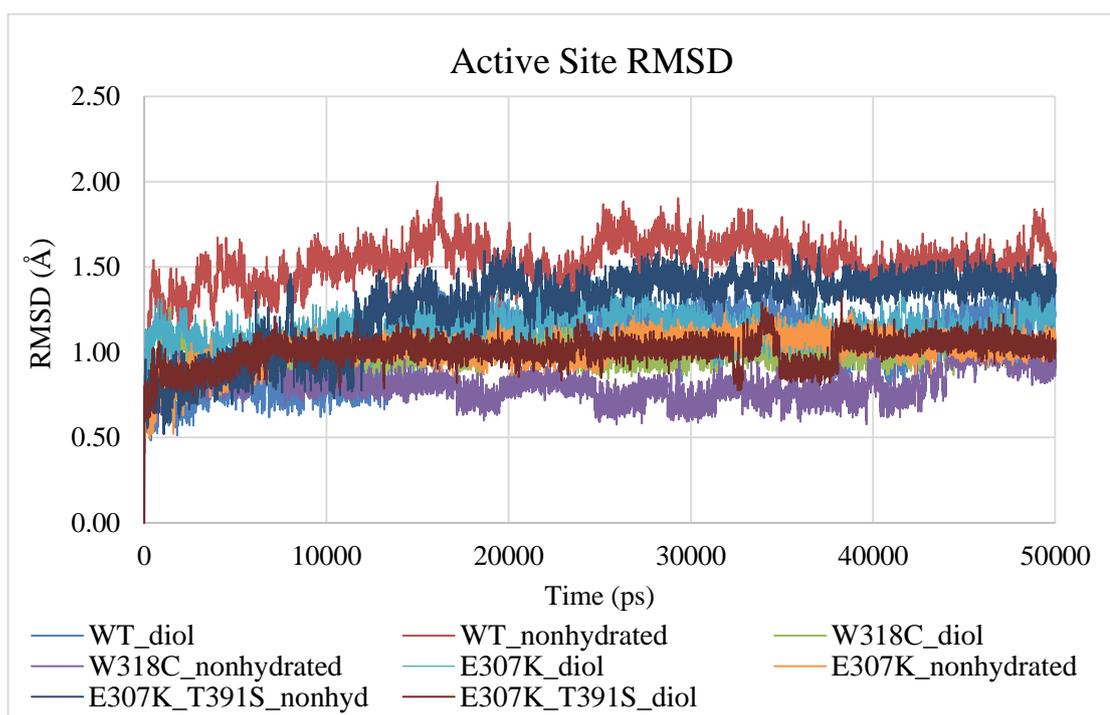


Figure 4.17. RMSD profiles of the active site residues for simulations at acidic pH.

Active site RMSD profile of the WT form with a nonhydrated FGL shows a significant difference from WT FGL in diol form of ARSA. There is a significant gap between W318C mutant forms with both FGL types and WT FGL in nonhydrated form of ARSA. Active sites of W318C types of ARSA stay rigid throughout the simulations. The RMSD profiles of the active sites of E307K+T391S and only E307K mutant forms of ARSA follow the similar trend of the RMSD profile of the active site of WT FGL in diol form of ARSA. The overall results do not differ significantly from the results of the simulations at neutral pH but the mutant forms of ARSA show lower trends of active site RMSD profiles than the ones performed at neutral pH. Results show that the most rigid active site structure is present in the simulations on W318C mutant form for its both types and E307K+T391S mutant form with its FGL in nonhydrated form. The general trend is that when FGL becomes a diol, the active site becomes more stable. This can be associated with the reaction mechanism of ARSA. FGL becomes diol when it is to catalyze the desulfurization reaction. On the other hand, the hydration state of FGL does not affect the overall backbone stability of the protein that the backbone trends are similar for all forms of ARSA at acidic pH. As for the neutral

pH simulations, RMSD profiles of the other specific regions on ARSA forms were calculated for the acidic pH simulations.

### 4.3.2. RMSD Calculations of the Specific Regions in Detail

Specific regions of ARSA were determined to be dimerization, octamerization and glycosylation sites other than the active site. The RMSD profiles of these regions from all simulations on eight different ARSA forms in acidic pH are given in Figure 4.18, Figure 4.27, Figure 4.33, Figure 4.38, and Figure 4.45. Most significant changes in the profiles are in dimerization1, dimerization2 and octamerization regions' RMSD profiles while in the RMSD profiles of 335-370 region including glycosylation site there are few sharp increases. The RMSD profiles of the dimerization1 regions of all forms are given in Figure 4.18.

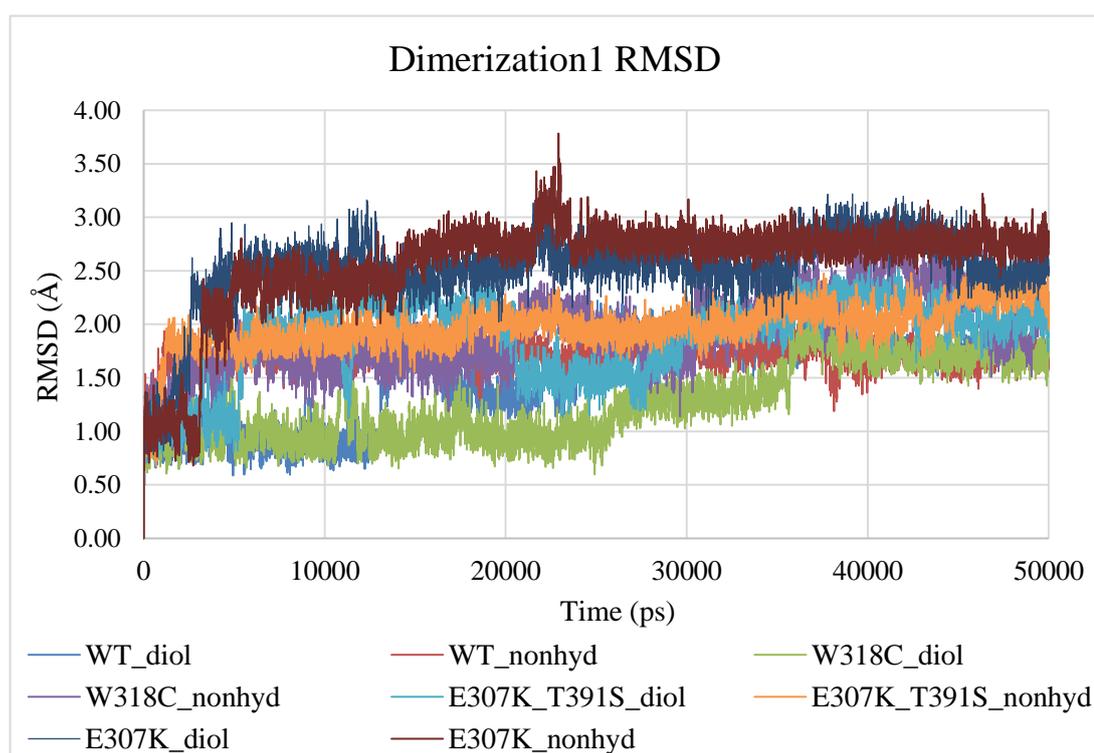


Figure 4.18. RMSD profiles of the dimerization1 regions of all eight ARSA forms.

There is a significant RMSD difference ( $\sim 1.1 \text{ \AA}$ ) between the top and bottom RMSD profiles after the trends reach a plateau. The structures whose dimerization1 loops deviate mostly from the initial crystal structure are the two types (FGL in diol and nonhydrated

states) of E307K mutant form of ARSA. In the last 10 ns, W318C-nonhydrated form of ARSA shows an increase and then a decrease again. Its profile cannot reach an equilibrium while, the trend of the E307K+T391S-nonhydrated form of ARSA quickly reaches an equilibrium at the first 2 ns and after 17 ns it starts to increase slightly but gradually. When only these two profiles are graphed it is seen that they follow a very similar trend which is also similar to the RMSD profile of WT-nonhydrated form as shown in Figure 4.19.

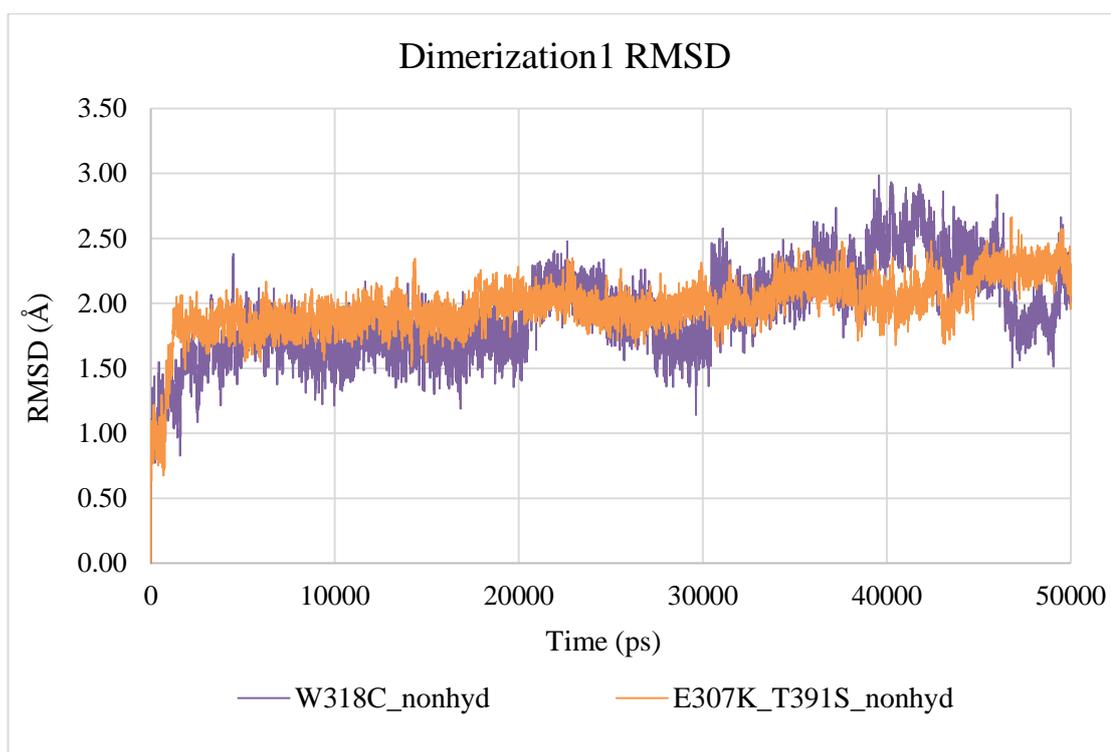


Figure 4.19. Dimerization1 RMSD profiles of the nonhydrated forms of mutant ARSA.

Diol form of the WT structure first follows a flat trend around 1.0 Å and then jumps to 1.5 and 1.7 Å band. When it reaches an equilibrium after 25 ns it follows almost the same trend with the nonhydrated form of WT structure's. The hydration state of FGL in the mutant structures do not affect the dimerization1 region's stability significantly. Only in the case of W318C mutant, diol form of W318C first follows a flat trend around 1.0 Å as the diol form of WT structure but again as similar it reaches an equilibrium around 1.6 after 35 ns. The RMSD profiles of the four forms of ARSA are given in Figure 4.20 and Figure 4.23 separately for the two states of FGL.

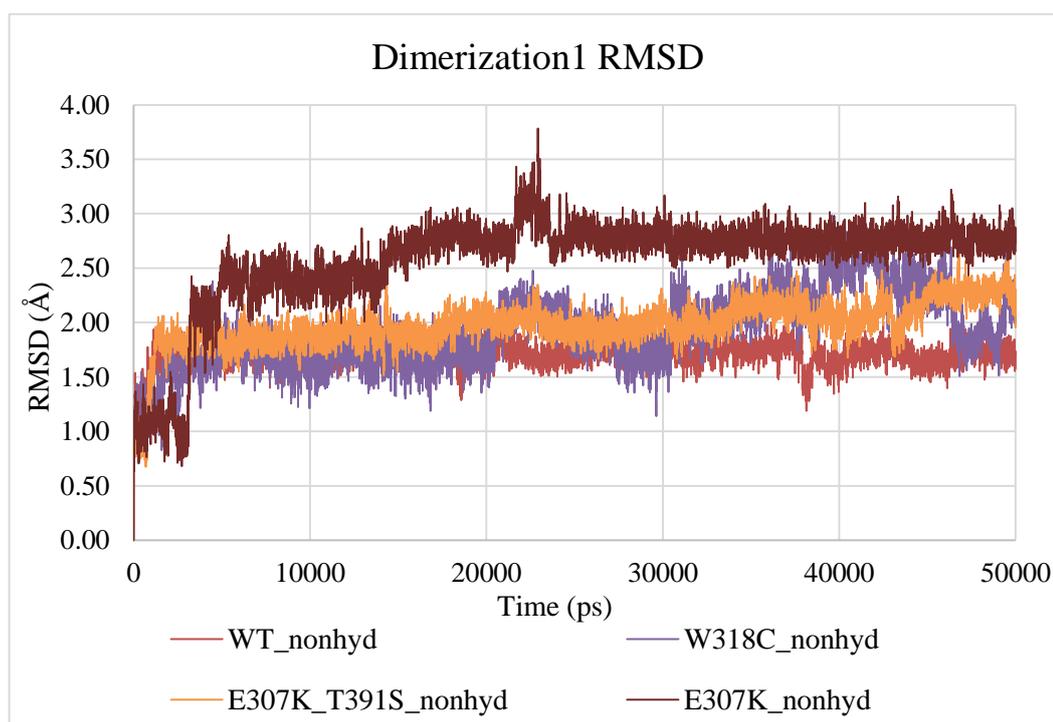


Figure 4.20. RMSD profiles of the dimerization1 region of nonhydrated forms of WT, W318C and E307K+T391S mutant forms of ARSA.

The RMSD profiles of the nonhydrated forms show that addition of T391S mutation makes dimerization1 region fluctuate less from its initial coordinates than the dimerization1 region of E307K only mutated form of ARSA. The increasing trend in the RMSD profile of W318C mutant between 35 and 45 ns can be linked to the conformational change of the dimerization1 loop shown in Figure 4.21.

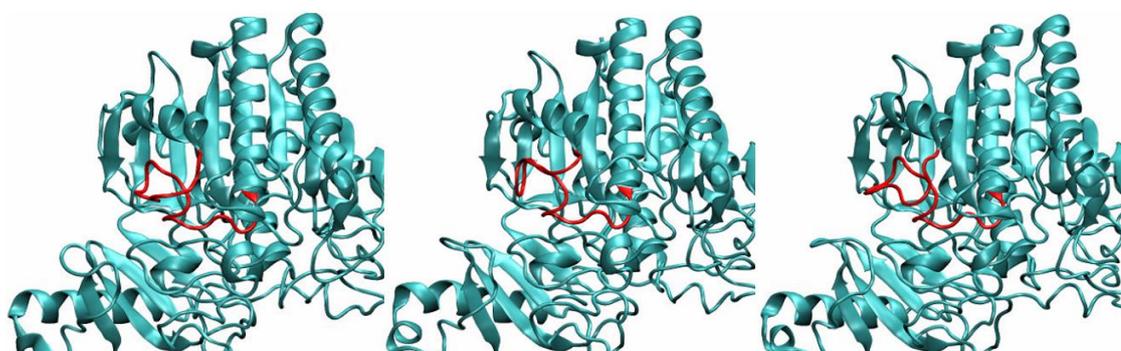


Figure 4.21. Dimerization1 loop (red) of W318C-nonhydrated mutant type of ARSA at 0 ns, 32 ns and 45 ns of the simulation.

This conformational change is due to the movement of Pro42 which turns its side chain upwards as shown in Figure 4.22.

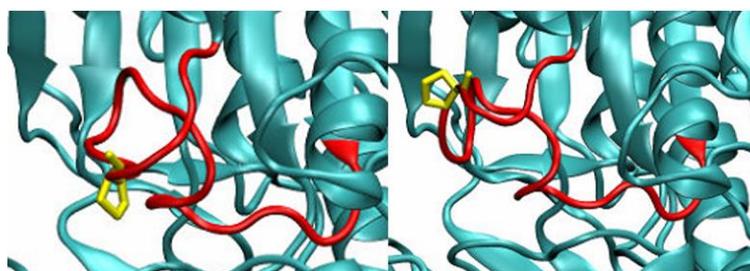


Figure 4.22. The conformational change of Pro42 (yellow) on the dimerization1 region (red) of W318C-nonhydrated mutant form.

The conformational change of Pro42, which is a hydrophobic residue, in the monomer-monomer interface can cause a disruption on the hydrophobic interactions between the monomers causing a hindrance on the dimerization of ARSA. On the other hand, in Figure 4.23 the RMSD profiles of the dimerization1 loop are given for the hydrated state of the structures in the position of FGL69.

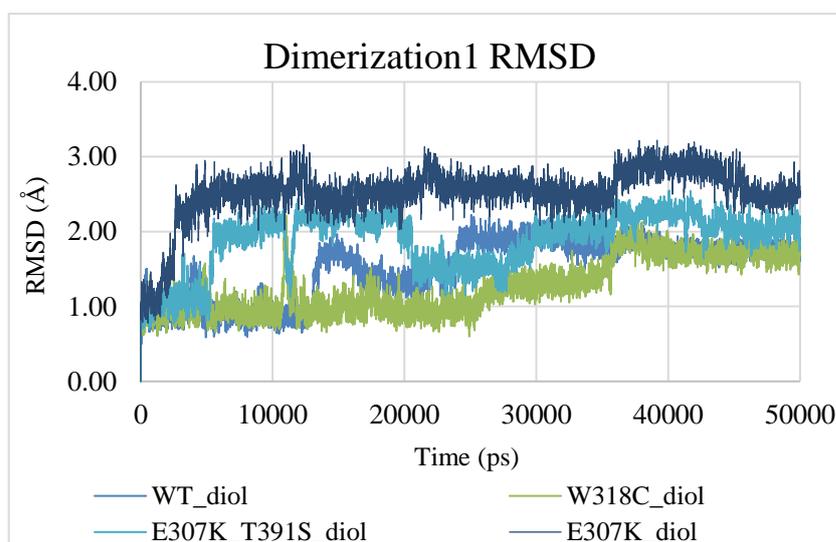


Figure 4.23. RMSD profiles of the dimerization1 region of diol forms of WT, W318C and E307K+T391S mutant forms of ARSA.

Dimerization1 region consists of 18 residues as Asp29-Thr46. Figure 4.18 shows that this time W318C mutant forms of ARSA and WT ARSA forms show the least deviations from the initial structure for the dimerization1 region. E307K mutant structure cannot reach an equilibrium for this region while, E307K+T391S mutant form follows a similar trend to the WT form's after 20 ns. Between 5 ns and 20 ns, the RMSD difference of the dimerization1 loop profiles between WT and E307K+T391S diol forms become significant. When the trajectory of E307K+T391S-diol mutant form was observed, conformational changes on the dimerization1 region were observed as shown in Figure 4.24.

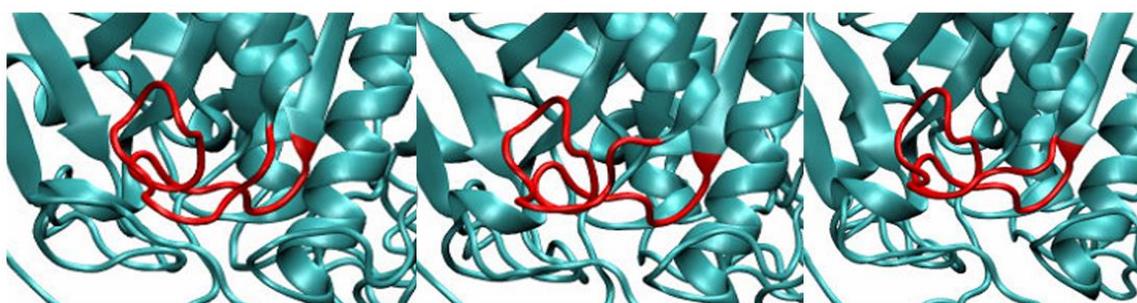


Figure 4.24. Dimerization1 region of E307K+T391S-diol form at 0 ns, 7 ns, and 28 ns of the simulation.

The conformational changes of the dimerization1 region are due to the movement of Thr45, shown in Figure 4.25. After 5 ns Thr45 turns its sidechain upwards and stays in that position for 15 ns. Thr45 is a polar, hydroxylic residue and its movement can cause a disturbance on the electrostatic forces between the monomers. This disruption can cause ARSA octamers to dissociate in the lysosome.

Dimerization1 region was found to change its conformation throughout the simulations on mutant forms in neutral pH, too. If the mutant monomer forms can be transferred into the lysosome as monomers they may not be forming dimer forms due to these conformational changes also observed in acidic pH.

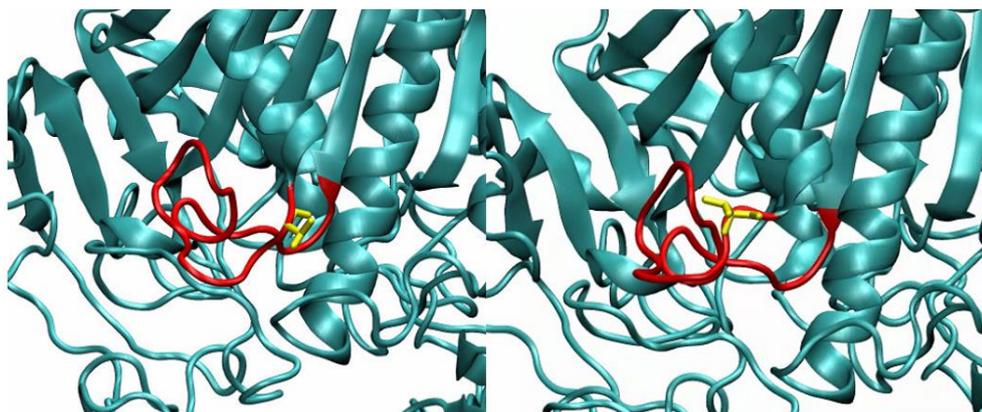


Figure 4.25. The conformational change of Thr45 (yellow) on the dimerization1 region (red) of E307K+T391S-diol mutant form.

Hydration state of FGL changes very fast from nonhydrated state to diol during the reaction but, during this period the stability of ARSA can be disrupted due to the disturbances on the oligomerization sites. In addition to these, examination of the trajectories of simulations on the two WT forms of ARSA showed that such significant conformational changes do not exist on Cys38 and Pro42 in these simulations. Cys38, Pro42 or Thr45 do not change conformation in WT-nonhydrated simulation. Thr45 turns its head from right to left (shown in Figure 4.26) in the simulation on WT-diol form of ARSA. This movement can be linked to the jump in the RMSD profile of the dimerization1 loop of WT-diol form around 13 ns in Figure 4.23. This change is not a major change like the movements that were observed in the mutant simulations as stated above.

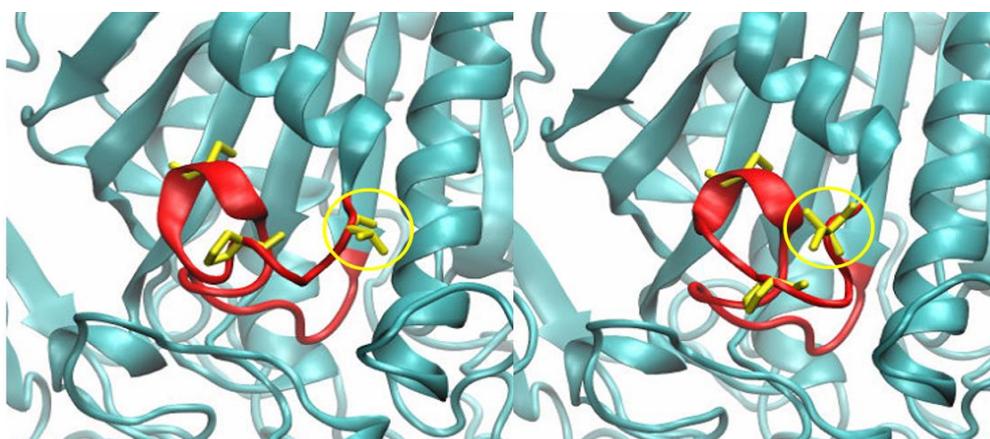


Figure 4.26. Movement of Thr45 (circled) on the dimerization1 loop (red) of WT-diol.

RMSD results of the dimerization1 region showed that fluctuations occur mostly in the simulations on E307K mutant for its both states while in the neutral pH E307K+T391S and W318C mutants show more significant deviations in their RMSD profiles for this region. It also supports the information of ARSA being more stable in acidic pH [102], and it is expected since ARSA oligomerizes in acidic milieu. Since the dimerization1 loop is not the only region on the dimerization site of ARSA we also inspected the dynamics of the other dimerization regions. The RMSD profiles of the dimerization2 region are given in Figure 4.27.

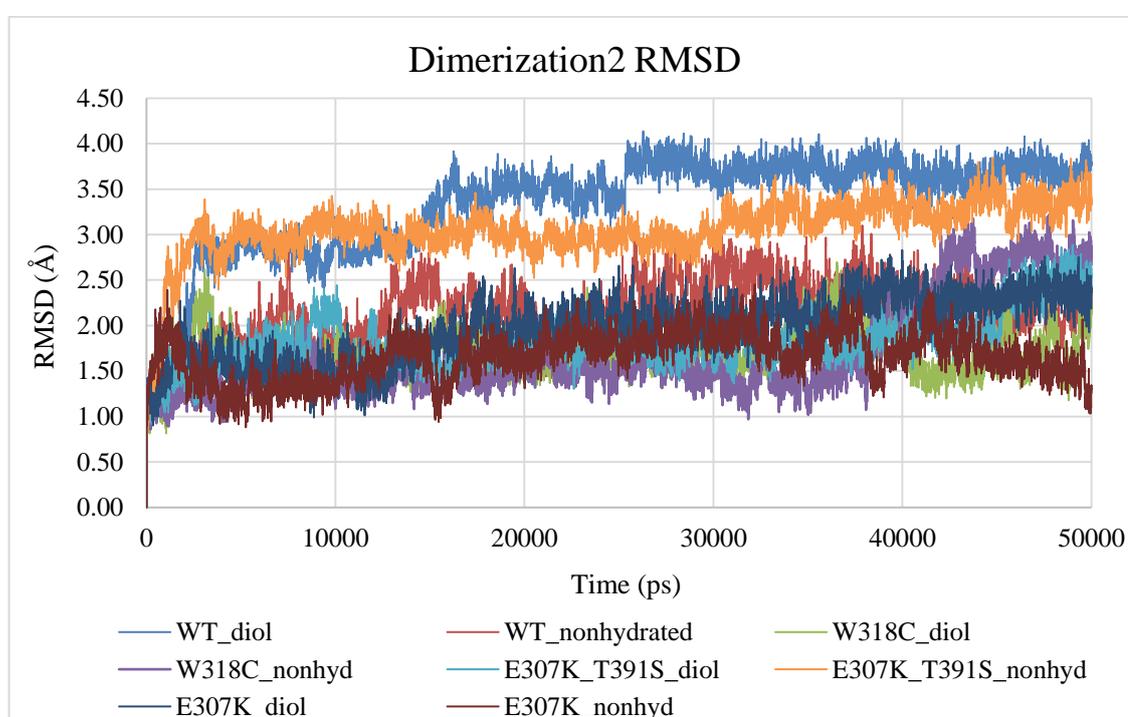


Figure 4.27. RMSD profiles of the dimerization2 regions from all simulations at acidic pH.

Dimerization2 region consists of 14 residues starting from the position 286 ending at 300. The highest RMSD profile of the dimerization2 region belongs to the WT-diol form of ARSA. This region reaches an equilibrium after 25 ns for the WT-diol form while the RMSD profiles of W318C-nonhydrated and E307K+T391S-diol forms do not reach an equilibrium in 50 ns. The state of FGL seems to affect the dimerization2 stability differently for the WT and mutant forms of ARSA. The difference between the RMSD profiles of this region is the most significant for WT forms of ARSA as shown in Figure 4.28.

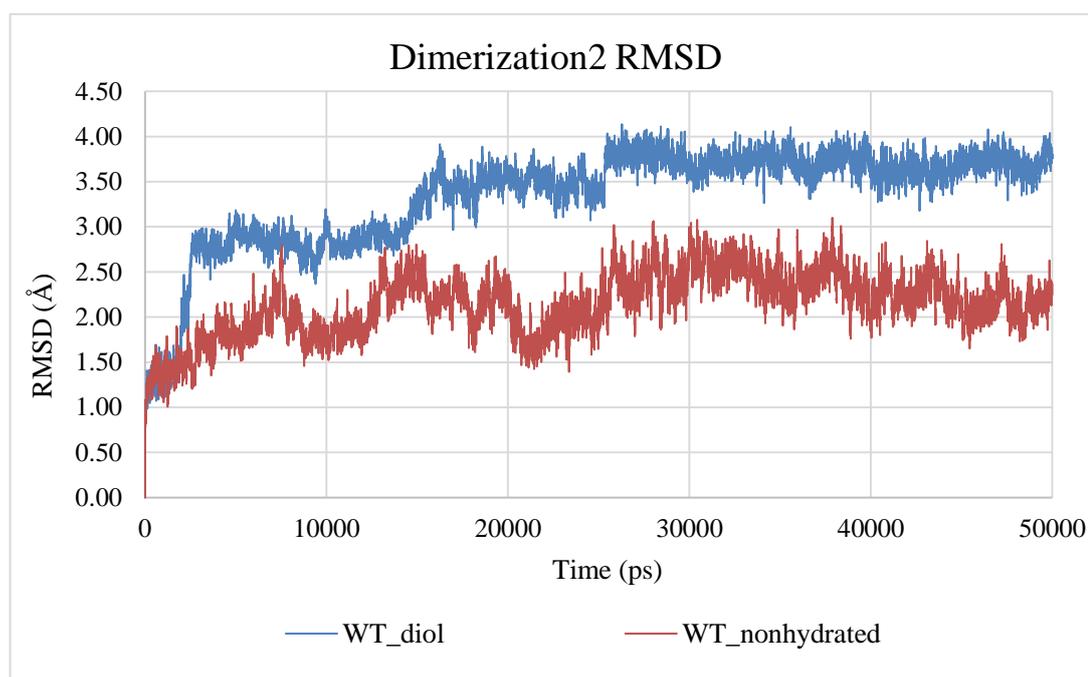


Figure 4.28. RMSD profiles of the dimerization2 regions of two WT forms of ARSA

There are 2 differences between the RMSD profiles of the dimerization2 region (residues: 286-300) of the two forms of WT ARSA monomers. These different movements of the dimerization2 region for the WT-diol and WT-nonhydrated form of ARSA are shown in Figure 4.29. One is the upper part of the loop indicated with black arrows as well as the left upper side of the loop, and the other is the turning angle of the bottom right part of this region (Figure 4.29). When we inspected the trajectories of WT simulations we saw that Cys294 turns its side chain down towards the solvents in the simulation on WT-diol form of ARSA. The similar movement was also observed in the WT-nonhydrated form of ARSA but its movement is minor compared to the WT-diol's. According to the conformational change of Cys294, loop part of the dimerization2 region changes its conformation as shown in Figure 4.29. The turns of the loops are indicated with black arrows.

This free cysteine, Cys294, forms a cysteine knot with the free cysteine of other monomer, Cys38 thus, this movement may be necessary for the dimer contact and stability. When the other simulations were analyzed this movement was not observed.

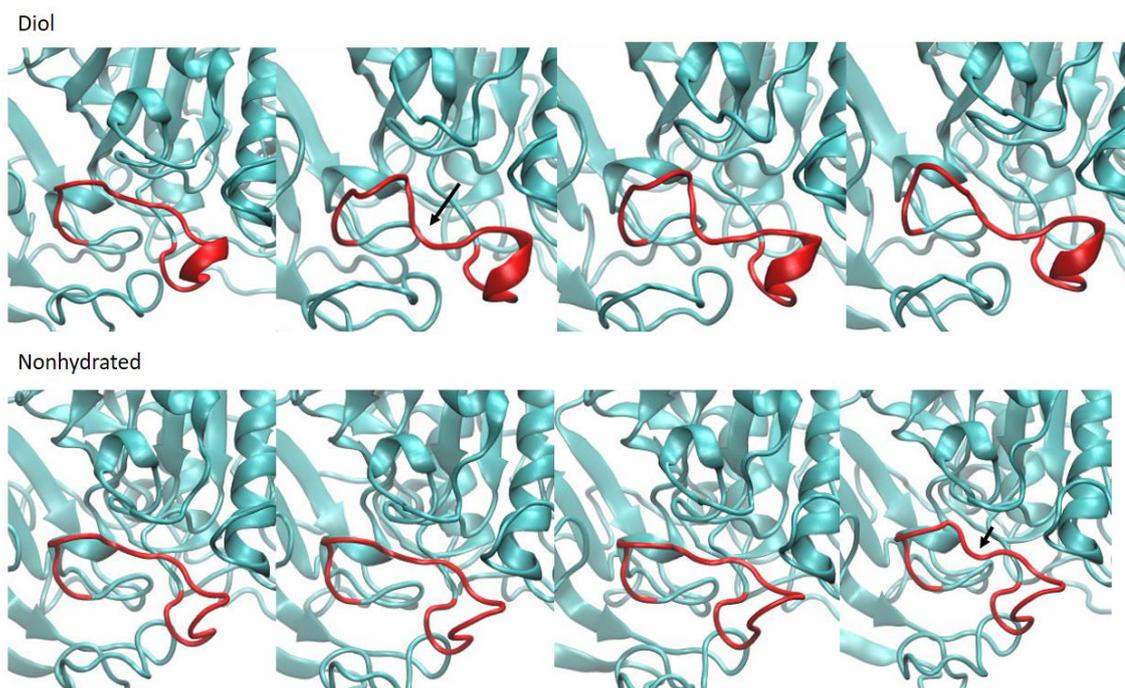


Figure 4.29. Dimerization2 regions of both WT ARSA forms (upper panel: diol and lower panel: nonhydrated) at 0 ns, 25 ns, 36 ns and 50 ns.

To understand the differences clearly, dimerization2 RMSD profiles are shown for diol and nonhydrated states of FGL in Figure 4.30 and Figure 4.31. In Figure 4.30, the results support the rigidity of Cys294 and dimerization2 region that all the diol type mutants' RMSD profiles follow a flat trend around 1.70 Å. In Figure 4.31, the highest RMSD profile belongs to the E307K+T391S-nonhydrated form of ARSA. None of the profiles of the nonhydrated types of ARSA seems to reach an equilibrium during the simulations. The jump in the profile of W318C-nonhydrated mutant around 40 ns can be linked to the movement of Met289 as shown in Figure 4.32 while other nonhydrated mutant forms do not show a sudden change in their RMSD trends for dimerization3 region. Their trajectories were also inspected, and the secondary structure of dimerization3 region or any conformational change of the residues on this region were not observed.

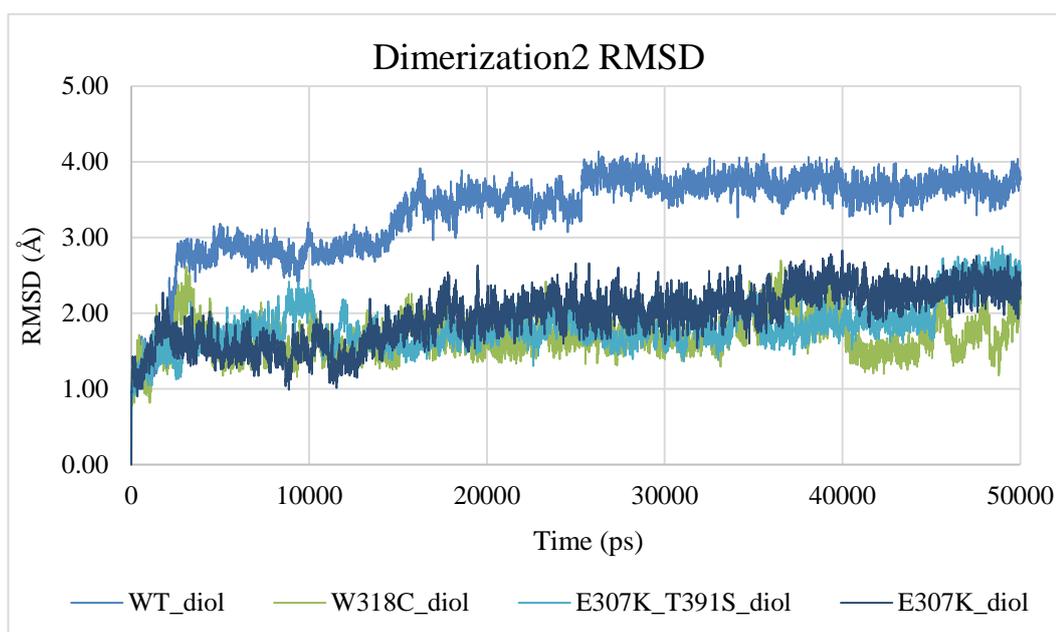


Figure 4.30. RMSD profiles of the dimerization2 regions of four diol forms of ARSA.

There is a significant difference in the RMSD profiles of WT-diol form and the other diol forms of ARSA in Figure 4.30. The profiles of WT-diol and E307K+T391S-diol mutant forms become closer but the difference between WT and W318C mutant forms become wider although W318C-diol forms shows a fluctuating trend after 40 ns. To understand the differences we also compared the nonhydrated forms as shown in Figure 4.31.

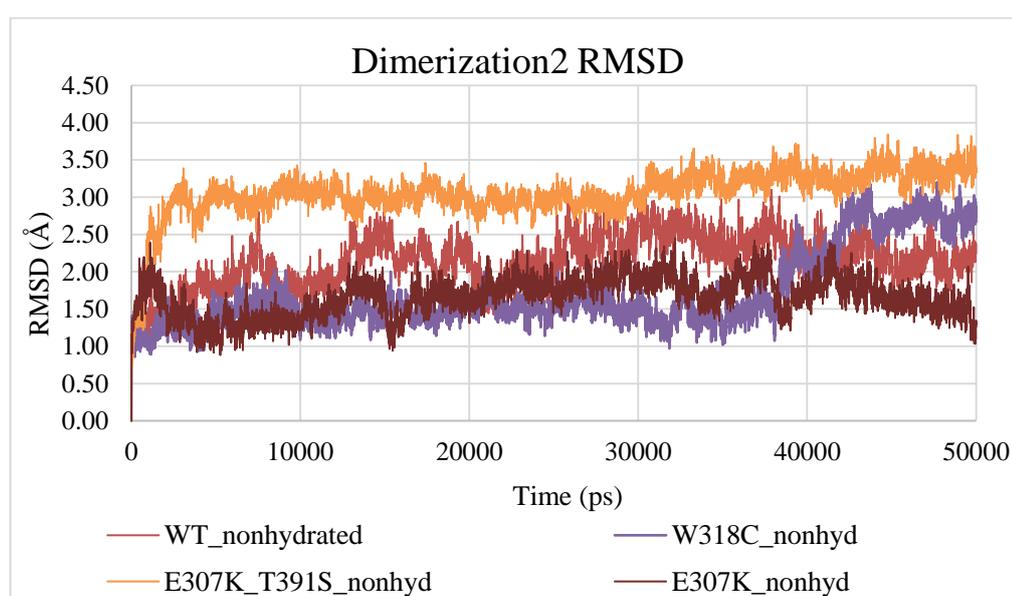


Figure 4.31. RMSD profiles of the dimerization2 regions of four nonhydrated forms

In the simulations on the nonhydrated states of the ARSA forms, the RMSD profile of the dimerization2 region of E307K+T391S-nonhydrated form follows a significantly higher trend than the other forms. For all simulations, the trends cannot reach an equilibrium. Even the profile of the WT-nonhydrated form starts to decrease slightly and constantly after the 35th ns of the simulation. These fluctuating trends can be linked to the movements of the free cysteine, Cys294, in the search of a disulfide link in this region.

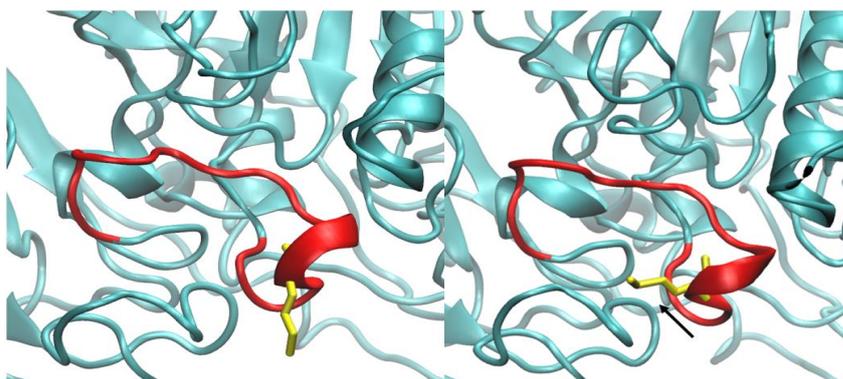


Figure 4.32. Met289 on the dimerization3 region of W318C-nonhydrated form of ARSA at 26 ns and 40 ns.

The jump in the RMSD profile of W318C-nonhydrated form is around 40 ns. As shown in Figure 4.32, Met289, which is a hydrophobic residue, turns its sidechain to left and the upper loop part of the dimerization2 region becomes flatter (indicated with a black arrow). These structural changes cause an increase in the RMSD profile. In this simulation on W318C-nonhydrated form Cys294 do not change its conformation. If its conformational change on the monomer form is necessary for the dimerization of ARSA, its lack can prevent the mutant proteins from oligomerization. On the other hand, other two nonhydrated state of mutant forms show a lower trend in their RMSD profiles for the dimerization2 region and when their trajectories were inspected we saw that in the simulation on E307K+T391S-nonhydrated form, Met289 turns its side chain to left as in the simulation on W318C-nonhydrated form (Figure 4.32) at the very beginning of the simulation, and this residue preserves its position like that. Also, after 32 ns, Cys294 changes its conformation as in the simulation on WT-nonhydrated form but this movement does not cause a major structural change on the loop part of the dimerization2 region (Figure 4.29). In the simulation on E307K-nonhydrated form, neither Cys294 nor Met289 change their conformations

throughout the simulation but Ser295 turns its side chain both towards the protein and the solvent. Its movements cause a minor folding on the loop part of the dimerization2 region.

Since the dimerization is very important for ARSA activity, the dynamics of the last part of the dimerization region, dimerization3 region, was analyzed. The RMSD profiles of the dimerization3 region of all forms are given in Figure 4.33.

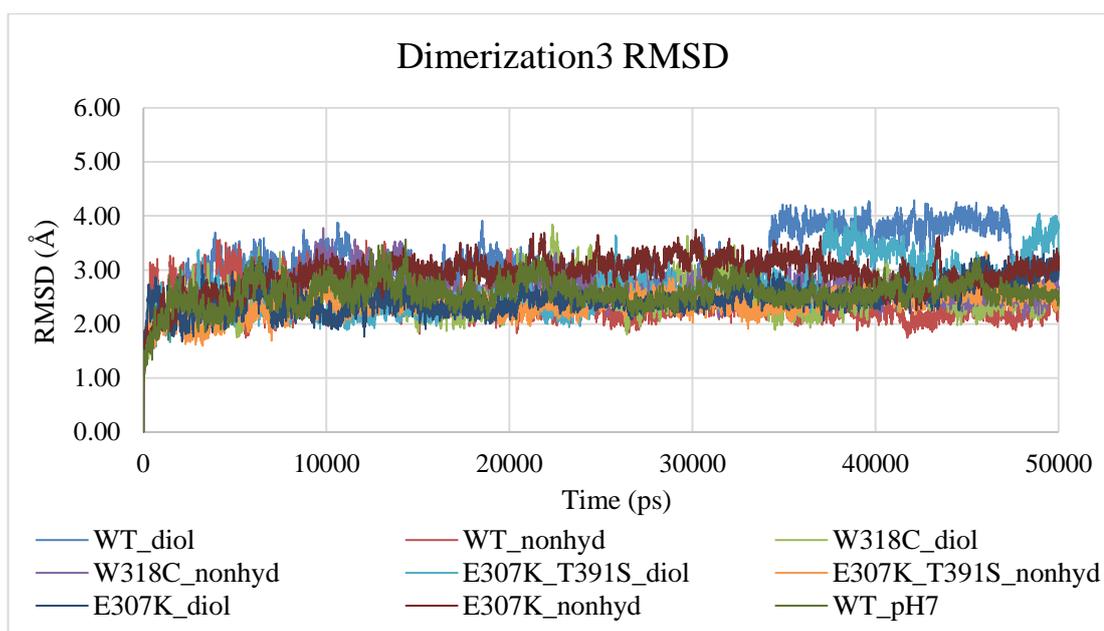


Figure 4.33. RMSD profiles of the dimerization3 regions of the all forms of ARSA.

The trends of the RMSD profiles of the dimerization3 regions are similar but there are jumps in the RMSD profiles of the diol forms of WT and E307K+T391S forms of ARSA around 35 ns in Figure 4.33. Only the diol forms of WT and E307K+T391S forms show the deviations while, in the nonhydrated forms of ARSA only the RMSD profile of the E307K-nonhydrated form cannot reach a plateau. When the trajectories of WT and E307K+T391S both diol forms of ARSA were inspected, we saw that residues: Thr409, Ala410, Asp411, Lys433, Gly443, Val445, Ala446, and Gly447 change their conformations leading to structural changes on this region as indicated with arrows in Figure 4.34.

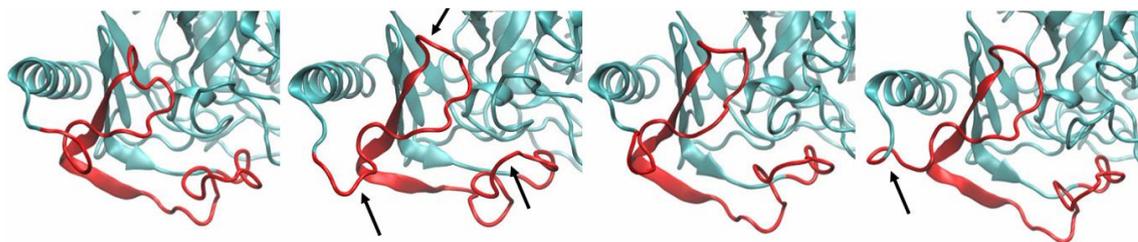


Figure 4.34. Dimerization3 region of E307K+T391S-diol form of ARSA.

Arrows in Figure 4.34 indicate the major changes on this region. They also show the locations of the residues: Thr409, Ala410, Asp411, Lys433, Gly443, Val445, Ala446, and Gly447. The screen shots were taken from 0 ns, 6 ns, 26 ns, and 43 ns of the simulation.

The deviations in the RMSD profile of E307K+T391S-diol mutant form between 35 and 46 ns are not drastic as the WT-diol form of ARSA's. When the trajectory of WT-diol form was inspected, we saw that the conformational change on the loop part indicated with the black arrow in the second image in Figure 4.35 is a consequence of the hydrogen bond formed between Thr409 and Asp411. This happens as the same in the simulation on WT-diol form.

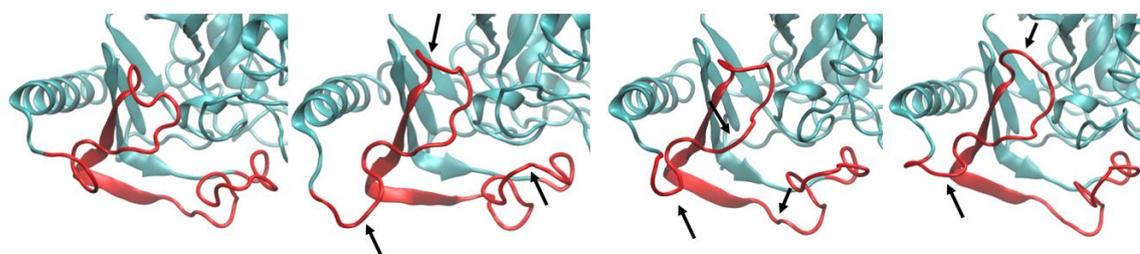


Figure 4.35. Dimerization3 region of WT-diol form of ARSA.

Arrows in Figure 4.35 indicate the major changes on this region. They also show the locations of the residues: Thr409, Asp411, Ser419, Lys433, Thr439, and Val445. In Figure 4.35 the dimerization3 region of WT-diol form of ARSA is shown for the same time frames (0 ns, 6 ns, 26 ns, and 43 ns). The step between 35 and 46 ns in the RMSD profile of WT-diol form can be linked to the severe movements of Val445 and Ala446 leading to winding and unwinding movement (Figure 4.36) of the Gly447 end of the dimerization3 region (left end of the red regions in Figure 4.35).

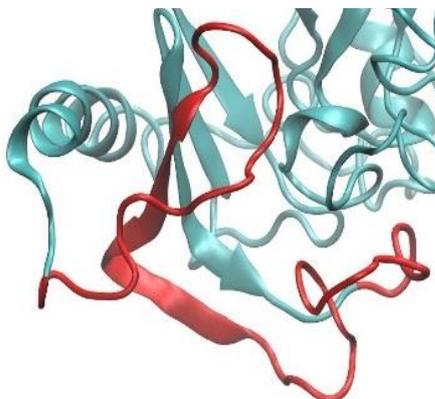


Figure 4.36. Dimerization3 region of WT-diol form at 35 ns.

In the monomer form of diol state, these fluctuations can be expected since the average of the RMSD values of the simulations run in acidic pH are higher than the ones in neutral pH. From the results for dimerization3 region, the outcome can be that this region is very flexible and its movements are independent from the mutations. These structural changes can be unimportant for the stability of the dimer form, because ARSA is expected to be more stable in the acidic environment [102]. On the other hand, hydration of FGL can depend on the oligomerization state of ARSA. In other words, to be a stable diol form, ARSA may require forming its octameric structure, and naturally its dimer form first. The stability of the diol ARSA forms may be better analyzed after the oligomeric structure is obtained thus the fluctuations on the specific regions for diol ARSA monomers may be conceivable. Another result is that the RMSD profile of the WT-nonhydrated form shows the lowest trend and after 35 ns it reaches an equilibrium around 2.1 Å. The difference between the RMSD profiles of the WT-diol and WT-nonhydrated forms are shown in Figure 4.37 along with the profile of the WT form at pH 7 for comparison. After 35 ns, their average RMSD values are 3.65 Å for WT-diol and 2.21 Å for WT-nonhydrated taking the first frame coordinates as reference. The average RMSD values for the same region for the last 15 ns become 1.97 Å for both WT-diol and WT-nonhydrated forms when taking the 35th ns coordinates as the reference coordinates. Also, when the coordinates at 35 ns were taken as reference, the RMSD profile of the WT-nonhydrated form was observed to be in a slight constant increase. These results show that this region even in the simulation on WT-nonhydrated form may not reach an equilibrium and can be considered very flexible for all cases.

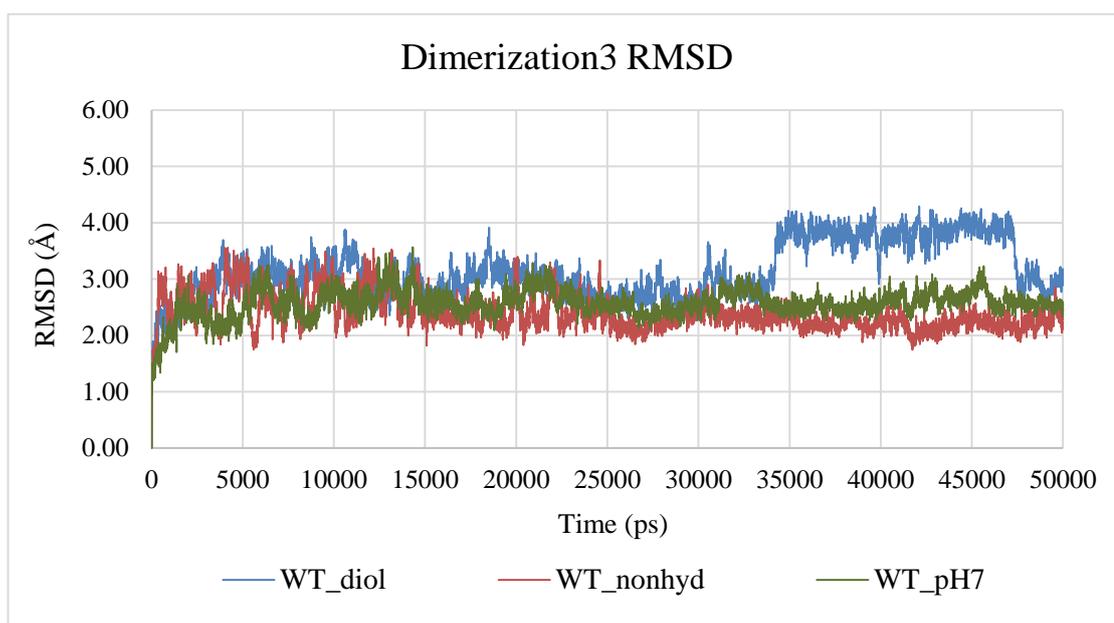


Figure 4.37. RMSD profiles of the dimerization3 regions of WT structures: two at acidic pH and one at neutral pH

The results from Figure 4.37 also show that the RMSD profiles of the dimerization3 region follow similar trends at first but in the trajectory of the WT-diol, it deviates from the average for 10 ns. This change may not be important since this region is flexible as so in the neutral pH simulations, and most of the RMSD profiles follow a similar trend.

Another important region on ARSA is its octamerization helix. It mediates the octamerization by hydrophobic attractions thus, its stability probably plays a crucial role in the stability of the oligomeric structure. The RMSD calculations were also performed for this helix (Pro450-Val470) and the RMSD profiles of the octamerization helix from the all simulations in acidic pH are shown in Figure 4.38. When we look at the results we see that the largest gap between the trends of the profiles is between the profiles of E307K+T391S-nonhydrated form and WT-diol form of ARSA. WT-nonhydrated form has a fluctuating trend between 20 and 40 ns but it reaches an equilibrium after 40 ns around 2.9 Å. The RMSD profile of the WT-diol form follows a flat trend after 25 ns around 2.3 Å.

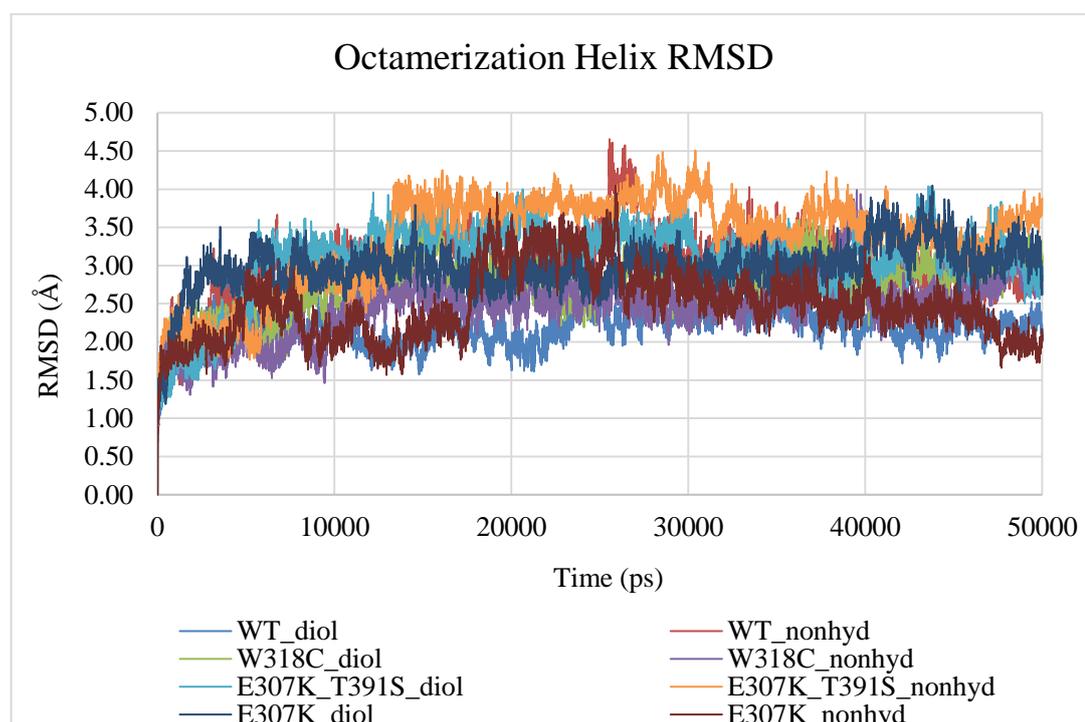


Figure 4.38. RMSD profiles of the octamerization helices of all forms of ARSA.

When the trajectory of WT-diol form was inspected, we saw that octamerization helix structure do not change but only minor differences. The RMSD profile of the WT-nonhydrated form follows a similar trend with the RMSD profile of WT ARSA in neutral pH and reaches a plateau around 2.0 Å as in the neutral simulation on WT form. The general trends of the profiles show that octamerization helix becomes more mobile in acidic pH, in which ARSA monomers tend to form octameric structure. This result also shows that octamer helix, especially the Val470 end of the helix, needs ARSA monomers to form the octameric structure to stay stable. For E307K+T391S-nonhydrated form, octamerization helix changes its secondary structure on the Val470 end of the helix significantly. The movements of this region are shown in Figure 4.39.

The structural changes shown for 0 ns, 19 ns, 27 ns, and 50 ns of the simulation in Figure 4.39 is linked to the conformational changes of Leu466. It makes almost a 180° turn towards the solvent. Leu466 is an aliphatic and hydrophobic residue. This exposure to solvent phase can cause a disruption in the hydrophobic interactions between the two octamer helices of the ARSA monomers.



Figure 4.39. Structural changes on the octamerization helix of E307K+T391S-nonhydrated mutant form of ARSA.

When we inspected the trajectory of E3007K+T391S-nonhydrated form we saw that there may be a connection between the movements of mutant Lys307 and this movement of Leu466. Their turns and fluctuations are shown for several time frames of the simulation in Figure 4.40.

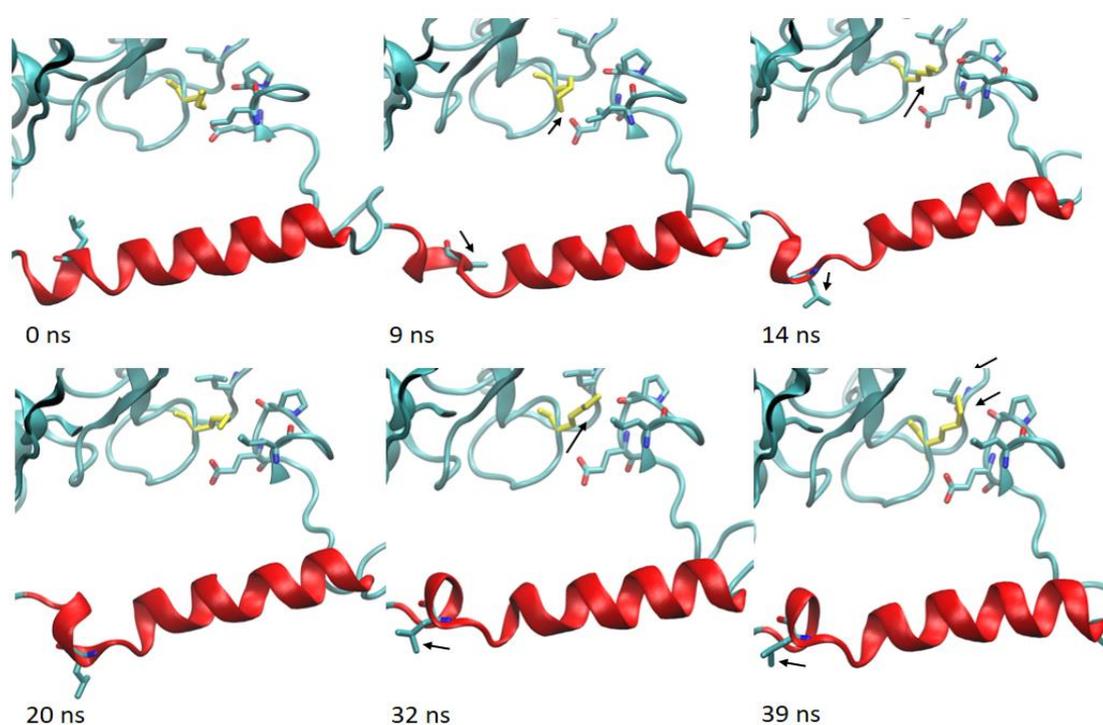


Figure 4.40. Leu466 and Lys307 (yellow) are indicated with black arrows at 0 ns, 9 ns, 14 ns, 20 ns, 32 ns, 39 ns of the simulation on E307K+T391S-nonhydrated form.

These movements cannot be correlated but the general trend shows that as the mutant Lys307 residue moves toward Leu298, Leu466 seems to turn towards the solvent. Also, Lys307 is surrounded with hydrophobic residues like Pro435, Leu437 and Leu431. The significant electrical change from glutamic acid to lysine in this region may cause an electrostatic disruption. This disruption may act like a force on the backbone throughout the protein and cause a conformational change of Leu466. In the simulations performed at neutral pH, the octamerization helix of E307K+T391S was observed to deviate most from its initial structure among the three simulations. The secondary structure change is not severe as it is in the simulation on E307K+T391S-nonhydrated form at acidic pH, but again Leu466 at the neutral pH simulation turns its side chain towards the solvent and then again towards the protein center. These results show that E307K+T391S mutant form of ARSA may not form octameric structure at acidic pH due to mutations.

First, we started our comparison by examining the octamerization helix of WT-nonhydrated form of ARSA. The conformational changes and the related movements of the residues on the octamerization helix of WT-nonhydrated form are shown in Figure 4.41 and Figure 4.42. Second, to understand the effect of the hydration state of FGL69, the RMSD profiles of the octamerization helices of two states of the ARSA forms are plotted separately as in Figure 4.43 and Figure 4.44. WT-diol ARSA monomer has a more stabilized profile than its nonhydrated form. It also has the lowest trend among the diol forms of ARSA. The sharp changes in the profile of WT-nonhydrated form between 25 and 30 ns can be explained by the movement of Asp467 and the turning of Leu466 by extension. Also, the helix gains a bent structure around the 25th ns of the simulation on WT-nonhydrated form due to the tension on dimerization3 region. These movements of the helix and the two residues (Leu466 and Asp467) are shown in Figure 4.41 and Figure 4.42, respectively.

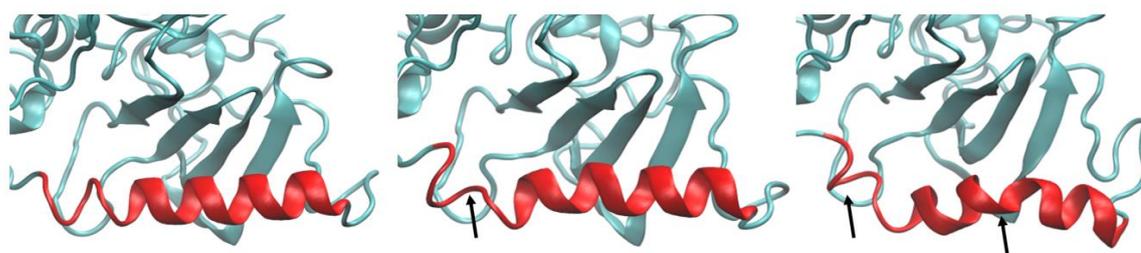


Figure 4.41. Octamerization helix of WT-nonhydrated form at 0 ns, 4 ns and 26 ns.

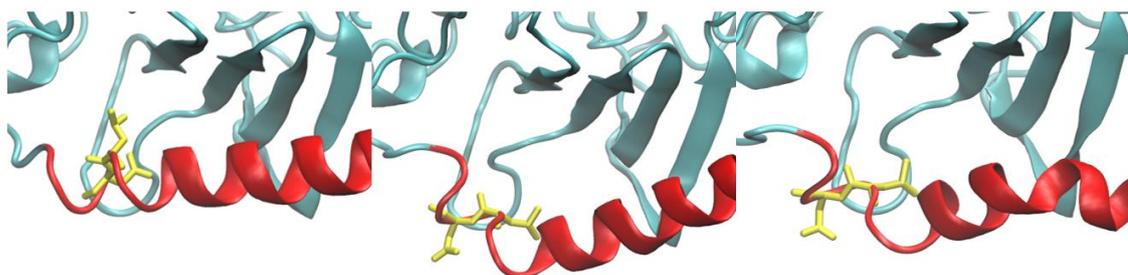


Figure 4.42. Leu466 (right yellow residue) and Asp467 (left yellow residue) on the octamerization helix (red) of WT-nonhydrated form at 0 ns, 18 ns and 26 ns.

The conformational changes shown in Figure 4.42 are different than those shown before for the octamerization regions of mutant types of ARSA in a way that Leu466 do not become solvent exposed and only Asp467 turns its side chain towards the solvent. Aspartic acid (Asp467) is a charged residue and its exposure to solvent may not cause any distribution in the electrostatic balance between the dimers. Asp467 can form various hydrogen bonds with the surrounding water molecules. This is the simulation on WT form of ARSA in its nonhydrated form at acidic milieu, we can conclude that these types of movements of Leu467 and its surrounding may not cause a hindrance in the octamerization of ARSA as long as they do not cause Leu467 to turn its side chain toward the solvent. To gain a better understanding the effects of the hydration state of FGL and the mutations, dimer simulations should be performed and analyzed to see whether the dimerization is also affecting the behavior of the residues on the octamerization helix. For the monomer part of our simulations we also showed the RMSD profiles of the octamerization helix for two states of the ARSA monomer in Figure 4.43 and Figure 4.44 to analyze the effect of the hydration state of FGL on the stability of the helix.

The main difference between Figure 4.43 and Figure 4.44 is that nonhydrated forms of ARSA have higher trends and not stabilized profiles for the octamerization helix. This can be expected since ARSA should be in its octameric form for catalysis, and during the reaction FGL becomes hydrated for a short time.

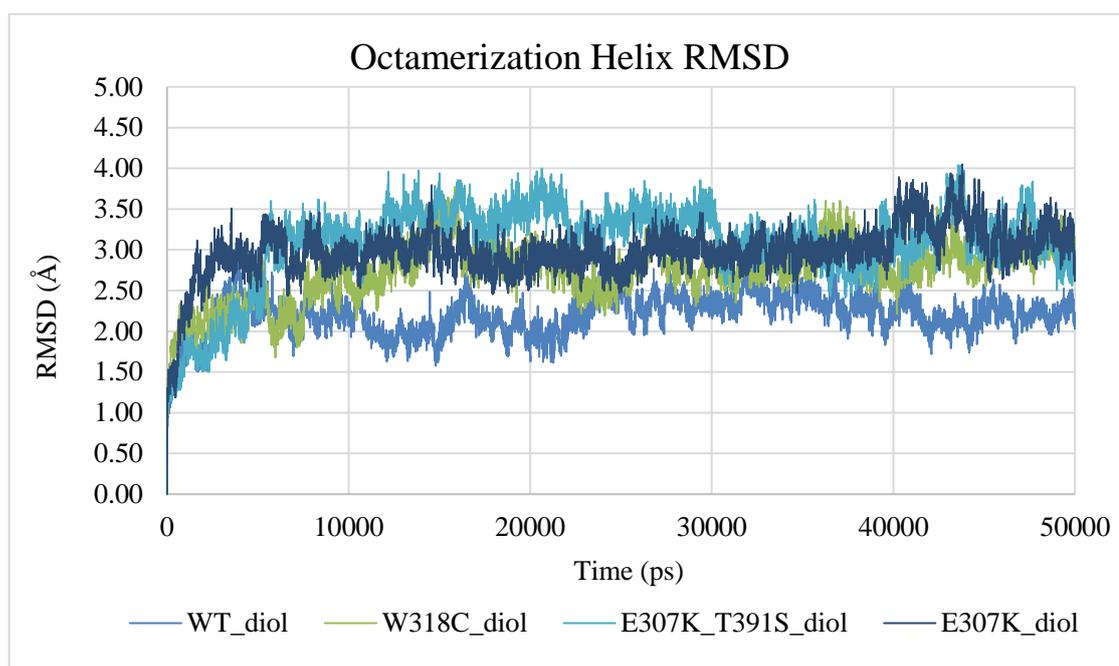


Figure 4.43. RMSD profiles of the octamerization helices of all forms of ARSA with FGL residues in diol state.

For both states, E307K+T391S shows the highest RMSD profile with sudden decreases and increases in its trend. On the other hand, E307K single mutant form in its nonhydrated state shows a sharply fluctuating profile but it has the lowest trend among nonhydrated forms. Even though the diol forms have lower and in equilibrium trends when we compare the WT and mutant forms we see that the profiles of the mutant forms of ARSA follow a higher trend than that of the WT's. In addition to these the fluctuations after 30 ns in Figure 4.43 are more significant in the diol forms of the mutants. Considering these outcomes, we can say that mutant ARSA dimers may not form octameric structure after being transferred into the lysosome and this situation may cause the inactivity of the enzyme. On the other hand, this situation does not hold for the nonhydrated forms of ARSA. According to the trends of the RMSD profiles of the octamerization helices we can conclude that this region becomes more mobile and flexible at acidic pH depending on the hydration state of FGL.

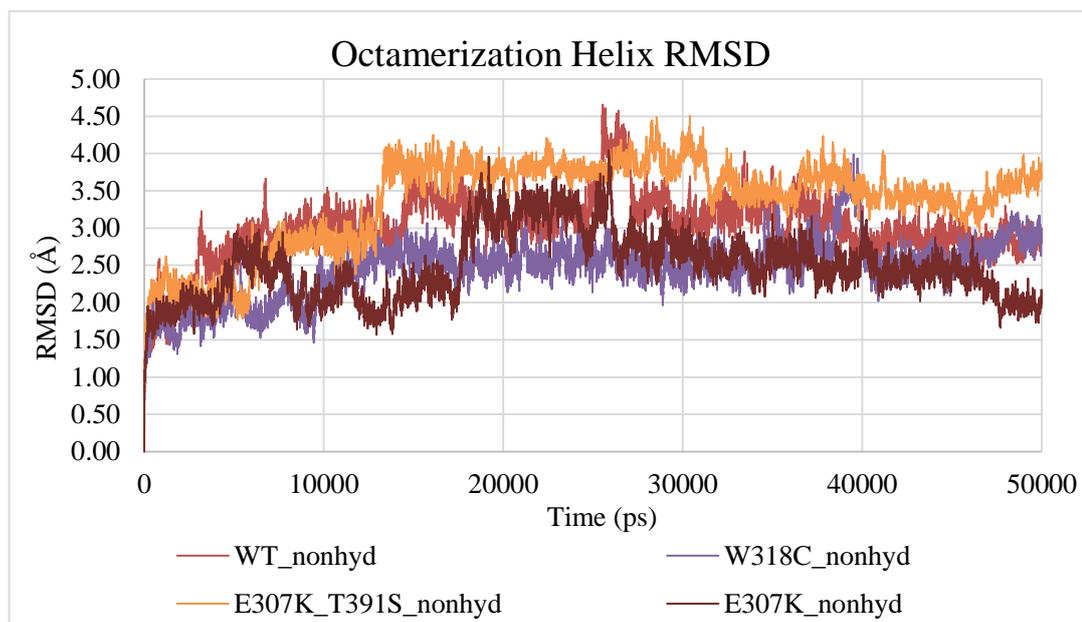


Figure 4.44. RMSD profiles of the octamerization helices of all forms of ARSA with FGL residues in nonhydrated state.

W318C mutant forms show similar trends for their octamerization helices in both pH. Only the nonhydrated form of W318C mutant at acidic pH has a peak on its RMSD profile around 37-40 ns and after 47 ns its trend starts to rise again. This change occurs due to the movements of Leu466 and Lys463 on the octamerization helix of W318C-nonhydrated mutant form. Leu466 turns its side chain towards the center of the protein causing the last turn of the helix unwind. In addition to this, Lys463 turns its side chain from left to right and forms a hydrogen bond with Glu424 on the loop at the back of the octamerization helix. It also forms hydrogen bond with Gln460. Glu424 is the key residue in the oligomerization state of ARSA. It forms an intramolecular hydrogen bond with Gln460 which favors the dimer form of ARSA, whereas in the octamer state, Glu424 chooses to form an intermolecular hydrogen bond with Phe398. The movement of Lys463 towards Glu424 makes the turns at the Val470 end of the octamerization helix unwind and causes an increase in the RMSD profile of W318C-nonhydrated form. On the other hand, W318C-diol form has a stable RMSD profile which means that its octamerization helix shows a more rigid profile throughout the simulation. If the hydration of FGL makes the octamer structure more stable for the reaction, this outcome should be expected.

Lastly, the RMSD profiles of the Asp335-Arg370 region were calculated and depicted in Figure 4.45. This region includes alpha helices and loops which make it compact and flexible at the same time. This region generally preserves its secondary structure throughout the all simulations. It includes the N-glycosylation region (Asn350) and its importance has been attributed [102]. In the both neutral and acidic simulations, this region remains as one of the least fluctuating regions.

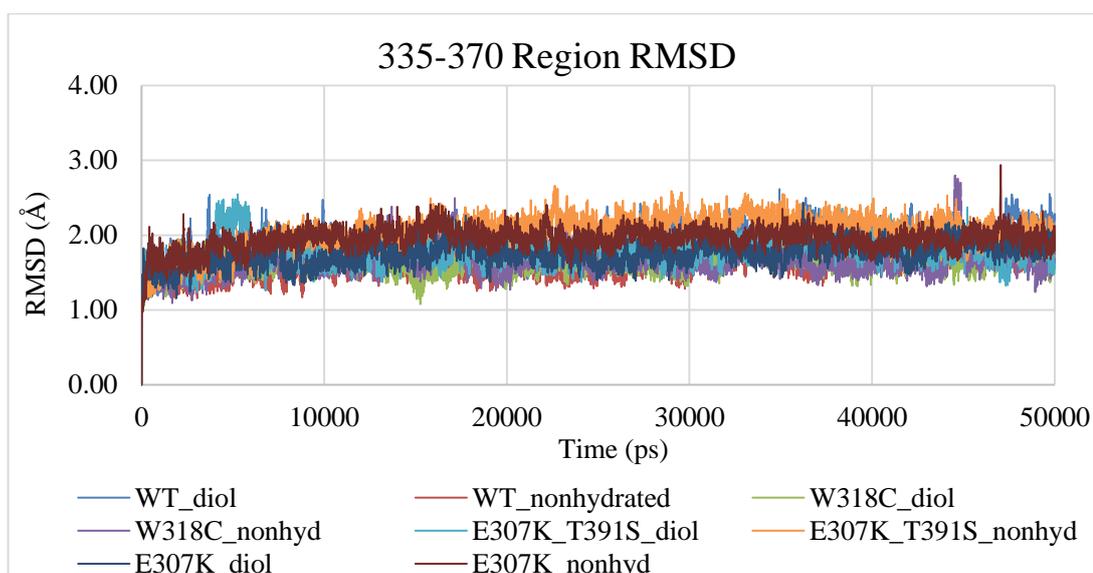


Figure 4.45. Asp335-Arg370 region RMSD profiles of all simulations at acidic pH

The trends of the RMSD profiles of the Asp335-Arg370 regions are similar for all simulations including the simulations in neutral milieu. They reach an equilibrium fast and the fluctuations are not significant for the most of the RMSD profiles. Since the glycosylation is the step before the transfer of the enzyme into the lysosome, the simulations in neutral pH have a higher significance for the stability of this region. The trajectories of the simulations in acidic pH were analyzed based on the trends of the RMSD profiles for Asp335-Arg370 region and the changes on the structures were detected accordingly. In the RMSD profiles of the nonhydrated forms of W318C and E307K mutant forms there are sudden sharp increases around 45 and 47 ns. The jump on the 45<sup>th</sup> ns of the simulation on W318C-nonhydrated form can be linked to the conformational change on Thr352 which is shown in Figure 4.46.

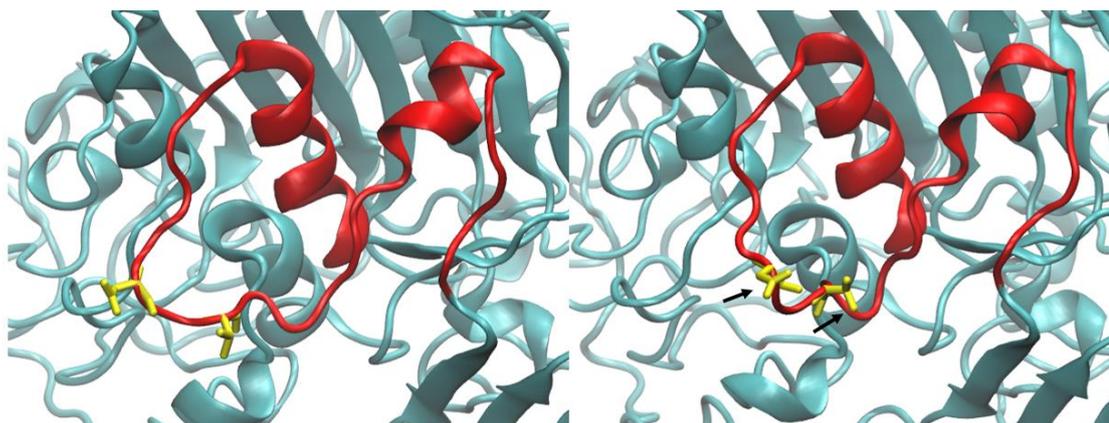


Figure 4.46. Thr352 (right yellow residue) and Asn350 (left yellow residue) at 14th ns and 45th ns of the simulation on W318C-nonhydrated mutant.

When we look at the trends of the RMSD profiles for two states of E307K+T391S mutant form we see that around 5 ns there is a jump in the profile of E307K+T391S-diol form while the profiles of the nonhydrated form gradually increases for 35 ns and then, constantly decreases for the rest of the simulation as shown in Figure 4.47.

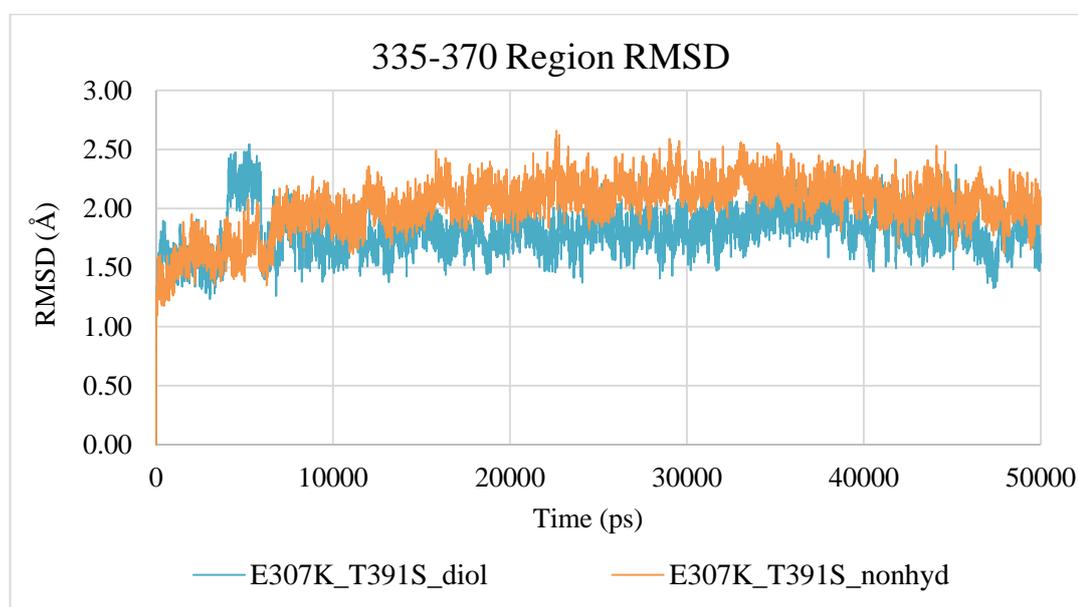


Figure 4.47. RMSD profiles of the Asp335-Arg370 region for the two states of E307K+T391S mutant.

The jump around 5 ns in the RMSD profile of E307K+T391S-diol mutant form can be explained with the conformational change of Thr365. It turns its side chain from left to right as shown in Figure 4.48 at 5th ns of the simulation on E307K+T391S-diol form. This happens because Leu363 come between Thr365 and Pro360 and pushes Thr365 away for 2 ns. This movement causes a secondary structure change in this region as in Figure 4.48 and a jump in the RMSD trend.

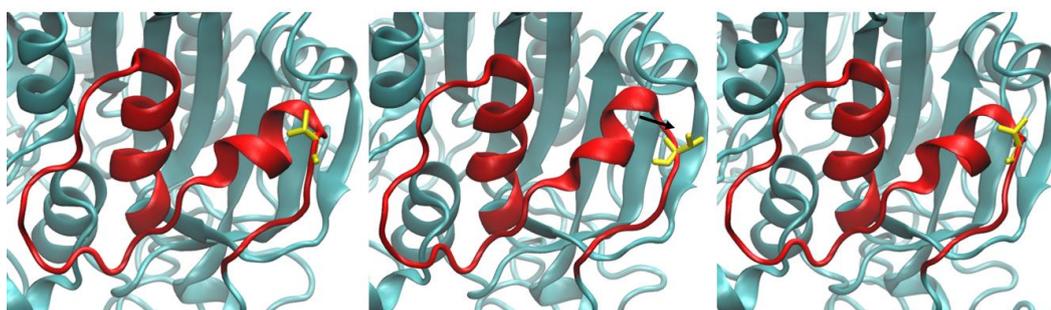


Figure 4.48. Thr365 (yellow) is shown on the Asp335-Arg370 region of E307K+T391S-diol mutant.

The arrow in Figure 4.48 indicates the direction of the sidechain and the snapshots are given for the times: 0 ns, 5 ns, 43 ns from left to right. This movement of Thr365 in the simulation on E307K+T391S-diol mutant does not affect the conformation of Asn350 as in the simulation on W318C-nonhydrated mutant form, thus this movement may not be critical or be the reason of any hindrance in the glycosylation of this mutant form. The jump in the RMSD profile of E307K+T391S mutant can only be a random fluctuation of Thr365 since it gains its initial conformation back and remains its position for the rest of the simulation on E307K+T391S-diol form after the 6<sup>th</sup> ns of the simulation. The state of FGL in the mutant E307K+T391S seems not affecting the trend of the RMSD profiles. For E307K+T391S mutant, they follow a similar trend after 7 ns. When we observed the trajectory of the nonhydrated form of E307K+T391S mutant form, we did not see any significant change on the Asp335-Arg370 region.

The jump for one instance in the RMSD profile (after 45 ns) of the E307K-nonhydrated form was neglected since its trend has reached an equilibrium after 15 ns, and preserves its profile even after this jump. This mutant form was analyzed to understand the additional

effect of the polymorphism, T391S; but for Asp335-Arg370 region, any significant effect of the polymorphism could not be observed. In all the simulations on monomer forms of ARSA Asp335-Arg370 region was left unbound to any sugar entity. The overall results showed that the acidity did not affect the stability, in other words, RMSD results of this region. In addition, the average RMSD values of the whole protein and of the specific regions are given in Table 4.8. When we compare the results for the whole trajectories, we see some significant difference in the average RMSD results of the mutant structures for the active site and the dimerization sites of the enzyme.

Table 4.8. Average RMSD values with their corresponding standard deviation values at acidic pH.

	<b>Overall Backbone</b>							
	<b>WT</b>		<b>W318C</b>		<b>E307K+T391S</b>		<b>E307K</b>	
	diol	nonhyd	diol	nonhyd	diol	nonhyd	diol	nonhyd
Avg	1.59	1.53	1.44	1.40	1.50	1.59	1.66	1.49
$\pm$ Std. Dev.	0.15	0.13	0.10	0.12	0.15	0.17	0.22	0.13
	<b>Active Site</b>							
	<b>WT</b>		<b>W318C</b>		<b>E307K+T391S</b>		<b>E307K</b>	
	diol	nonhyd	diol	nonhyd	diol	nonhyd	diol	nonhyd
Avg	1.06	1.53	0.99	0.81	1.00	1.26	1.13	1.01
$\pm$ Std. Dev.	0.19	0.13	0.07	0.09	0.08	0.22	0.09	0.09
	<b>Octamerization Site</b>							
	<b>WT</b>		<b>W318C</b>		<b>E307K+T391S</b>		<b>E307K</b>	
	diol	nonhyd	diol	nonhyd	diol	nonhyd	diol	nonhyd
Avg	2.20	3.01	2.74	2.46	3.09	3.36	2.98	2.48
$\pm$ Std. Dev.	0.24	0.44	0.35	0.40	0.50	0.59	0.31	0.45
	<b>Dimerization1 Site</b>							
	<b>WT</b>		<b>W318C</b>		<b>E307K+T391S</b>		<b>E307K</b>	
	diol	nonhyd	diol	nonhyd	diol	nonhyd	diol	nonhyd
Avg	1.51	1.72	1.26	1.93	1.88	1.98	2.51	2.57
$\pm$ Std. Dev.	0.40	0.12	0.36	0.34	0.38	0.22	0.36	0.46
	<b>Dimerization2 Site</b>							
	<b>WT</b>		<b>W318C</b>		<b>E307K+T391S</b>		<b>E307K</b>	
	diol	nonhyd	diol	nonhyd	diol	nonhyd	diol	nonhyd
Avg	3.34	2.14	1.70	1.70	1.82	3.06	1.96	1.66
$\pm$ Std. Dev.	0.57	0.35	0.24	0.52	0.29	0.32	0.36	0.26

Table 4.8. Average RMSD values with their corresponding standard deviation values at acidic pH (cont.).

	<b>Dimerization3 Site</b>							
	<b>WT</b>		<b>W318C</b>		<b>E307K+T391S</b>		<b>E307K</b>	
	diol	nonhyd	diol	nonhyd	diol	nonhyd	diol	nonhyd
Avg	3.14	2.40	2.50	2.61	2.63	2.37	2.47	2.92
$\pm$ Std. Dev.	0.50	0.32	0.31	0.27	0.45	0.23	0.26	0.31
	<b>335-370 Site</b>							
	<b>WT</b>		<b>W318C</b>		<b>E307K+T391S</b>		<b>E307K</b>	
	diol	nonhyd	diol	nonhyd	diol	nonhyd	diol	nonhyd
Avg	1.87	1.62	1.65	1.66	1.81	2.02	1.75	1.95
$\pm$ Std. Dev.	0.21	0.16	0.14	0.18	0.19	0.23	0.15	0.15

From the results in Table 4.8, we can conclude the following:

- The most rigid active site structures were observed in the simulations on both states of W318C mutant form of ARSA. This was not observed at the neutral pH-simulation.
- The stability (especially the Val470 end stability) of the octamerization helix seems to be affected by the hydration state of FGL in the WT form of ARSA which can be linked to the octamerization affinity of the enzyme. The Val470 end of the octamerization helix becomes more mobile due to the mutations E307K and T391S, but when the mutation E307K is present alone, the diol state of this mutant exhibits less rigid structure on the octamerization helix. The difference between the states of the mutants are similar also for W318C mutant form but it is less significant than the difference between the states of E307K only mutant.
- Dimerization1 region has the highest RMSD values for the mutant E307K in both cases. The difference is significant when the results for this region are compared to the WT's. On the other hand, W318C mutant seems to have the least average RMSD values of the dimerization1 region for its both states. W318C mutation may be affecting the folding which causes ARSA monomers to be retained in ER because the stability of the whole protein and of the specific regions become increased in the case of the simulations on W318C mutant in acidic pH. In the neutral pH, average RMSD

results of the mutant forms are greater than the ones at the acidic pH. It is the same for the WT form of ARSA even though the difference is smaller.

- There are significant differences between the average RMSD values of the dimerization<sup>2</sup> regions of ARSA forms. The hydration state of FGL seems to affect the stability of this region considerably for WT and E307K+T391S mutant forms in an opposite manner. The movements of Cys294 were discussed above and they were concluded to be important for dimer contact. The main outcome can be that the movement of Cys294 is necessary thus an increase in the RMSD profile may be necessary for this region for dimerization. E307K+T391S mutant form, showing the opposite of the trend, may not be able to oligomerize or to maintain its dimer form during catalysis. To understand the effects of the mutations on dimerization, dimer simulations should be analyzed.
- Average RMSD values for the dimerization<sup>3</sup> region are similar to the results for dimerization<sup>2</sup> region especially of the WT forms. The highest average RMSD results belongs to the WT-diol form of ARSA. This outcome can be interpreted as the fluctuations in the monomer form are expected for this region. The conformational changes in the simulation on WT-diol form can be seen also in the simulations on dimer and octamer forms. To gain a better understanding simulations on oligomeric structures should be performed.
- The results for the octamerization site show that the nonhydrated form of WT ARSA and two forms of E307K+T391S mutant ARSA have the highest average RMSD results. We can conclude that E307K and T391S mutants may hinder the octamerization when they are present together. When we look at the results of only E307K mutant, its diol form also shows a fluctuating octamerization region for this mutant. WT-nonhydrated form also has a very high average RMSD for the octamerization helix. The hydration state can be efficacious on the stability of the octamer structure as the mutations.
- We can conclude that E307K+T391S-nonhydrated mutant form has higher average RMSD values for almost all regions of the enzyme. Especially for the dimerization<sup>2</sup>

region, the difference between two states of E307K+T391S becomes very significant (1.2 Å). On the other hand, for W318C mutant the results do not differ at all with changing hydration state of FGL.

- Single E307K mutant shows the highest difference of the average RMSD results for the dimerization1 region. We can say that the effect of E307K mutation is enhanced by the addition of T391S mutation causing more mobile structures on the specific regions of ARSA.

### 4.3.3. RMSF Calculations

Mobility of the individual residues are important as much as the stability of the specific regions in MD simulations. Since the movements of the specific residues in dimerization or octamerization sites can affect the oligomerization of ARSA, we examined the movements of the all residues individually in terms of RMSF calculations to understand the regional effects of the mutations on the mobility of the residues because the side chains are generally more mobile than the backbone. The results are shown in Figure 4.49. The residues with RMSF values higher than 2.0 Å were considered as very mobile. The RMSF values were also scaled to B-factor and the results are given with the experimental B-factor values of ARSA for comparison. Coordinates of the carbon alpha atoms of the residues were used in the calculations. Average RMSF values of the carbon alpha atoms of the whole protein and the specific regions are given in Table 4.9.

From Figure 4.49, the significantly fluctuating regions can be regarded as His41-Thr45 for nonhydrated form of W318C and diol form of E307K+T391S mutants, Gln465- Ala469 for nonhydrated state of E307K+T391S mutant, and Gly171-Asn173 for nonhydrated form of WT ARSA. The later built part of the protein (Gly444-Gly447) has the highest RMSF results within each form of ARSA but the diol states of WT, E307K+T391S and E307K forms of ARSA have the highest peaks for this region. It may be a result of the exposure of this region in simulation on monomer forms and may be expected. The hydrogen bond occupancies of these residues with high RMSF results will be evaluated in terms of structural stability.

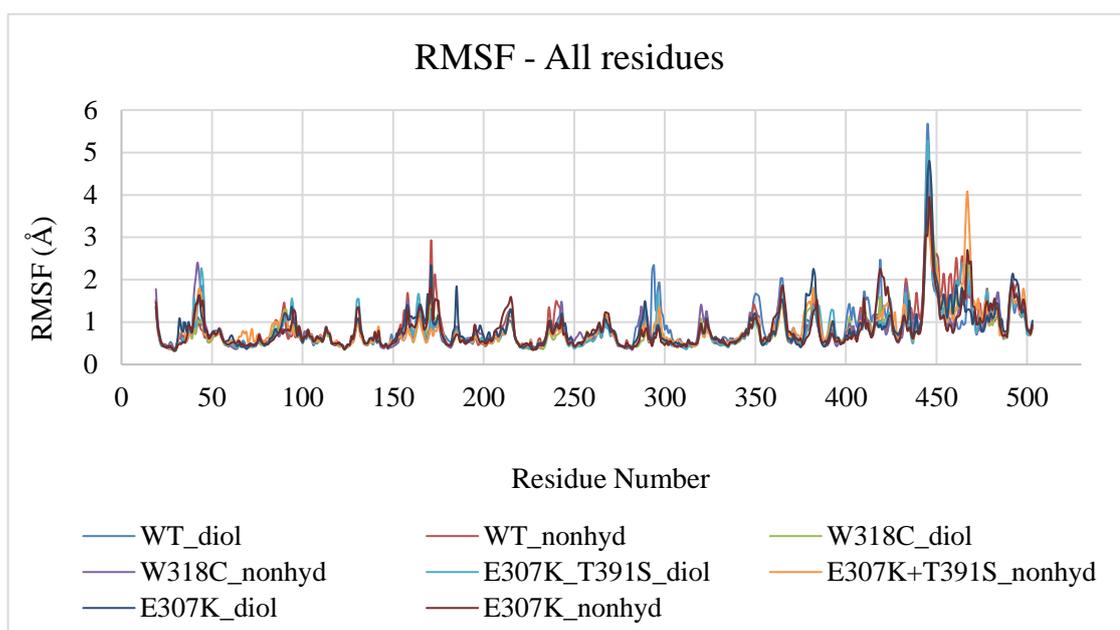


Figure 4.49. Residue based RMSF values at acidic pH.

Table 4.9. Average RMSF values and their corresponding standard deviations all residues and the specific regions of ARSA.

	Whole Protein RMSF (Å)							
	WT		W318C		E307K+T391S		E307K	
	diol	nonhyd	diol	nonhyd	diol	nonhyd	diol	nonhyd
Avg	0.84	0.85	0.77	0.80	0.81	0.83	0.85	0.82
±Std	0.51	0.50	0.46	0.44	0.50	0.45	0.49	0.46
	Active Site RMSF (Å)							
	WT		W318C		E307K+T391S		E307K	
	diol	nonhyd	diol	nonhyd	diol	nonhyd	diol	nonhyd
Avg	0.45	0.43	0.40	0.39	0.41	0.48	0.43	0.41
±Std	0.11	0.07	0.07	0.05	0.07	0.12	0.07	0.06
	Octamerization Helix RMSF (Å)							
	WT		W318C		E307K+T391S		E307K	
	diol	nonhyd	diol	nonhyd	diol	nonhyd	diol	nonhyd
Avg	1.24	1.91	1.50	1.25	1.67	1.84	1.46	1.38
±Std	0.42	0.41	0.37	0.28	0.41	0.81	0.27	0.54

Table 4.9. Average RMSF values and their corresponding standard deviations all residues and the specific regions of ARSA (cont.).

<b>Core Beta Sheets RMSF (Å)</b>								
	<b>WT</b>		<b>W318C</b>		<b>E307K+T391S</b>		<b>E307K</b>	
	diol	nonhyd	diol	nonhyd	diol	nonhyd	diol	nonhyd
Avg	0.45	0.47	0.44	0.44	0.45	0.47	0.46	0.48
±Std	0.06	0.04	0.07	0.05	0.06	0.09	0.05	0.07
<b>Dimerization1 RMSFD (Å)</b>								
	<b>WT</b>		<b>W318C</b>		<b>E307K+T391S</b>		<b>E307K</b>	
	diol	nonhyd	diol	nonhyd	diol	nonhyd	diol	nonhyd
Avg	0.83	0.67	0.73	1.08	0.93	0.87	0.95	0.87
±Std	0.29	0.17	0.34	0.66	0.56	0.40	0.33	0.43
<b>Dimerization2 RMSF (Å)</b>								
	<b>WT</b>		<b>W318C</b>		<b>E307K+T391S</b>		<b>E307K</b>	
	diol	nonhyd	diol	nonhyd	diol	nonhyd	diol	nonhyd
Avg	1.24	0.75	0.65	0.79	0.82	0.82	0.83	0.64
±Std	0.54	0.20	0.12	0.28	0.34	0.22	0.26	0.13
<b>Dimerization3 RMSF (Å)</b>								
	<b>WT</b>		<b>W318C</b>		<b>E307K+T391S</b>		<b>E307K</b>	
	diol	nonhyd	diol	nonhyd	diol	nonhyd	diol	nonhyd
Avg	1.52	1.32	1.24	1.21	1.21	1.19	1.23	1.28
±Std	1.05	0.79	0.84	0.84	1.02	0.63	0.96	0.82

The results in Table 4.9 show that the overall backbone mobility does not change significantly due to mutations but the least mobile backbone structure is W318C mutant forms, especially the diol form of W318C mutant, according to these results. This trend is also valid for the active site results that W318C forms have the lowest RMSF values for the active site residues. When we look at the results of the octamerization helices, we see that the most significant difference in the average RMSF values is between the two states of WT form of ARSA. Diol form of WT ARSA has the lowest average RMSF result which supports the hydration state is achieved when the octamerization is occurred or vice versa. In addition, the residues with the RMSF values higher than 2.0 Å are listed in Table 4.10.

Table 4.10. Residues, with RMSF values higher than 2.0 Å.

<b>Simulation</b>	<b>Residues with &gt; 2.0 Å RMSF</b>
WT_diol	293, 294, 364, 365, 419, 443, 444, 445, 446, 447, 448, 468
WT_nonhyd	171, 173, 433, 444, 445, 446, 447, 448, 449, 450, 451, 454, 457, 458, 460, 461, 464, 465
W318C_diol	444, 445, 446, 447, 448, 449, 450, 467, 468
W318C_nonhyd	41, 42, 43, 444, 445, 446, 447
E307K+T391S_diol	44, 45, 444, 445, 446, 447, 448, 463, 464, 465, 466, 467, 468, 469
E307K+T391S_nonhyd	444, 445, 446, 447, 448, 449, 450, 451, 465, 466, 467, 468, 469
E307K_diol	171, 382, 383, 444, 445, 446, 447, 448, 449, 450, 492, 493
E307K_nonhyd	418, 419, 420, 421, 443, 444, 445, 446, 447, 448, 449, 467, 468, 469

The residues in Table 4.10 are in agreement with the RMSD results because the regions showing sharp changes in their RMSD profiles shown previously consist of mainly the residues in the table. WT-nonhydrated form has nine residues with RMSF values greater than 2.0 on its octamerization helix (Pro450-Gln465). When we look at the residues of W318C-nonhydrated and E307K+T391S-diol mutant forms, we that there are residues with high RMSF values on the dimerization1 region of these structures (Asp29-Thr46). E307K+T391S-nonhydrated form also has residues with high RMSF values on its octamerization helix. The residues of the hinge-like loop on the dimerization3 region (Leu441-Ala448) showed up in Table 4.10 for each simulation. This region was originally missing in the crystal structure and was built later in the protein preparation step. That is why the only point we can directly conclude to is that this region is very flexible and the random movements of the residues on this region may not be significant for the protein stability or dimerization. The results should be revisited after the dimer simulations are completed.

The RMSF results were also shown for all carbon alpha atoms of the residues of all forms in B-scale along with the experimental B-factor data of ARSA for comparison. These results were plotted in Figure 4.50 for all ARSA forms at acidic pH. In fact, it would be better to prioritize the nonhydrated forms in terms of protein stability, because it is the state in that ARSA dimers or octamers preserve their structural integrity in the lysosome before

the reaction, thus nonhydrated forms should be considered more in the case of severe MLD causing mutants.

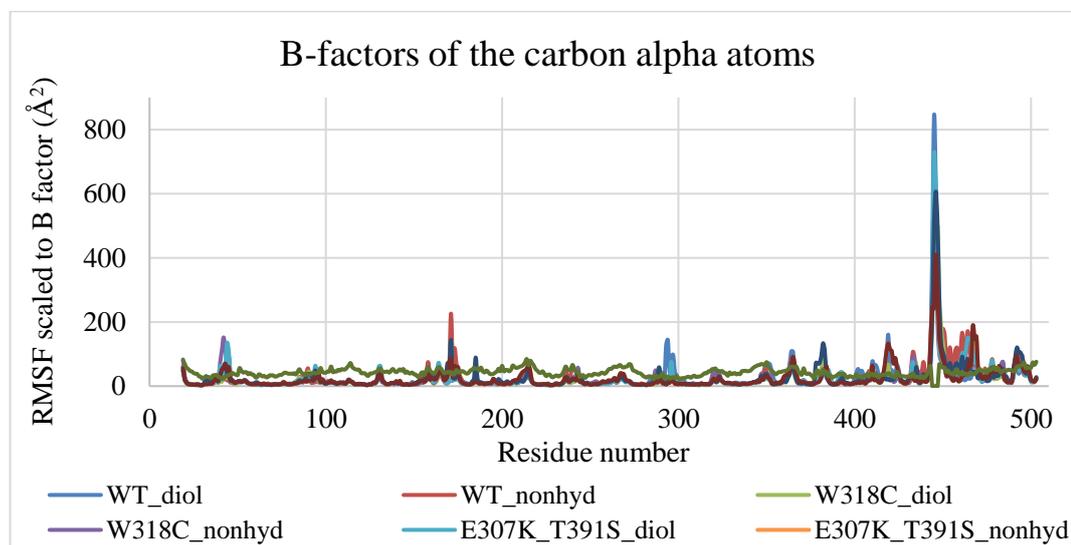


Figure 4.50. Experimental and calculated B-factors of the carbon alpha atoms.

The results in Figure 4.50 represents RMSF results of the residues as scaled to B-factor along with the experimental B-factor data. The residues with B-factor values higher than 100 correspond to the ones with RMSF values higher than  $2.0 \text{ \AA}$ . This is another way of showing which regions are fluctuating more and what is the ratio of this fluctuations to the measured experimental data.

In addition, the hydration state of ARSA was considered as an effect on the stability of the enzyme. When we plotted the RMSF results of all atoms for WT forms we saw that hydration makes Val445 and its surrounding more fluctuate whereas, Gly171 and Cys172 become more mobile in the diol forms of ARSA as shown in Figure 4.51. Comparison for other forms of ARSA were done and discussed further as in Figure A.2, Figure A.3, and Figure A.4.

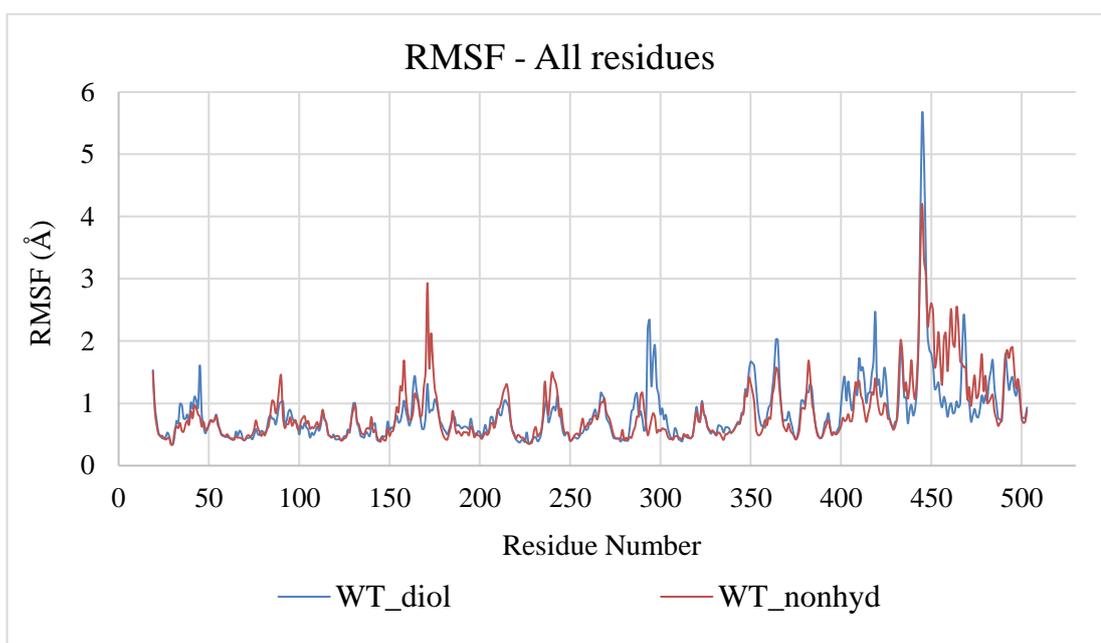


Figure 4.51. Comparison of the RMSF results of both forms of WT ARSA at acidic pH

From the results and other simulations on ligand bound and dimer forms we concluded that the state of FGL is not a determinant of the dynamics of the enzyme because even in the simulations on WT dimer ARSA forms we saw that protein dynamics can change significantly. Although it is a big sized protein, ARSA has very mobile regions that can show different dynamics and RMSD or RMSF results in each simulation.

#### 4.3.4. State of NAG+NDG Sugar Moiety at the Glycosylation Site

ARSA crystal structure in PDB (code: 1AUK) has a sugar entity (NDG+NAG) which is covalently bonded to Asn184, one of the three glycosylation sites [49,14,100,101]. When we observed all the simulations, we saw that the sugar entity changes its conformation and turns its body from up (His143) to down (towards His138) as shown in Figure 4.52 from the simulation on E307K+T391S-diol mutant form. The conformational changes occur differently only for the simulations on W318C in neutral pH but there are still major changes observed such that NDG+NAG complex becomes closer to His138 as shown in Figure 4.53.

Since the movements of the sugar moiety were observed in all simulations on all types of ARSA forms it can be deduced that the natural form of NAG+NDG in its glycosylation

site is near His138 rather than His143. It may depend on the molecule's strain energy in other words, inner lowest energy conformation. Bond angles and planarity may be driving the molecule to gain such conformations shown in Figure 4.52 and Figure 4.53.

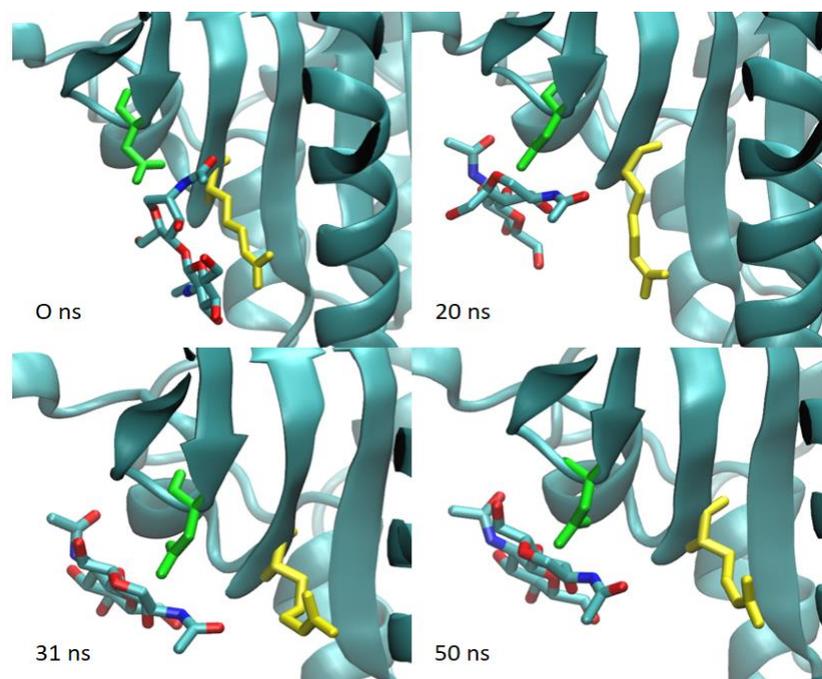


Figure 4.52. NDG+NAG sugar moiety (shown with sticks representation at 0 ns, 20 ns, 31 ns, 50 ns of the simulation on E307K+T391S-diol ARSA at neutral pH.

Asn184 is shown with sticks representation in neon green and His143 is shown in yellow in Figure 4.52. The covalent bonding between the sugar moiety and Asn184 is not shown for the snapshots. The movements of the sugar moiety in the simulation on W318C mutant is also shown in Figure 4.53. The coloring was done as the same as in Figure 4.52 for the same residues.

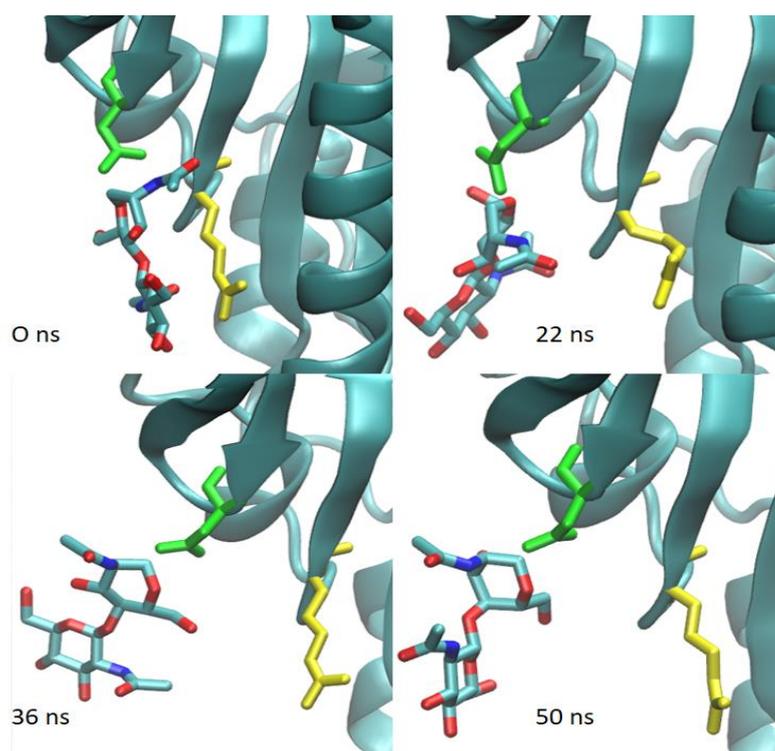


Figure 4.53. NDG+NAG sugar moiety (shown with sticks representation) at 0 ns, 22 ns, 36 ns, 50 ns of the simulation on W318C mutant ARSA at neutral pH.

NDG+NAG sugar moiety is covalently bonded to Asn184 but it is apparent that the best conformation for this group is to be away from His143. In all simulations, it moves towards His138. The acidity of the environment did not change the movements of the sugar group significantly.

#### 4.3.5. Salt Bridges and Hydrogen Bonding Occupancies

The secondary structure of the protein and possible salt bridges and the hydrogen bond occupancies between certain residues were calculated for all simulations at acidic pH. When all the possible salt bridges were compared for all simulations at acidic pH, we saw that some of the results were missing in other simulations. Certain possible salt bridges were formed in mutants whereas, some of them in the simulation on WT forms have gone missing in the simulations on mutant forms. These may be due to mutations and hydration state of FGL. The possible salt bridges with the average distances between the residues are summarized in Table 4.11.

Table 4.11. Average distances (whole trajectory) between possible salt bridge pairs.

Residue Pair	WT		W318C		E307K+T391S		E307K	
	Diol	Nonhyd.	Diol	Nonhyd.	Diol	Nonhyd.	Diol	Nonhyd.
ASP169-ARG288	4.81 Å	7.53 Å	-	7.34 Å <sup>+</sup>	7.30 Å	4.48 Å	-	-
ASP29-HIS229	4.70 Å	-	-	-	-	4.75 Å	-	-
ASP354-ARG80	7.40 Å	-	-	-	-	-	-	-
ASP434-LYS433	6.32 Å <sup>*</sup>	8.97 Å	5.21 Å <sup>***</sup>	7.65 Å	-	-	-	-
ASP50-ARG58	9.37 Å	-	-	-	-	-	-	-
GLU131-ARG97	10.07 Å	-	9.63 Å	10.84 Å	6.94 Å <sup>++</sup>	8.00 Å	12.13 Å	9.75 Å
GLU307-ARG311	4.23 Å	4.27 Å	4.25 Å	4.25 Å	-	-	-	-
GLU307-LYS395	3.62 Å	3.61 Å	3.60 Å	3.60 Å	-	-	-	-
ASP482-ARG479	-	12.20 Å	-	-	-	-	-	-
ASP61-LYS393	-	7.87 Å <sup>**</sup>	12.0 Å	-	-	-	-	-
GLU272-ARG19	-	10.07 Å	-	-	-	-	-	11.55 Å
GLU329-LYS393	-	8.71 Å	-	-	-	-	-	-
GLU103-ARG479	-	-	9.15 Å	-	8.83 Å	-	9.85 Å	9.72 Å
GLU240-ARG241	-	-	11.83 Å	-	11.61 Å	11.32 Å	-	-
ASP211-ARG217	-	-	-	-	-	-	-	7.20 Å <sup>•</sup>
ASP467-ARG84	-	-	-	-	-	10.27 Å	-	-
GLU451-LYS433	-	-	-	-	-	-	10.74 Å	-

\*: Asp434 and Lys433 pair in the simulation on WT-diol form starts with 9.0 Å but gets closer after 41 ns and stays in proximity around 3.5 Å.

\*\* : Asp61 and Lys393 pair in the simulation on WT-nonhydrated form starts with 12.9 Å but gets closer after 39 ns and stays in proximity around 4.2 Å and after 49 ns around 2.2 Å.

\*\*\*: Asp434 and Lys433 pair in the simulation on W318C-diol form starts with 9.5 Å but gets closer after 44 ns and stays in proximity around 3.0 Å.

+: Asp169 and Arg288 pair in the simulation on W318C-nonhydrated form starts with 6.5 Å but gets closer after 42 ns and stays in proximity around 4.2 Å.

++: Glu131 and Arg97 pair in the simulation on E307K+T391S-diol form starts with 14 Å but gets closer after 31 ns and stays in proximity around 4.5 Å.

•: Asp211 and Arg217 pair in the simulation on E307K-nonhydrated form starts with 10 Å but gets closer after 25 ns and stays in proximity around 4.1 Å.

Hydrogen bond occupancies between the certain residue pairs and between the residues in the specific regions of ARSA were calculated. In addition to these some of the residues were selected from other parts of the enzyme due to their movements after inspecting the trajectories. They were selected to be important for the protein folding and stability since the disruption of some of the hydrogen bonds in the backbone or between two secondary structures can be considered as an onset of the protein unfolding process. In this case of ARSA, also the epitope regions gain remarkable importance for the protein folding in terms of having certain hydrogen bonds. The hydrogen bond occupancies that were calculated are summarized in Table A.3 and discussed as below:

- Arg19 is the first residue after the signal peptide and one of the early forming epitopes. It is also located near Trp318 such that the distance between the carbon alpha atoms of Arg19 and Trp318 in the original crystal structure of ARSA is 14 Å. The main difference is that in the simulations on W318C forms the hydrogen bond occupancy between Arg19 and Glu273 decrease and the occupancy of the hydrogen bond between Arg19 and Asp216 increase when compared to the other simulations. It is similar for the simulations on both forms of E307K+T391S mutant form. On the other hand, the occupancy for the bond between Arg19 and Asp216 becomes decreased in the simulation on E307K-nonhydrated form. The other difference is between the nonhydrated and diol states. In the simulations on the nonhydrated forms of WT, E307K mutant, and W318C mutant ARSA enzymes the occupancy of the hydrogen bond between Arg19 and Glu273 is greater than the same pairs in the simulations on diol form of the same forms. On the other hand, the occupancy of the hydrogen bond between Arg19 and Asp216 is less in the simulations on the nonhydrated forms of the same ARSA structures.
- Asp29 is one of the active site residues and it is also the first residue of dimerization1 region. The occupancies of the hydrogen bonds between Asp29-Lys123 and Asp29-His227 are almost the same for all simulations. In the simulation on E307K+T391S-diol form the occupancy of the hydrogen bond between Asp29 and Arg73 becomes significantly increased when compared to the other simulations. Average occupancy of the hydrogen bond between Arg73 and Asp29 for WT and W318C forms is around 3% whereas, it is 41.43% in the simulation on E307K+T391S-diol form. The

occupancy of the hydrogen bond between Arg73 and Asp29 is higher for the both forms E307K mutant than the two forms of WT and W318C mutant ARSA structures. When E307K mutant form becomes hydrated the hydrogen bond occupancy between Asp29 and Arg73 becomes higher even though this increase is not significant. Also, in the simulation on E307K+T391S-nonhydrated form Asp29 does not form a hydrogen bond with Arg73. The distance between the alpha carbons of Asp29 and Glu307 in the original crystal structure of ARSA is 17.5 Å.

- Asp61 is a negatively charged residues which is located close to the dimerization site of ARSA. The distance between the carbon alpha atoms of Asp61 and Glu307 is 9.90 Å. When we compare the occupancies of the hydrogen bond between Asp61 and Arg390, we see that it falls to 20% in the simulations on both forms of E307K+T391S mutant. It is around 72% for WT and W318C mutant forms. On the other hand, the occupancy of the bond between Asp61 and Arg311 becomes greater in the simulations on both forms of E307K+T391S mutant. The occupancy for this pair seems insignificant for WT and W318C mutant forms but it increases at least 10-fold in the simulations on E307K+T391S mutant. The occupancy between Asp61 and Arg390 becomes a little higher in the simulations on both forms of only mutant E307K while, occupancies between Asp61 and other residues remain similar to the ones of E307K+T391S mutant forms. The hydrogen bond between Asp61 and Arg390 (instead of Arg311) may not mean something important since Arg311 and Arg390 are located very close to each other and Asp61 as shown in Figure 4.54. Small movements can cause this difference which may not have a structural significance.

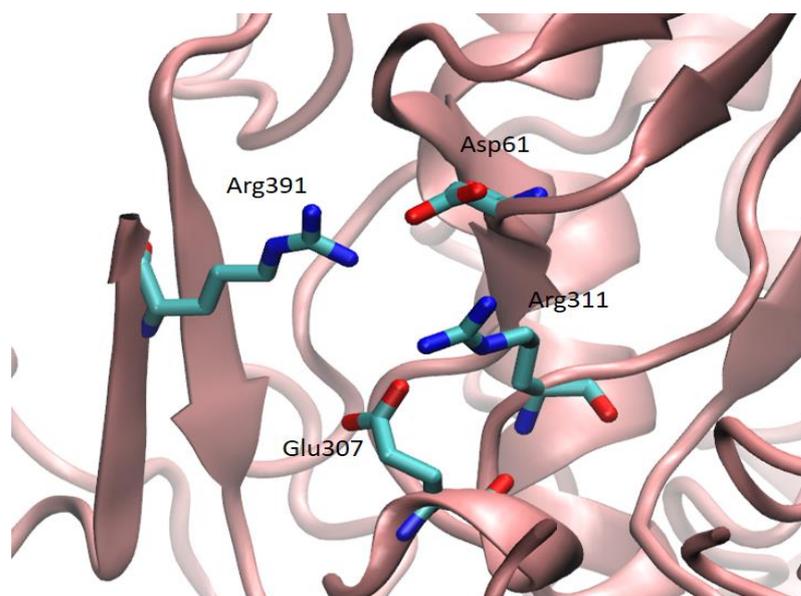


Figure 4.54. Asp61, Glu307, Arg311, and Arg390 on the original crystal structure of ARSA (PDB code: 1AUK).

#### 4.3.6. Secondary Structure Analysis

When performing MD simulations on mature ARSA the effects of mutations on folding and the stability of the enzyme is not easy to observe, because the protein is big in size (485 amino acids), and the already folded structure is not easily affected. Even in the simulation on known to be severely misfolded, and ER retention occasioned T274M mutant form at neutral pH, mutant ARSA preserved its overall structure and its specific regions' during the simulation, but only the Val470 end of the octamerization helix and the dimerization1 region [4,7]. Since the mutation, T274M, causes late infantile MLD and ER retention of ARSA we expected to see that the protein structure would be affected and severe secondary structure changes would be observed. When we analyzed the trajectory of the simulation on T274M mutant monomer ARSA we saw that the only significant difference on the structure is the deterioration of the Val440 end of the octamerization helix and the conformational changes of the residues in the dimerization1 region. The RMSD results of the simulation on T274M mutant showed that at the last 10 ns the octamerization helix deviates away from the initial structure significantly, whereas the changes on the dimerization1 region are not severe as in the simulations on W318C and E307K+T391S mutant forms at neutral pH. We see that Pro42 flips its sidechain upward direction slightly.

In addition, the RMSD profile and the movements of the dimerization<sup>2</sup> region of T274M become closer to that of W318C for 10 ns between 30th and 40th ns of the simulation. The other regions like dimerization<sup>3</sup> region or Asp335-Arg370 region do not show significant differences for T274M mutant as for the other mutants when compared to WT. These similarities imply that W318C and E307K+T391S mutations may also be causing misfolding of ARSA leading to degradation of the enzyme in ER. Accordingly, the secondary structure maps of the forms at acidic pH were analyzed. The maps showed that mutant forms, W318C-nonhydrated, both forms of E307K+T391S, both forms of E307K have some changes on the secondary structures. The biggest changes on alpha helices, especially the octamerization helix, were observed in the simulation on E307K+T391S-nonhydrated form. In addition to this, the diol form of E307K+T391S mutant has less deterioration in its secondary structure map. The changes on the alpha helices in the map of E307K+T391S-nonhydrated form is shown in Figure 4.55. This figure is given as example output from the Simulation Interactions Diagram (SID) tool of Schrödinger. The difference between the diol and nonhydrated forms is also seen in the case of single mutant form of E307K. Its diol form has less disturbance on the bands in its secondary structure map.

Red bands in Figure 4.55 stand for to indicate helical regions, blue for strand, white for loop and grey for turn. The arrows indicate the changes on the secondary structure in Figure 4.55. The severe changes in the map of E307K+T391S-nonhydrated form are not observed in the map of W318C-nonhydrated mutant. The secondary structure of the octamerization helix of the W318C-nonhydrated mutant form has the closest secondary structure map to WT ARSA forms. For comparison, the secondary structure map of W318C-nonhydrated mutant form is also shown in Figure 4.55.

The existence of E307K mutation makes the secondary structure of certain helices and little beta strands deteriorate during the simulations. On the other hand, the diol forms show less disturbance in their secondary structure maps for E307K single mutation as well. Another example for this situation is that the little alpha helix (Pro360-Gly364) stays rigid during the simulation on E307K-diol form whereas, it loses its helical structure in the simulation on the nonhydrated form. The most observable changes occur on the two helical regions. First one is one of the longest alpha helices on the right of the front side of ARSA, and the other one is the octamerization helix. They are formed by the residues Leu194-

Gln215 (the long alpha helix on the front) and Pro450-Val470 (octamerization helix). The sequence of ARSA starts from Arg19 but the secondary structure maps starts counting the residues from 1 that is what there is a shift for 18 residues in the maps.

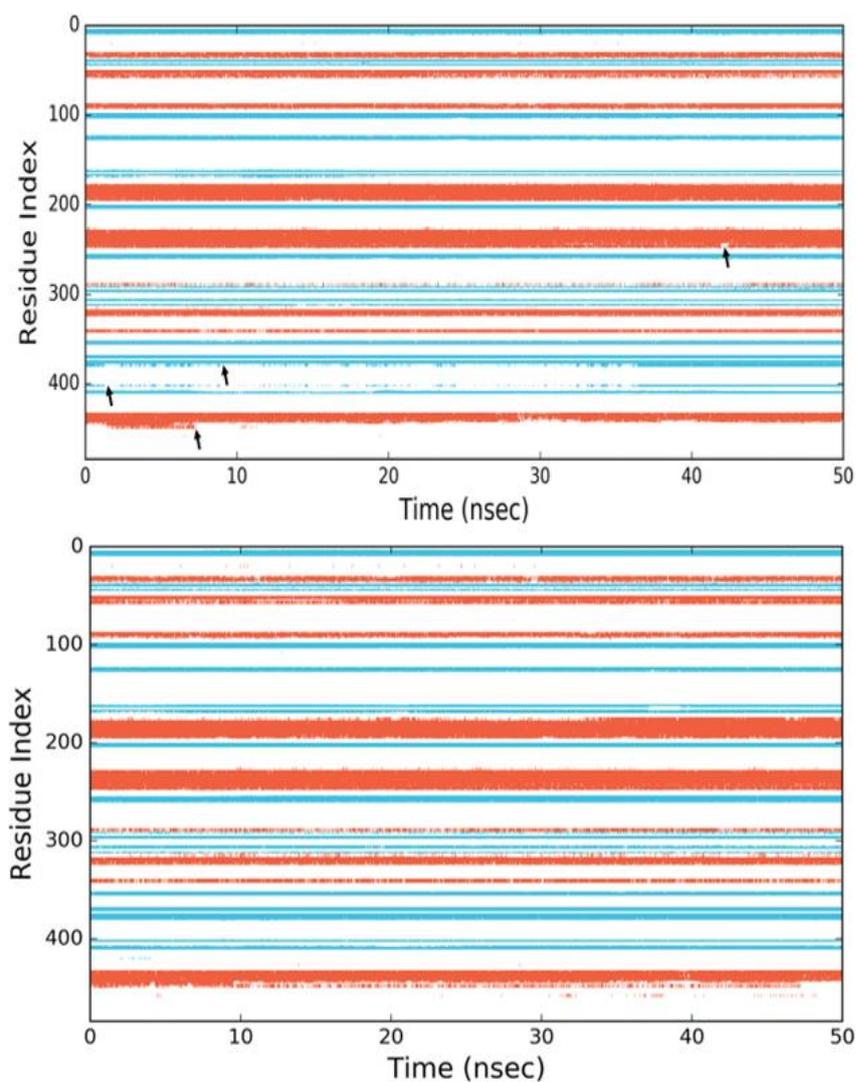


Figure 4.55. The secondary structure maps of E307K+T391S-nonhydrated (top) and W318C-nonhydrated (bottom) forms.

#### 4.3.7. Solvent Accessible Surface Area (SASA) Calculations

The changes in the solvent accessible surface area (SASA) of the protein are important for protein folding and recognition. First, the ER control mechanism is based on the

recognition of the hydrophobic residues which become solvent accessible since the hydrophobic residues are in the center parts of the protein their exposure to solvent signals misfolding of the protein. Second, the changes in the SASA of a protein can affect its oligomerization state due repulsion of the substructures during oligomerization. To understand the effects of the mutations on the SASA of ARSA we performed SASA calculations for each time step of the simulations. The results are given in Figure 4.56.

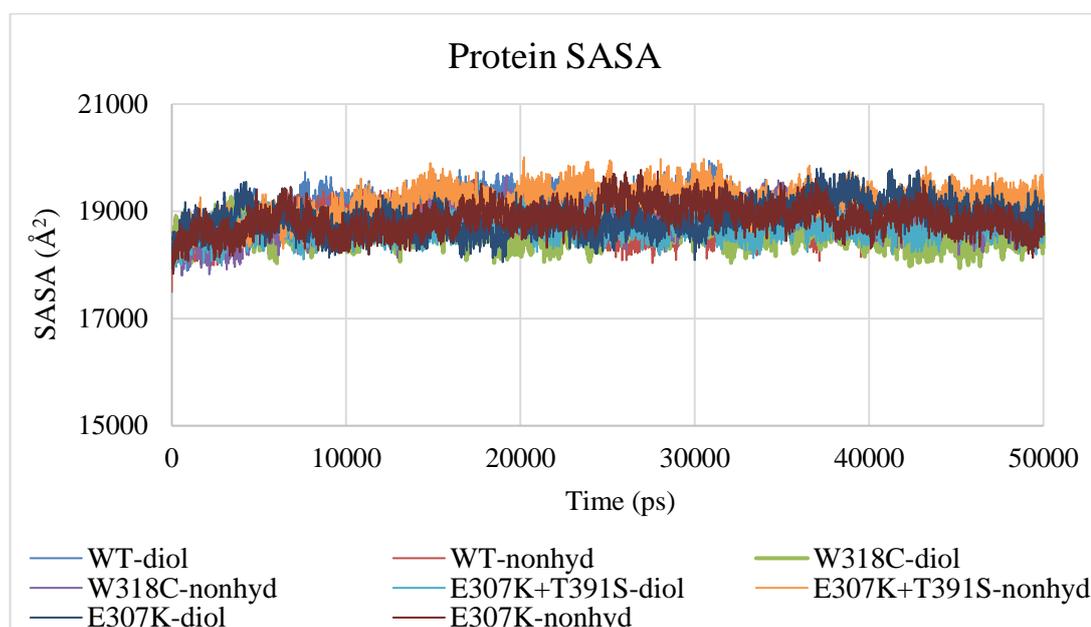


Figure 4.56. SASA profiles as a function of simulation time for simulations at acidic pH.

The results in Figure 4.56 show that mutations do not cause any major changes on the surface of the protein but it does not imply that this is not happening during folding since all the simulations were performed on the mature form of ARSA monomers. We also saw that overall protein dynamics do not change due to mutations (Figure 4.16) and SASA results also confirmed this outcome.

## 4.4. Pharmacophore Modeling

### 4.4.1. Ligand-Based Pharmacophore Modeling

The only known bioactive molecules of ARSA were retrieved from ChEMBL database. There are 653 molecules with standard potency values ranged from 5.4 nM to 42561.5 nM [82], [83]. These values were converted into pIC<sub>50</sub> values and their range is between 4.37 and 8.26. Since the number of the molecules is too high, first clustering was performed, and the cluster with the most populated one was chosen for QSAR analysis. This cluster includes 38 molecules. They include molecule between above pIC<sub>50</sub> values, and they were ranged as active, moderately active, and inactive due to selected threshold values 6.20 and 5.40 for active and inactive respectively. 11 of the molecules were selected randomly for the test set and the remaining 27 were selected for the training set. The model, AAADD, with the highest survival score among the all trials was selected for QSAR analysis. The results were insufficient to accept the model as accurate, because Q<sup>2</sup> and R<sup>2</sup> values were -0.52 and 0.76 respectively. These are under the threshold values and the standard deviation was 0.59, thus we rejected to build a model from ligand-based pharmacophore modeling.

### 4.4.2. Structure-Based Pharmacophore Modeling

QSAR analysis was performed as previously described in Section 3.5.2. The molecule sets from ZINC database, OTAVA database and ChEMBL database were used in VSW calculations. From the last step, SP docking, their XP descriptor files were used in E-Pharmacophore script to build structure-based models. Most of the molecules were more comfortable near loop regions of the active site, but not in the active site cavity as desired despite of still being in the grid box. Most of the binding poses were far away from the proposed binding site of the substrate. It is not a prerequisite for a candidate PC to bind as the substrate of the corresponding protein, but as their binding location moves away from Mg<sup>2+</sup> cation their binding free energies also increase. Since the models were built due to most of the resulting poses, the E-Pharmacophore scores were also very high for the used molecule databases for ARSA. Structured-based pharmacophore calculations were done on WT and W318C mutant monomers. Their results were listed in Table 4.12. The scores of

the sites in the hypotheses are too high to select for accepting a common scaffold for candidate chaperones of ARSA.

Table 4.12. Structure-based pharmacophore features by E-pharm of two forms of ARSA monomers at neutral pH.

<b>Protein Form</b>	<b>Molecules</b>	<b># of Molecules</b>	<b>Hypothesis</b>	<b>Scores</b>
W318C	OTAVA CNS Lib	214,300	APD	-1.66, -1.59, -1.34
W318C	Schöd Frag Lib	667 fragments	DHD	-1.66, -1.58, -1.26
W318C	ZINC Lib	4 millions	PNA	-2.84, -1.91, -1.60
WT	OTAVA CNS Lib	214,300	RDN	-3.11, -1.60, -1.37
WT	Schöd Frag Lib	667 fragments	ADD	-1.17, -0.90, -0.90
WT	CHEMBL Database	650	HDN	-2.24, -1.65, -1.27

Scores are given in the order of the corresponding letters in the hypothesis. The meaning of the letters in the hypotheses are as follows: A: Hydrogen Bond Acceptor, P: Positive Ionic Center, D: Hydrogen Bond Donor, H: Hydrophobe, N: Negative Ionic Center.

Due to the insufficient statistical results and bad/high scores of both ligand and structure based pharmacophore models we continued our docking calculations with other molecule sets (*de novo* synthesized molecules and aSMase inhibitors). If we were to have better values in our E-pharmacophore models, we would find matches in the same and other molecule databases for a start-up molecule base for candidate PCs.

#### 4.5. Docking Results at Neutral pH

The general working mechanism of pharmacological chaperones is to bind the unfolded or partially misfolded proteins to assist their folding and stability for their pass from ER control mechanism [27–33,104]. In this work, we searched pharmacological chaperone candidates for two types of ARSA mutants (W318C and E307K+T391S). The known aSMase inhibitors, the known bioactive molecules of ARSA, *de novo* synthesized molecules etc. were selected for docking calculations after the insufficient results of pharmacophore modeling. We evaluated the docked poses of these putative molecules in

terms of their binding modes and binding free energies in comparison with the results of the artificial substrate (NCS), natural ligands (sulfogalactosylceramide, SGC, and sulfogalactosylglycerolipid, SGG) which have shorter carbon tails than cerebroside 3-sulfate, and known bioactive molecules (ChEMBL molecule set of ARSA) of ARSA. Docking results of natural ligands and their interactions are given in Table 4.13 and Table 4.14, and the results of the top 10 compounds from known bioactive molecules of ARSA and known inhibitors of aSMase are given in Table 4.15 and Table 4.16.

#### 4.5.1. Docking of Natural Ligands

NCS is the artificial substrate of ARSA which is known as p-nitrocatechol. SGC and SGG are the variants of the natural substrates of ARSA namely, sulfogalactosylceramide, and sulfogalactosylglycerolipid respectively. SGC and SGG have the same head group of the natural substrates but only with shorter hydrocarbon tails [3]. BEI, SEI and LLE properties change according to the solvent accessibility and solubility properties of the different conformations. Since the substrates are not considered as pharmacological candidates and they are known to be binding to ARSA experimentally [3,45,50], only their GScore results were used for comparison in terms of binding affinity. The threshold values of BEI, SEI, LLE and QPlogBB properties of the natural substrates were not used for comparison but the efficiency results of other docked molecules were evaluated more in detail.

Table 4.13. IFD and rigid docking results of the ARSA substrates in the order of XP score (kcal/mol).

Molecule	WT -XP Rigid Docking			W318C - XP Rigid Docking		
	SGC	SGG	pNCS	SGC	SGG	pNCS
XP Gscore	-10.83	-6.86	-3.82	-5.76	-5.31	-3.87
Mol W	459.5	432.4	235.2	459.5	432.4	235.2
SASA	717.7	711.2	402.7	703.3	711.0	402.2
FOSA	376.7	453.7	0.0	367.3	444.2	0.0
FISA	321.7	254.8	289.2	319.9	264.6	291.5
PISA	16.9	0.0	110.9	13.7	0.0	108.2
#nonHatm	30	28	15	30	28	15

Table 4.13. IFD and rigid docking results of the ARSA substrates in the order of XP score (kcal/mol) (con.t).

	<b>WT -XP Rigid</b>			<b>W318C - XP Rigid</b>		
Molecule	SGC	SGG	pNCS	SGC	SGG	pNCS
RO5	2	1	0	2	1	0
PSA	206.0	180.8	138.5	214.7	186.0	136.8
QPlogBB	-3.83	-3.11	-2.18	-3.64	-3.18	-2.20
QPlogPo/w	-2.67	-0.83	-0.58	-2.53	-0.82	-0.61
glide LE	-0.36	-0.25	-0.25	-0.19	-0.19	-0.26
deltaG	-10.83	-6.86	-3.78	-5.76	-5.32	-3.84
p(Ki)	7.94	5.03	2.77	4.22	3.90	2.81
BEI	17.28	11.63	11.78	9.19	9.02	11.97
SEI	3.85	2.78	2.00	1.97	2.10	2.06
LLE	10.61	5.86	3.35	6.75	4.72	3.42
	<b>WT -IFD</b>			<b>E307K+T391S - XP Rigid</b>		
Molecule	SGC	SGG	pNCS	SGC	SGG	pNCS
XP Gscore	-11.72	-8.95	-6.16	-5.66	-4.71	-3.30
Mol W	459.5	432.4	235.2	459.5	432.4	235.2
SASA	741.2	733.3	733.3	706.0	739.3	404.5
FOSA	377.9	453.0	0.0	370.6	459.4	0.0
FISA	347.6	277.9	289.0	322.3	277.5	294.8
PISA	13.3	0.0	106.2	10.1	0.0	107.1
#nonHatm	30	28	15	30	28	15
RO5	2	1	0	2	1	0
PSA	222.1	182.6	135.7	216.8	183.5	137.9
QPlogBB	-4.17	-3.44	-2.16	-3.68	-3.47	-2.25
QPlogPo/w	-2.66	-0.84	-0.62	-2.53	-0.82	-0.63
glide LE	-0.39	-0.32	-0.30	-0.19	-0.17	-0.22
deltaG	-11.73	-8.96	-4.50	-5.67	-4.70	-3.27
p(Ki)	8.60	6.57	3.30	4.16	3.45	2.40
BEI	18.71	15.19	14.02	9.05	7.97	10.19
SEI	3.87	3.60	2.43	1.92	1.88	1.74
LLE	11.26	7.41	3.92	6.69	4.27	3.02

Physiochemical properties (BEI, SEI, LLE) of the compounds were calculated using the equations given in the methods. Other properties such as ligand efficiency or predicted blood brain barrier coefficient (QPlogBB) were calculated by QikProp program. Increasing Gscore values in Table 4.13 show that mutations affect the binding strength of the substrates, although the RMSD profiles of the active sites of the mutant forms did not differ from the WT form. When the coordinates with the best poses of the SGC were aligned (RMSD = 1.532 Å), the coordinates of the active site residues overlapped without any significant difference in conformation. The poses of rigidly docked SGC compounds do not coincide with the proposed binding mode of cerebroside 3-sulfate, the natural substrate of ARSA. Sulfate groups of these SGC compounds stayed out of the active site cavities. In the case of WT form, the compound positioned at the entrance of the active site forming several hydrogen bonds between its O and OH groups on its tail and the active site residues such as Ser150 and His125. These interactions lead to better (more negative) GScore values. On the other hand, SGC compounds stay even further away from the active site cavity (Table 4.14) and interactions were formed with other residues such as Arg288 Asp173, rather than the active site residues.

Table 4.14. The distances between the sulfate and nitrogen atoms of SGC molecules and  $Mg^{2+}$  in the active site cavities of the WT and mutant ARSA forms and their interactions.

<b>Protein</b>	<b>SGC Nitrogen- Mg<sup>2+</sup> dist.</b>	<b>SGC Sulfate- Mg<sup>2+</sup> dist.</b>	<b>Interactions</b>
WT	11.74	11.72	Ser150: Hbond(s)x2, Lys302: Hbond(s), His125: Hbond(s), Arg288: Hbond(s), His229: Hbond(b)
E307K + T391S	9.53	15.00	Asp152: Hbond(s), Gln153: Hbond(s), Asp173: Habond(s), Arg288: Hbond(s) and salt bridge
W318C	12.53	11.55	Asp173: Hbond(s)x2, Gln153: Hbond(s), Asp152: Hbond(s), Lys302: Hbond(s), His229: Hbond(b)

#### 4.5.2. Docking of Known Bioactive Molecules of ARSA

Rigid docking results of known bioactive molecules of ARSA from ChEMBL database are listed in Table 4.15.

Table 4.15. Rigid docking results of the known bioactive molecules of ARSA from ChEMBL database in the order of XP score (kcal/mol).

WT										
ID #	15331	2514	32695	10969	3441	58	21991	148	13364	3351
	56	7	8	79	27		6		69	0
LLE	11.26	7.41	3.92	6.69	4.27	3.02	-9	-9	-8	-8
glide LE	0	0	0	-1	0	0	0	0	0	0
mol_M W	616	805	386	257	377	444	426	299	344	-
SASA	837	1021	675	498	752	774	652	560	608	-
FOSA	352	590	186	97	356	314	208	248	201	-
FISA	485	418	244	308	89	292	155	285	232	-
PISA	0	14	246	93	307	169	218	6	175	-
%HOA	0	0	44	12	100	17	84	43	50	-
#nonHat m	42	56	28	18	28	32	30	20	25	-
RO5	3	3	1	1	0	1	0	0	0	-
PSA	333	280	146	163	64	180	93	157	124	-
QPlogB B	-5	-5	-3	-2	-1	-3	-1	-3	-2	-
QPlogP o/w	-9	-3	0	-2	4	1	4	1	0	-
deltaG	-11	-12	-10	-9	-9	-9	-9	-9	-8	-
p(Ki)	8	9	7	7	7	7	6	6	6	-
BEI	13	11	18	26	17	15	15	21	18	-
SEI	2	3	5	4	10	4	7	4	5	-
LLE	17	12	7	9	2	6	2	6	6	-

Table 4.15. Rigid docking results of the known bioactive molecules of ARSA from ChEMBL database in the order of XP score (kcal/mol) (cont.).

<b>E307K+T391S</b>									
ID #	15331	58	33510	1096	1255	2730	3269	14173	2675
	56		7	979	94	94	58	92	48
Gscore	-11	-10	-9	-8	-8	-8	-7	-7	-7
glide LE	0	0	0	0	0	0	0	0	0
mol_MW	616	444	-	257	250	428	386	226	563
SASA	790	770	-	485	440	764	699	472	881
FOSA	342	312	-	114	57	439	182	112	502
FISA	448	313	-	288	186	193	262	263	196
PISA	0	146	-	83	123	131	255	97	183
%HOA	1	1	-	2	1	2	2	1	1
#nonHatm	42	32	-	18	15	31	28	16	42
RO5	3	1	-	1	0	0	1	1	2
PSA	315	185	-	160	85	148	151	125	130
QPlogBB	-4	-3	-	-2	-1	-2	-3	-1	-3
QPlogPo/w	-9	0	-	-2	-1	1	0	-2	6
deltaG	-8	-10	-	-8	-7	-7	-5	-7	-21
p(Ki)	6	7	-	6	5	5	4	5	16
BEI	10	16	-	22	22	12	9	24	27
SEI	2	4	-	3	6	4	2	4	12
LLE	15	7	-	7	7	4	4	8	9

Table 4.15. Rigid docking results of the known bioactive molecules of ARSA from ChEMBL database in the order of XP score (kcal/mol) (cont.)

W318C										
ID #	109697 9	1533 156	25147	132090 2	13364 69	3392 362	59	52	164	2450 19
Gscore	-11	-10	-10	-10	-10	-9	-9	-9	-9	-9
glide LE	-1	0	0	-1	0	0	-1	0	0	0
mol_M W	257	616	805	183	344	266	153	302	318	259
SASA	494	818	1024	393	651	457	357	547	523	465
FOSA	108	377	585	115	198	115	101	178	0	113
FISA	287	441	421	176	239	247	149	190	330	305
PISA	99	0	18	102	215	95	107	179	194	47
%HOA	16	0	0	54	50	21	57	81	28	22
#nonHat m	18	42	56	13	25	19	11	22	23	18
RO5	1	3	3	0	0	1	0	0	1	1
PSA	161	312	285	85	128	143	66	88	163	162
QPlogB B	-2	-4	-5	-1	-2	-1	-1	-2	-3	-2
QPlogPo /w	-2	-9	-3	-1	0	-2	-1	3	0	-2
deltaG	-11	-9	-10	-10	-10	-9	-9	-9	-9	-9
p(Ki)	8	6	8	7	7	7	7	7	7	7
BEI	30	10	10	41	21	26	44	22	21	25
SEI	5	2	3	9	6	5	10	8	4	4
LLE	9	15	10	8	7	9	8	4	7	9

Binding free energy results (Gscore) of the known bioactive molecules of ARSA are very good as expected. The potency values of the known bioactive molecules are given in Table B.1. Since their potency values were determined to be low experimentally, we decided to take the Gscore values in Table 4.15 as good binding scores for ARSA at neutral pH. Accordingly, the results of the docking calculations of the known bioactive molecules of ARSA are given as follows.

### 4.5.3. Docking Calculations of aSMase Inhibitors

We selected the inhibitors of acid sphingomyelinase because it also catalysis a hydrolysis reaction in the lysosome. Its deficiency causes another lysosomal storage disease, Niemann-Pick disease. Its secondary structure with 6 beta sheets at the center of the protein surrounded with alpha helices also resembles the secondary structure of ARSA. Both active sites of the proteins include His and Asp residues as well as a dimetal coordination site [28-30, 32].

Table 4.16. Rigid docking results of the known aSMase inhibitors at acidic pH in the order of XP score (kcal/mol).

<b>WT</b>										
Compound	5_R	9_A	7d_R	8_A	7c_R	8_R	16c_R	16a_R	6-OH-DL-dopamine	16e_R
XP Gscore	-10	-9	-8	-8	-7	-7	-7	-7	-7	-7
Mol W	559	1047	345	427	317	265	346	290	213	347
SASA	751	1496	622	570	556	422	653	493	412	613
FOSA	168	1077	409	65	308	110	426	233	44	294
FISA	568	369	211	408	245	310	223	256	280	314
PISA	0	44	0	97	0	0	0	0	88	0
%HOA	0	0	15	0	3	0	52	36	0	0
#nonHatm	31	69	21	27	19	15	21	17	15	21
RO5	3	4	1	2	1	1	0	0	1	1
PSA	302	290	122	232	133	159	125	137	136	163
QPlogBB	-6	-8	-2	-4	-2	-2	-3	-2	-2	-3
QPlogPo/w	-6	7	1	-2	0	-3	4	2	-3	0
glide LE	0	0	0	0	0	0	0	0	0	0
deltaG	-10	-9	-6	-5	-5	-5	-7	-7	-7	-6
p(Ki)	7	6	4	4	4	4	5	5	5	5
BEI	13	6	13	9	12	14	14	17	24	13
SEI	2	2	4	2	3	2	4	4	4	3
LLE	13	-1	3	5	4	7	1	3	8	5
<b>E307K+T391S</b>										
Compound	5_R	6-OH-DL-dopamine	8_A	6_R	16e_R	16d_R	13_R	8_R	9_R	18_R
XP Gscore	-8	-7	-7	-7	-6	-6	-6	-6	-6	-6
mol_MW	559	213	427	277	347	263	251	265	247	272

Table 4.17. Rigid docking results of the known aSMase inhibitors at acidic pH in the order of XP score (kcal/mol) (cont.).

<b>E307K+T391S</b>										
SASA	702	408	604	474	594	429	375	434	442	407
FOSA	116	42	87	142	272	108	73	119	192	25
FISA	574	271	404	329	320	317	302	311	247	270
PISA	0	95	103	0	0	0	0	0	0	109
%HOA	0	0	0	0	0	0	0	0	7	24
#nonHatm	31	15	27	16	21	15	14	15	14	16
RO5	3	1	2	1	1	1	1	1	0	0
PSA	304	132	223	164	167	162	163	167	133	148
QPlogPo/w	-6	-3	-2	-2	0	-2	-3	-3	-1	1
glide LE	0	0	0	0	0	0	0	0	0	0
deltaG	-7	-7	-4	-6	-6	-6	-4	-4	-4	-4
p(Ki)	5	5	3	5	4	4	3	3	3	3
BEI	9	24	7	16	12	17	12	11	12	12
SEI	2	4	1	3	3	3	2	2	2	2
LLE	11	8	5	6	4	6	6	6	4	3
<b>W318C</b>										
Compound	5_R	8_A	14_R	8_R	4_R	7b_R	15a_R	13_R	7a_R	16d_R
XP Gscore	-9	-8	-8	-7	-7	-7	-7	-7	-7	-7
mol_MW	559	427	219	265	299	289	274	251	261	263
SASA	717	596	358	399	447	496	479	397	455	420
FOSA	116	60	126	128	56	247	265	81	198	122
FISA	588	439	230	268	384	249	212	310	255	296
PISA	0	89	0	0	0	0	0	0	0	0
%HOA	0	0	1	0	0	0	44	0	0	0
#nonHatm	31	27	12	15	16	17	16	14	15	15
RO5	3	2	0	1	1	1	0	1	1	1
PSA	315	236	122	150	187	142	117	155	142	157
QPlogBB	-6	-2	-4	-3	-3	-2	-2	-2	-2	-2

Binding modes and strengths of the aSMase inhibitors and known bioactive molecules of ARSA were analyzed taking the natural ligand results as basis. If the GScore values are less than the natural ligands we assumed tighter binding due to more interactions between the compound and ARSA. The reference binding mode is the proposed mode of cerebroside sulfate in which the sulfate group enters to the active cavity and stays close the active site residues and the divalent cation. From our results, 5\_R has the best GScore values for all

ARSA forms but the octanol water partition coefficient falls below the set threshold value of -2.0 which means that 5\_R compound is highly hydrophilic. Predicted brain/blood partition coefficient of 5\_R also falls below its threshold value. A positive QPlogBB value is desired for drug delivery especially for the diseases related with CNS. On the other hand, LLE results of 5\_R compound are very positive which indicates that 5\_R can be a good drug candidate for ARSA. Regardless to these outliers in the specified properties, we selected 5\_R for further MD simulations due to its good rigid docking XP GScore result because we know that calculated BBB coefficient can change significantly for the same molecule with different conformations according to its PSA or SASA [105–107,66]. IFD was not performed because of the computational expense since the binding free energy results from the rigid docking calculations of 5\_R are significantly good. On the other hand, other compounds with values in the limits such as 16c\_R for WT or 15a\_R for W318C mutant should also be evaluated.

When we compare the results of the known bioactive molecules and aSMase inhibitors we can conveniently say that their binding to ARSA at neutral pH is strong according to their Gscore values. This is a promising start in the way of finding possible PCs for ARSA because the tight binding of the molecules at neutral pH can help mutant ARSA forms to become more stable and pass ER control mechanism or to oligomerize and these molecules should be considered further for improvement since aSMase is also a lysosomal enzyme.

#### **4.5.4. IF Docking of Natural Substrates**

Induced fit result of SGC shows that the conformation of SGC coincides very well with the proposed binding conformation of cerebroside 3-sulfate. For the comparison, the coordinates of both p-nitrocatechol sulfate (p-NCS) bound ARSA crystal structure and the top SGC bound WT ARSA were aligned as in Figure 4.57.

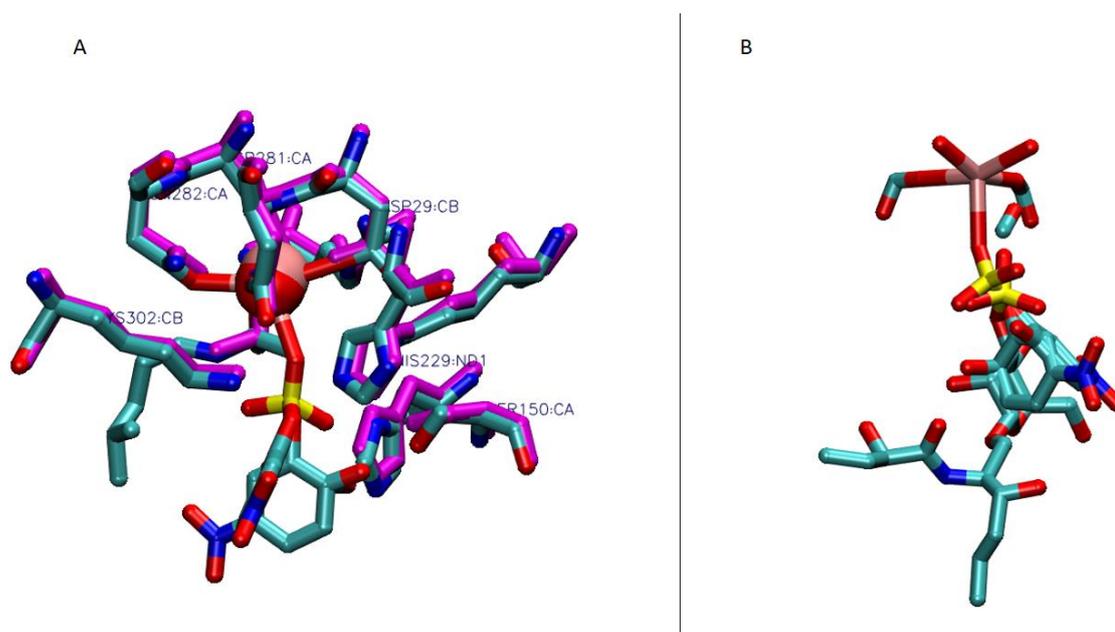


Figure 4.57. Active site of the p-nitrocatechol sulfate bound ARSA crystal structure (1E2S in blue) and SGC docked WT ARSA (1AUK in pink) structure.

These coordinates of SGC-ARSA complex in Figure 4.57 was taken from the induced fit docking results. The protein backbone alignment RMSD is 1.56 Å. Only p-nitrocatechol (P-NCS) is shown to emphasize the alignment of the active site residues. In 1E2S structure the Mg<sup>2+</sup> ion and CSN are covalently bonded. B shows the CSN and docked SGC structures after the protein alignment. The longer chain belongs to SGC. Both Mg<sup>2+</sup> ions are shown with sphere representation (pink: 1E2S, red: SGC docked 1AUK).

Figure 4.57 shows that IFD results agree with the experimental data. The sulfate groups of CSN and SGC overlap as well as the active site residues. The GScore result of this pose of SGC was -11.7 kcal/mol, which leads us to the conclusion of that candidate pharmacological chaperones should have similar or less negative GScore values when they are docked with the induced fit protocol. The interaction map between the SGC and ARSA is shown in Figure 4.58. Interacting residues are also in agreement with the proposed binding mode of the natural substrate.

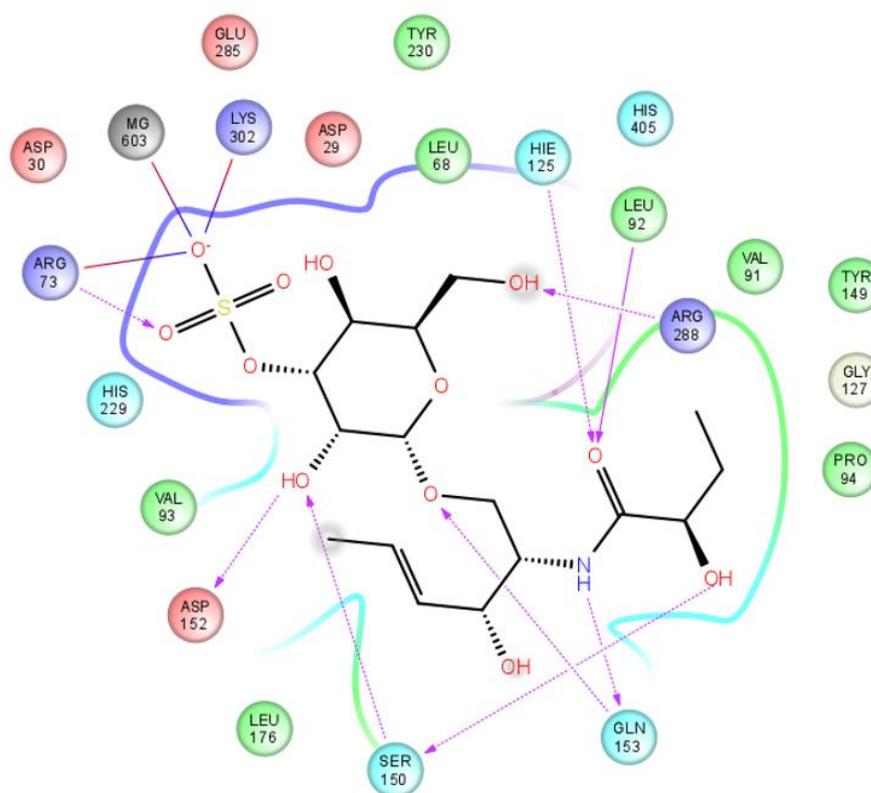


Figure 4.58. Interaction map between SGC and ARSA.

In Figure 4.58, three salt bridges are formed between the sulfate group and two active site residues (Arg73 and Lys302) and  $Mg^{2+}$  ion are shown. Two active site residues (Ser150 and His125- protonated) form hydrogen bonds with SGC.

According to binding mode and docking results of SGC, the results of the other docked compounds were analyzed in terms of having favorable binding and position. In the docking results of known bioactive molecules of ARSA (Table 4.16) some of the molecules have more SGC-like binding positions although having lower GScore values. These aspects should be considered along with the physicochemical properties of the molecules.

#### 4.5.5. Docking Calculations of Other Molecules

We also selected other inhibitor molecules like Verubecestat for Alzheimer [108] and GW4869 for neutral nSMase [109] as possible PC candidates. These molecules were docked

to three forms of ARSA in neutral pH with rigid XP docking protocol. The results are given in Table 4.17 and Table 4.18.

Table 4.17. The rigid docking results of Verubecestat in neutral pH are given for three forms of ARSA in the increasing order of GScore values.

Protein form	WT			W318C			E307K+T391S		
	#1	#2	#3	#1	#2	#3	#1	#2	#3
Poses									
XP Gscore	-4.00	-3.93	-3.89	-4.93	-4.24	-3.54	-5.81	-4.78	-2.96
MW	409.4	-	409.4	409.4	-	409.4	-	409.4	409.4
SASA	643.5	-	646.1	623.0	-	650.6	-	623.1	636.6
FOSA	170.1	-	169.0	166.6	-	169.7	-	162.0	160.6
FISA	176.6	-	178.7	168.3	-	179.0	-	163.7	194.2
PISA	225.6	-	232.9	215.2	-	231.1	-	224.4	206.3
%HOA	78.3	-	77.6	79.4	-	77.9	-	80.8	77.4
#nonHatm	28	-	28	28	-	28	-	28	28
RO5	0	-	0	0	-	0	-	0	0
PSA	119.7	-	120.6	120.6	-	121.0	-	119.7	121.9
QPlogBB	-1.186	-	-1.244	-1.04	-	-1.24	-	-0.97	-1.31
QPlogPo/w	1.679	-	1.615	1.62	-	1.68	-	1.74	2.04
glide LE	-0.14	-0.06	-0.04	-0.08	-0.07	-0.13	-0.12	-0.08	-0.11
deltaG	-3.98	-	-1.20	-2.24	-	-3.53	-	-2.10	-2.94
Ki	0.00	-	0.13	0.02	-	0.00	-	0.03	0.01
p(Ki)	2.91	-	0.88	1.64	-	2.59	-	1.54	2.16
BEI	7.12	-	2.16	4.01	-	6.32	-	3.76	5.26
SEI	2.43	-	0.73	1.36	-	2.14	-	1.29	1.77
LLE	1.24	-	-0.73	0.02	-	0.90	-	-0.20	0.12

Second poses of Verubecestat of WT, W318C mutant forms and the first pose of Verubecestat of E307K+T391S mutant form could not be processed due to the warning of its charged state or an odd number of electrons. The LLE values of these results are worse (less) when compared to the other docking calculations. The LLE and p(Ki) values of the results are below of the ideal value for optimal drug candidates [91,110]. Since the pose of the ligand is one of the determining factors for LLE we can conclude that these GScore values and poses of Verubecestat imply that Verubecestat may not be our first choice for PC candidates for ARSA.

Table 4.18. The rigid docking results of GW4869 in neutral pH are given for three forms of ARSA in the increasing order of GScore values.

Form	WT			W318C			E307K+T391S		
	#1	#2	#3	#1	#2	#3	#1	#2	#3
Poses	#1	#2	#3	#1	#2	#3	#1	#2	#3
Gscore	-6.56	-5.26	0.12	-7.19	-4.30	-1.11	-6.70	-6.01	-3.85
MW	-	-	504.6	-	504.6	-	-	-	504.6
SASA	-	-	913.5	-	836.8	-	-	-	829.9
FOSA	-	-	342.2	-	340.2	-	-	-	307.0
FISA	-	-	182.1	-	163.4	-	-	-	142.8
PISA	-	-	389.1	-	333.2	-	-	-	380.2
%HOA	-	-	81.0	-	81.9	-	-	-	87.3
#nonHatm	-	-	38	-	38	-	-	-	38
RO5	-	-	1	-	1	-	-	-	1
PSA	-	-	121.8	-	119.0	-	-	-	119.7
QPlogBB	-	-	-2.15	-	-1.72	-	-	-	-1.44
QPlogPo/w	-	-	4.50	-	4.11	-	-	-	4.44
Glide LE	-0.13	-0.10	0.01	-0.15	-0.11	0.02	-0.13	-0.11	-0.10
DeltaG	-	-	0.19	-	-4.22	-	-	-	-3.76
Ki	-	-	1.38	-	0.00	-	-	-	0.00
p(Ki)	-	-	-0.14	-	3.09	-	-	-	2.76
BEI	-	-	-0.28	-	6.13	-	-	-	5.46
SEI	-	-	-0.11	-	2.60	-	-	-	2.30
LLE	-	-	2.01	-	-1.02	-	-	-	-1.68

The first two poses of GW4869 for WT and E307K+T391S mutant forms, as well as the first and the last pose of GW4869 of W318C mutant form could not be processed for due to the warning of its charged state or an odd number of electrons. When we examined the poses, we saw that all resulting conformations of GW4869 could not reach the active site cavity and docked to the entrance of the active site. This is not the exact desired pose for the docked molecules to ARSA. For E307K+T391S, only the first pose is good in terms of entering the active site cavity and for W318C, all the poses of GW4869 can be said to be good in the same manner but when we looked at the interactions between the proteins and their corresponding docked GW4869 molecules, we did not see any significant interaction

between the molecules and the active site residues. Despite having good QPlogBB values the LLE values of the resulting conformations are bad for drug candidates.

#### 4.5.6. Docking of *de novo* Synthesized Molecules

Lastly, we performed docking calculations of *de novo* synthesized molecules from the Department of Pharmacy at Istanbul University on W318C mutant monomer at neutral pH. There are 8 molecules (SMILES codes and chemical structures are given in Appendix D), and 11 poses of these molecules were generated by using LigPrep. Rigid docking protocol with extra precision was performed on the coordinates that were taken from last frame of the simulation on apo form of W318C mutant monomer. The XP Gscore values were plotted as shown in Figure 4.59.

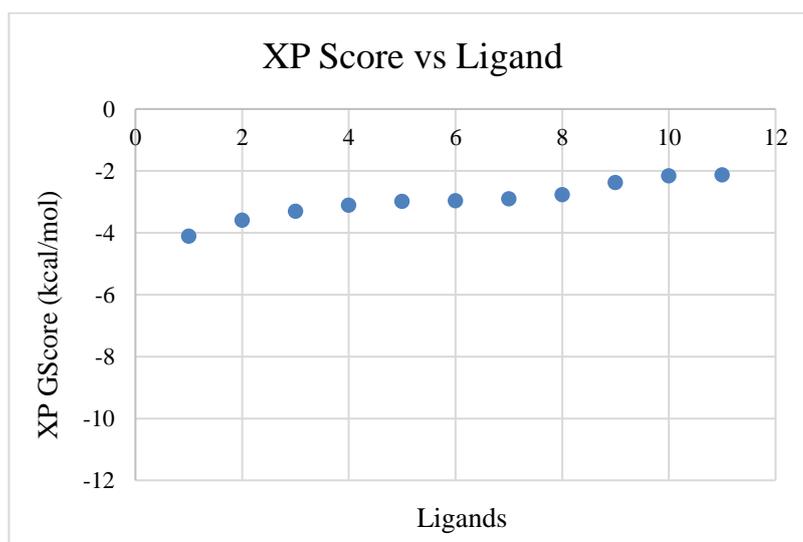


Figure 4.59. Binding free energy results for the *de novo* synthesized molecules after the docking calculations on W318C mutant.

As seen from Figure 4.59, the binding free energy results are very high when compared to the molecules that bind tightly to ARSA forms at neutral pH. Since, these molecules can be developed and synthesized we decided on performing an MD simulation after selecting one (D28) with the most promising predicted BBB coefficient and LLE value. The MD simulation results will be discussed in the corresponding section further, but as a preliminary

result we decided not moving on with the *de novo* synthesized molecules for ARSA due to their poor binding energies, and bad binding poses.

## 4.6. Docking Results at Acidic pH

### 4.6.1. Docking of Natural Ligands

Docking calculations were performed at acidic pH with the same compound datasets as in neutral ARSA analysis. The coordinates from the last frames of all simulations in acidic environment were taken and energy minimization was performed before docking. Accordingly, docking results of natural ligands and their interactions are given in Table 4.19 and Table 4.20, and the results of the top 10 compounds from known bioactive molecules of ARSA and known inhibitors of aSMase are given in Table 4.22 and Table 4.24.

Table 4.19. Rigid docking results of the substrates of ARSA at acidic pH as XP GScores (kcal/mol).

Protein Form	State	SGC	SGG	pNCS
WT	diol	-8.102	-7.237	-5.241
	nonhyd	-6.186	-6.126	-4.879
W318C	diol	-8.723	-8.533	-3.826
	nonhyd	-5.677	-4.59	-3.947
E307K+T391S	diol	-8.736	-8.534	-5.781
	nonhyd	-10.751	-8.637	-6.675
E307K	diol	-6.958	-7.548	-5.120
	nonhyd	-8.793	-6.538	-5.329

Results in Table 4.19 show that ligand binding is more favored in the hydrated state of FGL for WT and W318C mutant forms of ARSA while it is favored in the nonhydrated form of FGL for E307K+T391S mutant form of ARSA. The interactions of the resulting poses were summarized in Table 4.20 to understand how good the substrates of ARSA entered the active site cavity.

Table 4.20. The interactions between docked ARSA substrates and eight forms of ARSA at acidic pH.

<b>Protein Form</b>	<b>State</b>	<b>SGC</b>	<b>SGG</b>	<b>pNCS</b>
<b>WT</b>	Diol	Ser150 (H-s), polar: His125, Ser150, Gln153, His229	His125 (H-s), Lys123 (H-s x2), Gln153 (H-s), Polar: Thr70, His125, Ser150, Gln153, His229	His125 (pi-pi), Gln153 (H- s), Lys302 (SB, H-s), Polar: Thr70, His125, Ser150, Gln153, His229
	Nonhyd	Lys123 (H-s), Ser150 (H- s), Asn152 (H-b), Mg <sup>2+</sup> (MC), Polar: His125, Ser150, Gln153, Gln174, His229	Lys123 (SB), His125 (H-s), Asp152 (H-s), Polar: His125, Ser150, Gln153, Gln174, His229	Lys123 (SB), Ser150 (H-s), His229 (H-b), Polar: His125, Ser150, Gln153, His229
<b>W318C</b>	Diol	Polar: Thr70, Ser95, Ser96, His125, Ser150, Gln153, His229	His125 (H-s), Polar: Thr70, Ser95, Ser96, His125, Ser150, Gln153, His229	Polar: Thr70, His125, Ser150, Gln153, His229
	Nonhyd	His405 (H-s), Polar: Gln153, His229	Asp152 (H-s), Glu285 (H- s), Polar: Gln153, His229, Gln486	Lys123 (SB), Mg <sup>2+</sup> (SB), Lys302 (SB, pi-cat), His405 (pi-pi), Polar: Ser150, His229, Asn282
<b>E307K +T391S</b>	Diol	Lys123 (H-s), Polar: His125, Ser150, Gln153, His229	Lys123 (SB, H-s), Ser150 (H-s), Polar: Thr70, His125, Ser150, Gln153, Ser229	Lys123 (SB, H-s), His405 (H-s), Polar: Thr70, His125, Ser150, His229
	Nonhyd	Thr70 (H-b), Gly90 (H-b x2), Leu92 (H-b x2), Mg <sup>2+</sup> (MC, SB), Lys123 (SB, H-s), Lys302 (SB), Polar: Thr70, His125, Ser150, His229	Thr70 (H-b), Gly90 (H-b x2), Leu92 (H-b x2), Mg <sup>2+</sup> (MC, SB), Lys123 (H-s), Polar: Thr70, His125, Ser150, His229	Lys123 (H-s), Asn282 (H- s), Mg <sup>2+</sup> (MC), Lys302 (pi- cat), Polar: Thr70, His125, Ser150, His229, Asn282

Table 4.20. The interactions between docked ARSA substrates and eight forms of ARSA in acidic pH (cont.).

Protein Form	State	SGC	SGG	pNCS
E307K	Diol	FGL69 (H-s), Ser150 (H-s), Asp152 (H-b), Gln153 (H-s), Lys302 (SB, H-s x2), Mg <sup>2+</sup> (SB), Polar: His229, Asn282, Ser378	Lys123 (SB, H-s), His125 (H-s), Gln153 (H-s), Lys302 (SB), Mg <sup>2+</sup> (SB), Polar: Thr70, Ser150, His229	FGL69 (H-s), Ser150 (H-s), His125 (pi-pi), His229 (H-b), Lys302 (SB, H-s), Mg <sup>2+</sup> (SB), Polar: Thr70, His125, Gln153, Asn282
	Nonhyd	FGL69 (H-s), Leu92 (H-b), Lys123 (H-s), Asp152 (H-s x2, H-b), Asp173 (H-b), Lys302 (SB, H-s), Mg <sup>2+</sup> (SB), Polar: His125, Ser150, Gln153, His229	His125 (H-s), Ser150 (H-b), Gln153 (H-s), Lys302 (H-s), His405 (H-s), Polar: His125, Ser150, Gln153, His229	FGL69 (H-s), Lys123 (H-s), His125 (pi-pi), Ser150 (H-s), His229 (H-b), Lys302 (SB, H-s), Mg <sup>2+</sup> (SB), Polar: His125, Ser150, Gln153, Asn282

Especially the interactions between the substrates and both forms of WT ARSA, and E307K mutant structures show that GScore values around -6.0 kcal/mol may be the threshold for PC binding since almost all the active site residues and Mg<sup>2+</sup> cation form strong interactions with SGG, SGC and pNCS with both forms of E307K mutant. The results can be compared with the ones as previously done for the docking results of the WT ARSA form from neutral pH simulation. Properties of WT ARSA at acidic pH are given in Table 4.21.

Table 4.21. Physicochemical properties of the natural substrates of ARSA for the top resulting poses of the IF docking calculations on WT forms at acidic pH.

Protein Form	WT - Diol			WT - Nonhydrated		
	SGC	SGG	NCS	SGC	SGG	NCS
Molecules						
XP Gscore	-11.43	-10.05	-7.04	-10.179	-10.264	-7.191
MW	459.5	432.4	235.2	459.5	432.4	235.2
SASA	711.0	651.8	390.9	679.1	747.3	399.5
FOSA	367.8	369.6	0.0	367.3	444.8	0.0
FISA	327.1	279.8	282.8	298.7	299.9	288.3

Table 4.21. Physicochemical properties of the natural substrates of ARSA for the top resulting poses of the IF docking calculations on WT forms at acidic pH (cont.).

Protein Form	WT - Diol			WT - Nonhydrated		
	PISA	13.8	0.0	105.7	10.8	0.0
%HOA	0	21	36	0	18	35
#nonHatm	30	28	15	30	28	15
RO5	2	1	0	2	1	0
PSA	227.6	188.9	138.1	212.4	188.3	137.8
QPlogBB	-3.8	-3.0	-2.0	-3.3	-3.8	-2.2
QPlogPo/w	-2.6	-1.1	-0.6	-2.4	-0.9	-0.6
Glide LE	-0.4	-0.4	-0.5	-0.3	-0.4	-0.5
DeltaG	-11.4	-10.1	-7.0	-10.2	-10.3	-7.2
Ki	4.2E-09	4.3E-08	7.0E-06	3.5E-08	2.9E-08	5.4E-06
p(Ki)	8.38	7.37	5.16	7.45	7.53	5.27
BEI	18.23	17.04	21.93	16.22	17.42	22.40
SEI	3.68	3.90	3.73	3.51	4.00	3.82
LLE	10.93	8.51	5.72	9.85	8.48	5.86

The results in Table 4.21 show that the known substrates of ARSA have high ligand lipophilic efficiency (LLE) and high binding free energy values. These are expected because the substrate of ARSA binds to the protein in lysosome and this binding should be strong for the reaction to occur. These results can be selected as the upper limits in the PC search for ARSA because the aim is to find a molecule which will bind ARSA moderately tight in the active site and will be expelled when the natural ligand comes. In other words, we do not want a molecule to bind tighter than natural substrates in the lysosome. We also see that predicted brain/blood partition coefficients (QPlogBB) of the natural substrates are off limits. It means that for the search of PC we can select molecules with QPlogBB values less than -3.0 when the QikProp results are taken as basis.

#### 4.6.2. Docking of Known Bioactive Molecules of ARSA

Other than the natural substrates, the known bioactive molecules of ARSA, which were taken from ChEMBL database, were docked rigidly to eight forms of ARSA in acidic pH and the results are given in Table 4.22 and Table 4.23. The results in Table 4.22 were listed in the increasing order of GScore values for all forms of ARSA. We also summarized the docking results of ChEMBL's ARSA molecules in the increasing order of Gscore values only for WT-diol form, and sorted the other results of other forms according to the order of WT-diol form's. By this way, we can see the effect of the hydration state of FGL and the mutants on the binding free energy of the molecules from ChEMBL database as shown in Table 4.23.

Table 4.22. XP docking results of the known bioactive molecules of ARSA in increasing order of binding free energy (Gscore in kcal/mol) for both forms of all monomers at acidic pH.

WT				W318C				E307K + T391S			
Diol	Gscore	Nonhyd	Gscore	Diol	Gscore	Nonhyd	Gscore	Diol	Gscore	Nonhyd	Gscore
1162201	-11.1	1162201	-11.8	3351078	-11.8	1162201	-9.0	132722	-11.8	3351078	-14.4
25147	-10.2	3351078	-9.9	1162201	-11.6	3351078	-8.6	25147	-11.2	25147	-13.9
132722	-9.4	132722	-9.3	132722	-11.3	132722	-8.6	274619	-10.2	1533156	-13.4
1533156	-9.2	153057	-8.8	25147	-11.0	153057	-8.3	1162201	-9.9	1162201	-12.7
1622540	-8.8	58	-8.6	1533156	-10.4	58	-8.3	1533156	-9.9	132722	-11.8
1318812	-8.4	1096979	-8.6	52	-9.8	1096979	-8.1	240390	-9.7	780	-11.6
3351078	-8.4	1256283	-8.5	1474884	-9.3	1256283	-7.9	1318812	-9.7	1622540	-11.0
1231700	-8.4	723	-8.4	1096979	-9.2	723	-7.4	780	-9.6	1318812	-10.8
189382	-8.3	1318812	-8.2	265301	-9.1	1318812	-7.4	45068	-9.2	274440	-10.7
245019	-8.1	240390	-8.0	265301	-9.1	240390	-7.3	1474884	-8.9	262083	-10.7

The numbers under the titles of diol and nonhydrated are the ChEMBL ID numbers of the molecules in the tables of docking results for ChEMBL molecules.

Table 4.23. The rigid docking results of the known bioactive molecules of ARSA at acidic pH in the increasing order of Gscore values (kcal/mol) for WT-diol form.

WT				W318C				E307K+T391S			
Diol	Gscore	Nonhyd	Gscore	Diol	Gscore	Nonhyd	Gscore	Diol	Gscore	Nonhyd	Gscore
1162201	-11.1	1162201	-9.9	1162201	-11.6	1162201	-9.0	1162201	-9.9	1162201	-12.7
25147	-10.2	25147	-11.8	25147	-11.0	-	-	25147	-11.2	25147	-13.9
132722	-9.4	132722	-8.8	132722	-11.3	132722	-8.6	132722	-11.8	132722	-11.8
1533156	-9.2	1533156	-7.6	1533156	-10.4	1533156	-7.1	1533156	-9.9	1533156	-13.4
1622540	-8.8	1622540	-6.4	1622540	-6.8	1622540	-5.4	1622540	-8.1	1622540	-11.0
1318812	-8.4	1318812	-8.4	1318812	-9.0	1318812	-7.4	1318812	-9.7	1318812	-10.8
3351078	-8.4	3351078	-8.6	3351078	-11.8	3351078	-8.6	3351078	-8.0	3351078	-14.4
1231700	-8.4	1231700	-6.8	1231700	-7.9	1231700	-5.9	1231700	-7.8	1231700	-9.6
189382	-8.3	189382	-6.1	189382	-6.1	189382	-6.2	189382	-6.3	189382	-8.2
245019	-8.1	245019	-6.9	245019	-8.7	245019	-6.0	245019	-7.6	245019	-9.5
52	-7.7	52	-7.4	52	-9.8	52	-6.8	52	-7.7	52	-9.9
1416659	-7.2	1416659	-8.6	1416659	-8.4	1416659	-7.0	1416659	-7.9	1416659	-10.0
274619	-7.0	274619	-7.7	274619	-6.9	274619	-6.9	274619	-10.2	274619	-8.8
153057	-6.7	153057	-5.1	153057	-4.7	153057	-8.3	153057	-8.6	153057	-8.2
56393	-6.6	56393	-7.5	56393	-7.5	56393	-6.4	56393	-8.3	56393	-9.5
58	-6.5	58	-7.6	58	-8.0	58	-8.3	58	-7.9	58	-8.9
1256283	-5.6	1256283	-6.1	1256283	-7.3	1256283	-7.9	1256283	-7.1	1256283	-8.4
45068	-5.6	45068	-5.2	45068	-6.1	45068	-4.9	45068	-9.2	45068	-8.7
723	-5.5	723	-5.8	723	-6.1	723	-7.4	723	-6.0	723	-8.1
262083	-5.9	262083	-5.7	262083	-6.5	262083	-4.2	262083	-6.7	262083	-10.7

From Table 4.24, we see that most of the ChEMBL molecules bind better in terms of their free energies to the W318C-diol and E307K+T391S-nonhydrated forms than both states of WT ARSA form. We can say that, according to these results in Table 4.24, in the acidic milieu, the presence of E307K mutation makes the molecules bind more favorably to ARSA. On the other hand, the presence of W318C, especially in its nonhydrated form makes the GScore values increase around 1.0 kcal/mol. When we compared the results of the docking calculations for ChEMBL molecules in both pH, we saw that some molecules bind better in diol or nonhydrated forms in acidic pH than the nonhydrated forms in neutral milieu

like CHEMBL3351078 (IUPAC name: [[(2R,3R,4R,5R)-5-(6-aminopurin-9-yl)-3-hydroxy-4-phosphonooxy-tetrahydrofuran-2-yl]methoxy-hydroxy-phosphoryl] [(2S,5R)-5-(3-carbamothioylpyridin-1-ium-1-yl)-3,4-dihydroxy-tetrahydrofuran-2-yl]methyl phosphate).

The main aim is to find a molecule which will bind ARSA in neutral pH better than in the acidic pH and will not occupy the active site when the natural substrate comes. Among the results from the docking calculations on ChEMBL molecules two compounds among the top 10 can be suggested for E307K+T391S mutant and W318C mutant respectively as CHEMBL32658 (IUPAC name: 5-[(6,6-Dimethylheptyl)amino]-2-quinolinol), and CHEMBL1096979 (Benserazide Hydrochloride). MD simulations on these molecules docked complexes can lighten up better what happens on the dynamics of ARSA forms when a ligand is bound.

#### 4.6.3. Docking of aSMase Inhibitors

After the known bioactive molecules, we also performed rigid docking calculations with extra precision (XP) protocol for aSMase inhibitors at acidic pH. The results are given in Table 4.24. ZINC molecule is also an aSMase inhibitor. The name is different because it was retrieved from the ZINC database.

Table 4.24. Top different ten compounds from aSMase inhibitors dataset in increasing order of binding free energy values (GScore in kcal/mol).

WT				W318C			
Diol	Gscore	Nonhyd	Gscore	Diol	Gscore	Nonhyd	Gscore
16e_R	-11.12	8_A	-10.01	8_A	-11.08	5_R	-9.99
6_R	-10.57	13_R	-9.63	5_R	-10.81	8_A	-9.09
8_A	-10.57	5_R	-9.60	16b_R	-10.73	13_R	-8.42
16b_R	-10.16	16d_R	-9.47	16c_R	-10.62	6_R	-8.17
16e_R	-10.07	16a_R	-9.43	13_R	-10.31	4_R	-7.46
16d_R	-9.98	16e_R	-8.74	15c_R	-10.21	16d_R	-7.43
16c_R	-9.96	16b_R	-8.20	16a_R	-10.05	9_R	-7.28

Table 4.24. Top different 10 compounds from aSMase inhibitors dataset in increasing order of binding free energy values (GScore in kcal/mol) (cont.).

WT				W318C			
16a_R	-9.95	6_R	-8.14	15d_R	-9.76	1_R	-7.05
13_R	-9.75	15d_R	-7.88	8_R	-9.76	ZINC34577343	-6.96
18_R	-9.37	4_R	-7.76	4_R	-9.69	15b_R	-6.94
E307K+T391S				E307K			
Diol	Gscore	Nonhyd	Gscore	Diol	Gscore	Nonhyd	Gscore
5_R	-12.89	5_R	-15.246	8_A	-13.784	8_A	-13.156
8_A	-12.396	8_A	-14.785	5_R	-11.84	5_R	-12.851
13_R	-12.187	9_A	-13.397	9_A	-11.84	13_R	-11.699
16c_R	-11.797	16c_R	-12.581	13_R	-11.442	16c_R	-11.002
6_R	-11.433	13_R	-12.466	16b_R	-11.219	4_R	-10.99
16e_R	-11.395	8_R	-12.458	16c_R	-11.122	17b_R	-10.889
16b_R	-11.215	4_R	-12.199	16e_R	-11.118	15a_R	-10.468
4_R	-11.094	16e_R	-12.152	7c_R	-11.039	15c_R	-10.424
16d_R	-10.554	16b_R	-11.968	16a_R	-10.665	15d_R	-10.347
8_R	-10.54	18_R	-11.912	15c_R	-10.642	16d_R	-10.21

For WT-diol form, the top compound is 8\_A because it has the highest average GScore results although it is not the top compound in the list but the third. The situation is the same for W318C-nonhydrated form. For WT-nonhydrated form 8\_A is both the top and the most dispersed compound among the top results and the same is valid for the results of W318C-diol form. For the E307K+T391S-diol form, two compounds, 8\_A and 13\_R, are almost equally distributed among the top results while 8-A has a domination on the list of the top compounds with best GScore values among the results. This situation is similar to the results of the docking calculations in neutral pH in terms of the list of the top compounds.

The physicochemical properties of the top ten resulting aSMase inhibitors were calculated and summarized for each form of ARSA in Table 4.25-Table 4.30.

Table 4.25. The physicochemical properties of the top scoring molecules from the rigid docking calculations of aSMase inhibitors on WT-diol form.

Protein Form	WT-Diol									
	16e_R	6_R	8_A	16b_R	16e_R	16d_R	16c_R	16a_R	13_R	18_R
Molecules	16e_R	6_R	8_A	16b_R	16e_R	16d_R	16c_R	16a_R	13_R	18_R
XP Gscore	-11.12	-10.57	-10.57	-10.16	-10.07	-9.98	-9.96	-9.95	-9.75	-9.37
Mol W	347.3	277.2	427.2	318.2	347.3	263.1	346.3	290.2	251.1	272.1
SASA	529.8	458.9	571.5	523.8	554.0	416.1	544.1	498.3	381.8	395.3
FOSA	278.6	154.7	56.2	286.7	314.1	112.6	325.4	247.2	55.2	26.5
FISA	248.0	301.0	411.5	237.1	235.6	301.2	217.3	248.6	322.2	268.4
PISA	0.00	0.00	94.89	0.00	0.00	0.00	0.00	0.00	0.00	97.85
%HOA	1.42	0.00	0.00	43.31	5.25	0.00	50.43	37.67	0.00	23.38
#nonHatm	21	16	27	19	21	15	21	17	14	16
RO5	1	1	2	0	1	1	0	0	1	0
PSA	156.5	162.0	217.8	140.7	146.9	163.7	138.3	140.9	164.0	148.5
QPlogBB	-2.14	-2.34	-3.82	-2.32	-2.10	-2.10	-2.21	-2.36	-2.18	-2.01
QPlogPo/w	0.14	-1.42	-1.71	2.92	0.44	-1.73	3.56	2.29	-3.44	0.43
glide LE	-0.40	-0.49	-0.32	-0.39	-0.35	-0.49	-0.35	-0.43	-0.55	-0.45
deltaG	-8.42	-7.90	-8.64	-7.47	-7.39	-7.35	-7.27	-7.28	-7.66	-7.20
p(Ki)	6.17	5.79	6.33	5.47	5.42	5.39	5.33	5.33	5.61	5.28
BEI	17.77	20.90	14.82	17.20	15.60	20.48	15.38	18.38	22.36	19.40
SEI	3.94	3.58	2.91	3.89	3.69	3.29	3.85	3.79	3.42	3.55
LLE	6.03	7.22	8.04	2.55	4.98	7.12	1.76	3.04	9.05	4.85

When we look at the BBB partition coefficients of the molecules we see that for the same molecules have different QPlogBB results. This happens because their conformations, bond angles, backbone bending and consequently, solvent accessible surface areas are changing. Since the QPlogBB results of the known ligands, bioactive molecules (molecules from ChEMBL database), and substrates are also below the threshold value we can conclude that the results around -3.0 for QPlogBB can be accepted due to the similarity of the results.

Table 4.26. The physicochemical properties of the top scoring molecules from the rigid docking calculations of aSMase inhibitors on WT-nonhydrated form in increasing order of GScore values.

Protein Form	WT - Nonhydrated									
	8_A	13_R	5_R	16d_R	16a_R	16e_R	16b_R	6_R	15d_R	4_R
XP Gscore	-10.01	-9.63	-9.60	-9.47	-9.43	-8.74	-8.20	-8.14	-7.88	-7.76
Mol W	427.2	251.1	559.2	263.1	290.2	347.3	318.2	277.2	330.3	299.1
SASA	615.0	358.9	719.1	438.0	513.2	595.9	543.9	470.9	641.8	447.3
FOSA	79.5	83.3	118.7	113.6	245.0	289.7	321.9	156.9	438.1	52.1
FISA	423.6	273.2	589.9	320.3	265.0	304.2	218.3	311.6	199.9	389.4
PISA	105.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
%HOA	0.0	0.0	0.0	0.0	34.7	0.0	48.0	0.0	57.0	0.0
#nonHatm	27	14	31	15	17	21	19	16	20	16
RO5	2	1	3	1	0	1	0	1	0	1
PSA	230.4	146.7	316.8	164.6	139.3	166.3	134.0	157.7	109.3	190.7
QPlogBB	-4.27	-1.64	-5.99	-2.42	-2.58	-3.00	-2.24	-2.50	-2.48	-3.01
QPlogPo/w	-1.71	-3.12	-5.66	-1.80	2.26	0.04	3.18	-1.43	4.19	-3.75
glide LE	-0.31	-0.54	-0.29	-0.46	-0.40	-0.29	-0.29	-0.51	-0.39	-0.45
deltaG	-8.29	-7.55	-8.90	-6.84	-6.75	-6.05	-5.51	-8.13	-7.88	-7.18
p(Ki)	6.08	5.53	6.52	5.01	4.95	4.43	4.04	5.96	5.78	5.27
BEI	14.22	22.03	11.66	19.05	17.05	12.77	12.69	21.50	17.49	17.61
SEI	2.64	3.77	2.06	3.05	3.55	2.67	3.01	3.78	5.29	2.76
LLE	7.78	8.65	12.18	6.82	2.69	4.39	0.86	7.39	1.59	9.01

Table 4.27. The physicochemical properties of the top scoring molecules from the rigid docking calculations of aSMase inhibitors on W318C-diol form in increasing order of GScore values.

Protein Form	W318C - Diol									
	8_A	5_R	16b_R	16c_R	13_R	15c_R	16a_R	15d_R	8_R	4_R
Molecules	8_A	5_R	16b_R	16c_R	13_R	15c_R	16a_R	15d_R	8_R	4_R
XP Gscore	-11.08	-10.81	-10.73	-10.62	-10.31	-10.21	-10.05	-9.76	-9.76	-9.69
Mol W	427.2	559.2	318.2	346.3	251.1	316.3	290.2	330.3	265.1	299.1
SASA	564.7	691.0	588.3	619.3	385.8	576.5	482.8	604.9	401.0	444.4
FOSA	76.7	95.4	327.0	397.0	65.8	355.6	237.8	389.0	97.7	58.9
FISA	408.8	588.7	256.1	222.3	318.2	217.2	240.4	214.1	298.1	378.1
PISA	77.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
%HOA	0.0	0.0	40.9	51.3	0.0	50.2	38.6	53.4	0.0	0.0
#nonHatm	27	31	19	21	14	19	17	20	15	16
RO5	2	3	0	0	1	0	0	0	1	1
PSA	231.7	318.3	133.9	131.1	162.5	117.0	128.6	121.8	160.3	187.7
QPlogBB	-3.74	-5.65	-2.85	-2.61	-2.18	-2.37	-2.23	-2.42	-2.02	-2.91
QPlogPo/w	-1.71	-5.56	3.06	3.87	-3.41	3.52	2.21	3.98	-2.75	-3.68
glide LE	-0.32	-0.33	-0.42	-0.38	-0.59	-0.54	-0.43	-0.49	-0.53	-0.57
deltaG	-8.61	-10.11	-8.06	-7.94	-8.22	-10.20	-7.36	-9.76	-7.92	-9.12
p(Ki)	6.31	7.41	5.91	5.82	6.02	7.48	5.40	7.15	5.81	6.69
BEI	14.78	13.25	18.56	16.80	23.99	23.65	18.59	21.66	21.90	22.35
SEI	2.72	2.33	4.41	4.44	3.71	6.39	4.20	5.88	3.62	3.56
LLE	8.03	12.97	2.85	1.95	9.43	3.96	3.19	3.17	8.56	10.36

Table 4.28. The physicochemical properties of the top scoring molecules from the rigid docking calculations of aSMase inhibitors on W318C-nonhydrated form in increasing order of GScore values.

Protein Form	W318C - Nonhydrated									
	5_R	8_A	13_R	6_R	4_R	16d_R	9_R	1_R	34577343	15b_R
Molecules										
XP Gscore	-9.99	-9.09	-8.42	-8.17	-7.46	-7.43	-7.28	-7.05	-6.96	-6.94
Mol W	559.2	427.2	251.1	277.2	299.1	263.1	247.1	110.0	232.4	302.2
SASA	661.7	591.4	380.7	476.2	457.8	447.2	448.4	280.5	544.1	578.0
FOSA	126.9	50.9	56.6	157.5	62.9	121.3	197.6	131.9	364.3	336.0
FISA	522.4	427.0	319.0	315.0	385.1	322.6	247.3	145.7	109.9	240.0
PISA	0.0	106.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
%HOA	0.0	0.0	0.0	0.0	0.0	0.0	6.8	51.8	91.1	44.5
#nonHatm	31	27	14	16	16	15	14	6	15	18
RO5	3	2	1	1	1	1	0	0	0	0
PSA	294.4	232.4	158.8	159.3	189.6	160.8	130.4	64.0	51.6	118.8
QPlogBB	-4.89	-4.10	-2.16	-2.59	-3.05	-2.51	-1.78	-0.71	-0.98	-2.61
QPlogPo/w	-5.19	-1.75	-3.44	-1.46	-3.72	-1.80	-1.18	-0.10	3.75	3.21
glide LE	-0.30	-0.25	-0.45	-0.34	-0.43	-0.32	-0.45	-1.18	-0.43	-0.39
deltaG	-9.30	-6.78	-6.34	-5.50	-6.88	-4.80	-6.29	-7.05	-6.47	-6.95
p(Ki)	6.82	4.97	4.65	4.03	5.04	3.52	4.61	5.17	4.74	5.09
BEI	12.19	11.63	18.52	14.56	16.86	13.37	18.65	46.96	20.39	16.85
SEI	2.32	2.14	2.93	2.53	2.66	2.19	3.53	8.07	9.19	4.29
LLE	12.01	6.72	8.09	5.50	8.76	5.31	5.78	5.26	0.99	1.89

The second compound in Table 4.28 from the last is the one that was retrieved from ZINC database and the numbers are its ZINC ID.

Table 4.29. The physicochemical properties of the top scoring molecules from the rigid docking calculations of aSMase inhibitors on E307K+T391S-diol form in increasing order of GScore values.

Protein Form	E307K+T391S - Diol									
	5_R	8_A	13_R	16c_R	6_R	16e_R	16b_R	4_R	16d_R	8_R
XP Gscore	-12.89	-12.40	-12.19	-11.80	-11.43	-11.40	-11.22	-11.09	-10.55	-10.54
Mol W	559.2	427.2	251.1	346.3	277.2	347.3	318.2	299.1	263.1	265.1
SASA	637.1	569.4	376.8	601.2	463.4	618.6	551.6	438.6	441.5	430.3
FOSA	128.9	81.8	76.7	368.6	157.6	294.3	317.8	57.6	115.3	113.1
FISA	499.4	396.3	299.6	232.5	300.4	320.3	228.4	375.6	321.7	312.8
PISA	0.0	89.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
%HOA	0.0	0.0	0.0	48.1	0.0	0.0	45.4	0.0	0.0	0.0
#nonHatm	31	27	14	21	16	21	19	16	15	15
RO5	3	2	1	0	1	1	0	1	1	1
PSA	286.5	227.8	165.5	124.9	158.3	166.3	124.2	188.3	167.2	160.7
QPlogBB	-4.47	-3.63	-1.91	-2.66	-2.36	-3.28	-2.42	-2.82	-2.44	-2.30
QPlogPo/w	-4.95	-1.56	-3.20	3.61	-1.39	0.07	3.02	-3.62	-1.78	-2.71
glide LE	-0.41	-0.37	-0.86	-0.43	-0.71	-0.54	-0.45	-0.66	-0.70	-0.58
deltaG	-12.65	-10.10	-12.05	-9.11	-11.42	-11.38	-8.53	-10.58	-10.55	-8.72
p(Ki)	9.27	7.40	8.84	6.68	8.37	8.34	6.25	7.75	7.73	6.39
BEI	16.58	17.33	35.19	19.29	30.21	24.02	19.65	25.92	29.38	24.10
SEI	3.24	3.25	5.34	5.35	5.29	5.02	5.03	4.12	4.62	3.97
LLE	14.22	8.96	12.04	3.07	9.77	8.27	3.23	11.37	9.50	9.10

Table 4.30. The physicochemical properties of the top scoring molecules from the rigid docking calculations of aSMase inhibitors on E307K+T391S-nonhydrated form in increasing order of GScore values.

Protein Form	E307K+T391S - Nonhydrated									
	5_R	8_A	9_A	16c_R	13_R	8_R	4_R	16e_R	16b_R	18_R
XP Gscore	-15.25	-14.79	-13.40	-12.58	-12.47	-12.46	-12.20	-12.15	-11.97	-11.91
Mol W	559.2	427.2	1047.1	346.3	251.1	265.1	299.1	347.3	318.2	272.1
SASA	649.1	562.8	1599.3	558.5	380.2	414.0	441.8	604.8	549.3	410.4
FOSA	181.1	67.0	1081.2	326.3	80.2	121.9	47.2	307.6	319.5	28.4
FISA	456.2	406.8	455.4	227.0	296.2	288.5	389.4	292.0	229.9	267.8
PISA	0.0	82.8	55.6	0.0	0.0	0.0	0.0	0.0	0.0	110.5
%HOA	0.0	0.0	0.0	48.3	0.0	0.0	0.0	0.0	45.6	24.2
#nonHatm	31	27	69	21	14	15	16	21	19	16
RO5	3	2	4	0	1	1	1	1	0	0
PSA	282.2	223.9	311.5	126.3	156.1	159.4	191.6	156.5	135.6	148.4
QPlogBB	-4.14	-3.70	-9.40	-2.39	-1.96	-1.97	-2.96	-2.93	-2.38	-2.07
QPlogPo/w	-4.62	-1.69	7.62	3.48	-3.25	-2.56	-3.73	0.18	3.10	0.54
glide LE	-0.47	-0.46	-0.17	-0.60	-0.88	-0.82	-0.73	-0.58	-0.63	-0.74
deltaG	-14.54	-12.31	-11.66	-12.58	-12.33	-12.30	-11.62	-12.14	-11.97	-11.89
p(Ki)	10.66	9.02	8.55	9.22	9.04	9.02	8.51	8.90	8.77	8.71
BEI	19.06	21.13	8.16	26.63	36.01	34.01	28.47	25.62	27.57	32.03
SEI	3.78	4.03	2.74	7.30	5.79	5.66	4.44	5.68	6.47	5.87
LLE	15.27	10.71	0.93	5.74	12.29	11.57	12.24	8.72	5.68	8.17

The LLE and pKi values of the results are above the ideal value for optimal drug candidates [110,111]. This situation is expected since the most of the aSMase inhibitors are the variants of other known inhibitors and antidepressants [109,112,113].

The main outcome is that some molecules are binding ARSA forms tightly in the neutral pH but worse in the acidic pH when compared their GScore values. 5\_R for WT or 14\_R for W318C mutant can be given as examples for this situation. These compounds should be further investigated as the candidate PCs for ARSA. For the mutant structures,

5\_R may not be a proper PC chaperon but was selected for MD simulation on W318C mutant due to its high GScore result. The analysis of this MD should be done in great care and be compared with the results of another MD simulation on other ligand bound W318C mutant form.

#### 4.6.4. Docking Cerebroside 3-Sulfate at Acidic pH

The substrate of ARSA with longer carbon chains, cerebroside 3-sulfate (C3S), was docked to the ARSA forms at acidic pH. Induced fit docking protocol was also applied for the docking calculations of C3S to WT ARSA forms. The results were summarized in Table 4.31.

Table 4.31. Rigid docking results of C3S for the eight forms of ARSA at acidic pH.

<b>Protein Form</b>	<b>WT</b>	<b>WT</b>	<b>W318C</b>	<b>W318C</b>	<b>E307K+T391S</b>	<b>E307K+T391S</b>
State	Diol	Nonhyd	Diol	Nonhyd	Diol	Nonhyd
XP Gscore	-5.64	-7.83	-8.61	-6.06	-7.66	-9.79
Mol W	499.6	499.6	499.6	499.6	499.6	499.6
SASA	682.6	823.9	786.3	713.5	754.8	817.7
FOSA	428.5	515.3	484.6	449.1	480.4	532.4
FISA	247.1	297.0	295.3	256.4	267.3	279.0
PISA	6.55	8.78	3.93	5.57	4.42	3.92
%HOA	10.94	4.85	3.22	9.21	10.72	9.20
#nonHatm	33	33	33	33	33	33
RO5	2	2	2	2	2	2
PSA	187.6	192.6	198.7	187.2	185.2	199.8
QPlogBB	-2.77	-3.94	-3.71	-2.98	-3.29	-3.68
QPlogPo/w	-0.75	-0.69	-0.78	-0.75	-0.75	-0.54
glide LE	-0.17	-0.24	-0.26	-0.18	-0.23	-0.30
deltaG	-5.64	-7.82	-8.61	-6.07	-7.66	-9.80
p(Ki)	4.14	5.73	6.31	4.45	5.61	7.18
BEI	8.28	11.48	12.64	8.91	11.23	14.38

Table 4.31. Rigid docking results of C3S are given for the eight forms of ARSA in acidic pH (cont.).

Protein Form	WT	WT	W318C	W318C	E307K+T391S	E307K+T391S
State	Diol	Nonhyd	Diol	Nonhyd	Diol	Nonhyd
SEI	2.21	2.98	3.18	2.38	3.03	3.60
LLE	4.89	6.42	7.09	5.20	6.36	7.72

The poses of the docked C3S molecules are evaluated and the most similar poses to the proposed mechanism pose are the results for WT-nonhydrated and E307K+T391S-nonhydrated forms. The best GScore values belong to the WT-nonhydrated, W318C-diol and E307K+T391S-nonhydrated forms. According to these results we can take the scores of these poses as basis for further rigid docking calculations. The blood brain barriers partition coefficients of these two best poses are off the limits (less than -3.0) which means that QPlogBB results which are out of limits for other compounds may be still valid in the PC search for MLD.

The interaction maps between docked C3S and the proteins, WT-nonhydrated and E307K+T391S-nonhydrated forms, are shown in Figure 4.60 and Figure 4.61, respectively.

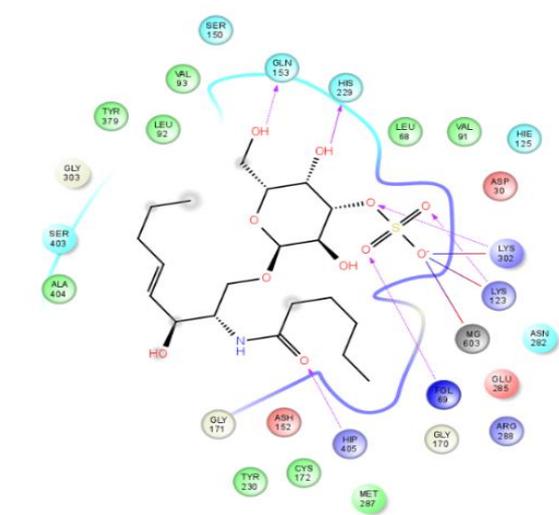


Figure 4.60. The interaction map between rigidly docked C3S (GScore: -7.83 kcal/mol) and the WT-nonhydrated form of ARSA.



summarized the results of the IFD calculations of C3S for also the mutant forms of ARSA through from Table 4.32 to Table 4.37.

Table 4.32. The induced fit docking results of C3S to WT-diol form are given with the calculated physicochemical properties of the all resulting poses in increasing order of GScore values.

Protein Form	WT - Diol							
	#1	#2	#3	#4	#5	#6	#7	#8
Poses of C3S								
XP Gscore	-10.57	-10.50	-10.32	-10.18	-9.94	-9.39	-9.29	-9.00
Mol W	499.6	499.6	499.6	499.6	499.6	499.6	499.6	499.6
SASA	852.3	791.8	735.6	793.1	743.2	764.0	795.6	687.3
FOSA	546.6	512.6	459.6	478.8	467.8	478.8	509.0	420.4
FISA	297.5	269.5	266.4	303.3	266.0	272.6	281.6	256.2
PISA	5.69	7.11	7.17	8.38	7.16	10.21	2.49	8.77
%HOA	5.43	9.62	8.40	2.45	8.84	8.60	8.03	12.16
#nonHatm	33	33	33	33	33	33	33	33
RO5	2	2	2	2	2	2	2	2
PSA	197.5	194.6	193.5	200.5	191.3	200.0	195.9	193.3
QPlogBB	-4.08	-3.47	-3.17	-3.77	-3.19	-3.32	-3.60	-2.86
QPlogPo/w	-0.57	-0.58	-0.72	-0.67	-0.68	-0.56	-0.62	-0.79
glide LE	-0.32	-0.32	-0.31	-0.31	-0.30	-0.28	-0.28	-0.27
deltaG	-10.56	-10.49	-10.33	-10.20	-9.93	-9.37	-9.31	-9.01
p(Ki)	7.74	7.69	7.57	7.47	7.28	6.87	6.82	6.60
BEI	15.49	15.40	15.16	14.96	14.57	13.75	13.65	13.22
SEI	3.92	3.95	3.91	3.73	3.81	3.44	3.48	3.42
LLE	8.31	8.27	8.29	8.14	7.96	7.43	7.44	7.39

There are significant differences between the poses of the 8<sup>th</sup> conformation of C3S and other resulting conformations in Table 4.33. The most similar conformation to the proposed mechanism conformation is the last pose with GScore value of -9.00 kcal/mol. The interaction maps of the first and the last poses are shown in Figure 4.62 and Figure 4.63.



In Figure 4.63, the purple dashed lines show hydrogen bonds between the atoms of the substrates and the side chains of the residues whereas solid purple lines are for the hydrogen bonds with the backbone of the residues. Blue-red transitive colored lines indicate the salt bridges.

We concluded that the last pose is the most similar pose to the conformation of C3S in the proposed reaction mechanism. It forms salt bridge and hydrogen bonds with Arg73 and Lys123 which are both active site residues as well as Ser150 as shown in Figure 4.63.

Table 4.33. The induced fit docking results for the top ten poses of C3S to WT-nonhydrated form with their physicochemical properties in increasing order of GScore.

Protein Form	WT - Nonhydrated									
	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10
Poses of C3S	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10
XP Gscore	-10.00	-9.94	-8.90	-8.60	-8.50	-8.42	-8.05	-7.94	-7.73	-7.69
Mol W	499.6	499.6	499.6	499.6	499.6	499.6	499.6	499.6	499.6	499.6
SASA	791.9	822.9	864.1	851.4	808.4	802.9	781.6	843.9	789.1	860.3
FOSA	518.4	514.0	552.9	551.2	507.7	518.2	497.3	531.5	511.6	540.7
FISA	261.3	293.7	304.0	294.0	291.5	277.1	273.3	302.3	268.0	308.3
PISA	9.76	12.89	4.83	3.51	6.83	5.13	8.85	7.49	7.03	8.71
%HOA	15.33	5.39	4.68	8.08	4.94	8.11	9.73	4.61	13.50	3.42
#nonHatm	33	33	33	33	33	33	33	33	33	33
RO5	2	2	2	2	2	2	2	2	2	2
PSA	195.1	195.1	198.3	199.3	189.9	188.1	194.9	193.9	192.9	203.3
QPlogBB	-3.32	-3.90	-4.22	-4.03	-3.80	-3.67	-3.43	-4.15	-3.40	-4.24
QPlogPo/w	-0.36	-0.69	-0.57	-0.54	-0.73	-0.77	-0.56	-0.75	-0.49	-0.59
glide LE	-0.30	-0.30	-0.27	-0.26	-0.26	-0.26	-0.24	-0.24	-0.23	-0.23
deltaG	-10.00	-9.93	-8.91	-8.61	-8.48	-8.42	-8.05	-7.95	-7.72	-7.69
p(Ki)	7.33	7.28	6.53	6.31	6.22	6.17	5.90	5.83	5.66	5.64
BEI	14.67	14.57	13.07	12.64	12.44	12.35	11.81	11.67	11.33	11.28
SEI	3.76	3.73	3.29	3.17	3.27	3.28	3.03	3.01	2.93	2.77
LLE	7.69	7.97	7.10	6.85	6.94	6.94	6.46	6.58	6.15	6.22

GScore values of the resulting poses of WT ARSA are higher than those of nonhydrated form but when we examined the poses we saw that C3S poses are better in terms of having the sulfate group entered in the active site cavity and located near to Mg<sup>2+</sup> cation and FGL69 residue. The Gscore difference between the last pose and the first pose of C3S molecules docked to diol form is 1.57 kcal/mol, and the results of the best poses from the docking calculations on the nonhydrated form of WT ARSA are closer to the last pose from the docking calculation on diol form. The interactions between the carbon chain and the residues in the active site cavity may be more favorable in terms of binding free energy, but we want sulfate group or other group of PC candidates to enter the active site cavity and position near to the cation. The results around -9.00 kcal/mol are still significantly good for the docking calculations on ARSA forms and can be selected as the upper limit for further docking calculations for PC search.

Table 4.34. The induced fit docking results of the top ten poses of C3S to W318C-diol form in increasing order of GScore.

Protein Form	W318C - Diol										
	6_1	1	2	3	4	5	6	7	8	9	10
Poses of C3S	6_1	1	2	3	4	5	6	7	8	9	10
XP Gscore	-10.99										
Docking score	-10.99	-5.39	-4.15	-4.06	-3.64	-3.49	-3.45	-3.27	-3.02	-2.99	-2.97
Mol W	499.6	499.6	499.6	499.6	499.6	499.6	499.6	499.6	499.6	499.6	499.6
SASA	808.6	828.9	789.4	836.3	848.5	843.1	859.8	786.9	851.4	806.7	810.1
FOSA	505.3	529.5	488.4	506.0	566.1	541.3	562.1	469.0	537.8	518.2	537.4
FISA	293.6	290.2	289.3	319.7	273.7	288.6	284.3	303.4	302.3	278.9	260.2
PISA	7.24	7.05	10.5	8.16	5.94	10.6	10.9	11.8	8.9	7.5	10.3
%HOA	3.92	7.62	5.88	0.44	9.44	6.84	7.75	2.86	4.72	10.89	14.35
#nonHatm	33	33	33	33	33	33	33	33	33	33	33
RO5	2	2	2	2	2	2	2	2	2	2	2
PSA	198.4	203.4	196.6	199.5	190.9	194.7	198.8	194.5	203.4	200.4	199.7
QPlogBB	-3.81	-3.87	-3.68	-4.26	-3.83	-3.99	-3.99	-3.76	-4.16	-3.64	-3.43
QPlogPo/w	-0.71	-0.58	-0.76	-0.81	-0.53	-0.66	-0.51	-0.75	-0.66	-0.61	-0.38
glide LE	-0.33	-0.16	-0.13	-0.12	-0.11	-0.11	-0.11	-0.10	-0.09	-0.09	-0.09

Table 4.34. The IFD results of the top ten poses of C3S to W318C-diol form (cont.).

Protein Form	W318C - Diol										
	deltaG	-10.99	-5.38	-4.16	-4.06	-3.63	-3.50	-3.47	-3.27	-3.04	-3.00
p(Ki)	8.05	3.94	3.05	2.98	2.66	2.56	2.54	2.39	2.23	2.20	2.18
BEI	16.12	7.89	6.10	5.96	5.33	5.13	5.08	4.79	4.45	4.41	4.36
SEI	4.06	1.94	1.55	1.49	1.39	1.32	1.28	1.23	1.09	1.10	1.09
LLE	8.77	4.52	3.80	3.79	3.19	3.23	3.05	3.14	2.88	2.81	2.56
glide hbond	-0.26	-1.42	-0.46	-0.87	-0.19	-0.39	-0.26	-0.12	-0.68	-0.67	-0.08

Table 4.35. The IFD results of the top ten poses of C3S to W318C-nonhydrated form.

Protein Form	W318C - Nonhydrated									
	1	2	3	4	5	6	7	8	9	10
Poses of C3S										
XP Gscore	-10.69	-10.54	-10.51	-9.84	-9.78	-9.76	-9.68	-9.60	-9.08	-8.77
Mol W	499.6	499.6	499.6	499.6	499.6	499.6	499.6	499.6	499.6	499.6
SASA	826.2	848.8	801.3	859.0	807.5	780.2	823.2	823.9	795.7	865.9
FOSA	541.7	536.6	526.1	539.6	492.2	488.9	531.3	531.3	495.6	534.8
FISA	275.8	298.0	266.7	305.0	302.6	278.9	281.3	277.4	287.3	317.3
PISA	6.71	11.55	5.84	11.86	10.14	9.78	7.79	12.61	10.29	11.35
%HOA	7.50	6.76	12.41	4.31	3.37	6.73	9.28	7.25	8.52	1.30
#nonHatm	33	33	33	33	33	33	33	33	33	33
RO5	2	2	2	2	2	2	2	2	2	2
PSA	189.7	202.2	196.8	197.3	191.9	192.6	199.5	190.0	188.7	201.2
QPlogBB	-3.74	-4.09	-3.50	-4.21	-3.96	-3.55	-3.80	-3.72	-3.67	-4.43
QPlogPo/w	-0.61	-0.63	-0.59	-0.60	-0.90	-0.79	-0.68	-0.54	-0.68	-0.74
glide LE	-0.32	-0.32	-0.32	-0.30	-0.30	-0.30	-0.29	-0.29	-0.28	-0.27
deltaG	-10.69	-10.53	-10.49	-9.83	-9.77	-9.77	-9.67	-9.60	-9.08	-8.78
p(Ki)	7.84	7.72	7.69	7.21	7.16	7.16	7.09	7.04	6.65	6.43
BEI	15.69	15.45	15.40	14.43	14.33	14.33	14.19	14.09	13.32	12.88
SEI	4.13	3.82	3.91	3.65	3.73	3.72	3.55	3.71	3.53	3.20
LLE	8.45	8.35	8.28	7.81	8.06	7.95	7.76	7.58	7.33	7.18

Table 4.36. The induced fit docking results of the top ten poses of C3S to E307K+T391S-nonhydrated form with their physicochemical properties in increasing order of GScore.

Protein Form	E307K+T391S - Diol									
	1	2	3	4	5	6	7	8	9	10
Poses of C3S										
XP Gscore	-10.97	-10.85	-10.78	-10.48	-10.46	-10.37	-10.32	-10.24	-9.85	-9.75
Mol W	499.6	499.6	499.6	499.6	499.6	499.6	499.6	499.6	499.6	499.6
SASA	793.8	790.4	722.9	815.4	810.8	860.8	835.5	828.6	766.1	844.2
FOSA	487.4	480.4	424.5	530.8	524.0	556.5	523.1	532.2	487.3	544.8
FISA	294.9	302.7	292.6	274.2	274.1	291.1	299.3	286.4	269.6	284.4
PISA	8.96	4.74	3.06	7.92	10.14	10.73	10.90	7.51	6.53	12.28
%HOA	5.97	3.45	4.02	8.96	12.14	8.05	5.60	7.76	9.38	9.89
#nonHatm	33	33	33	33	33	33	33	33	33	33
RO5	2	2	2	2	2	2	2	2	2	2
PSA	202.7	201.6	186.3	195.0	200.9	199.7	195.6	188.9	180.9	198.3
QPlogBB	-3.7	-3.82	-3.439	-3.609	-3.577	-4.063	-4.012	-3.899	-3.386	-3.914
QPlogPo/w	-0.644	-0.847	-1.165	-0.482	-0.473	-0.518	-0.637	-0.732	-0.772	-0.561
glide LE	-0.333	-0.329	-0.327	-0.318	-0.317	-0.314	-0.313	-0.31	-0.298	-0.296
deltaG	-10.99	-10.86	-10.79	-10.49	-10.46	-10.36	-10.33	-10.23	-9.83	-9.77
p(Ki)	8.05	7.96	7.91	7.69	7.67	7.60	7.57	7.50	7.21	7.16
BEI	16.12	15.93	15.83	15.40	15.35	15.20	15.16	15.01	14.43	14.33
SEI	3.97	3.95	4.24	3.95	3.82	3.80	3.87	3.97	3.98	3.61
LLE	8.70	8.81	9.07	8.17	8.14	8.11	8.21	8.23	7.98	7.72

Table 4.37. IF docking results of top ten poses of C3S to E307K+T391S-nonhydrated form with their calculated physicochemical properties in increasing order of GScore.

Protein Form	E307K+T391S - Nonhydrated									
	1	2	3	4	5	6	7	8	9	10
Poses of C3S										
XP Gscore	-11.77	-11.61	-11.57	-11.50	-11.28	-11.11	-10.70	-10.64	-10.55	-9.97
Mol W	499.6	499.6	499.6	499.6	499.6	499.6	499.6	499.6	499.6	499.6
SASA	809.6	860.4	799.0	827.2	868.0	782.9	793.9	756.7	819.3	842.0
FOSA	518.5	544.9	515.5	547.3	541.3	492.8	512.9	486.4	527.2	538.2
FISA	278.6	307.6	273.6	274.8	317.1	278.0	273.9	259.1	277.4	289.9
PISA	10.22	5.47	7.27	2.31	7.29	9.13	4.27	11.19	11.98	11.25
%HOA	8.68	2.57	10.36	9.83	1.18	8.17	9.08	11.43	7.26	7.77
#nonHatm	33	33	33	33	33	33	33	33	33	33
RO5	2	2	2	2	2	2	2	2	2	2
PSA	190.5	197.5	195.5	191.7	204.4	194.1	182.8	194.9	190.5	197.8
QPlogBB	-3.67	-4.26	-3.55	-3.75	-4.39	-3.51	-3.54	-3.17	-3.72	-3.95
QPlogPo/w	-0.63	-0.64	-0.61	-0.65	-0.64	-0.66	-0.65	-0.54	-0.63	-0.58
glide LE	-0.36	-0.35	-0.35	-0.35	-0.34	-0.34	-0.32	-0.32	-0.32	-0.30
deltaG	-11.78	-11.62	-11.58	-11.48	-11.29	-11.12	-10.69	-10.66	-10.56	-9.97
p(Ki)	8.64	8.51	8.49	8.42	8.27	8.15	7.84	7.81	7.74	7.31
BEI	17.29	17.04	17.00	16.85	16.56	16.32	15.69	15.64	15.49	14.62
SEI	4.53	4.31	4.34	4.39	4.05	4.20	4.29	4.01	4.06	3.69
LLE	9.26	9.16	9.10	9.06	8.91	8.81	8.49	8.35	8.37	7.88

WT-diol form binds to C3S with better binding free energy score than the WT-nonhydrated form does. The differences between the GScore results, especially of the top poses, are not significantly high but when the reaction mechanism is considered we expect to see that diol form has higher GScore values. The only question in the case of binding of C3S to WT ARSA is that the resulting poses from the docking calculations on WT-diol form are not overlapping with the known binding pose of CNS in the PDB structure of C69A mutant ARSA (PDB code: 1E2S). The only pose that is close to the proposed binding pose of C3S or the pose of CNS in the PDB structure is the last pose of C3S in Table 4.32.

The Gscore values of the resulting poses are missing from docking calculations on W318C-diol form because the difference between the top two poses (6\_1 and 1) is very high thus only the first pose was generated as the output. Instead of their GScore values, docking scores of the results were listed. When we examined the results from this docking, we selected the poses 1, 2, 3, 7, 9, 10, 11, 13, and 17 as the good ones for MD simulations on W318C-diol form, even though only the top ten results are shown in Table 4.34.

The average GScore results of C3S binding calculations become better when the mutation E307K is present. This is the general outcome of the docking calculations meaning that the presence of E307K mutation may be causing conformational changes on the backbone such that the ligand can enter the active site easier, and bind with more interactions. This is not expected, since E307K causes MLD.

#### **4.7. What Happens When a Ligand is Docked to Mutant Structure?**

After deciding on the protein-ligand complexes we performed MD simulations on these structures. The analyses of the MD simulations were done as previously shown in the sections on MD simulations on ARSA monomers at neutral and acidic pH. The calculations of RMSD, RMSF, possible salt bridges, hydrogen bond occupancies and secondary structure were performed and discussed further.

Eight simulations were performed on WT-ligand and W318C mutant-ligand complexes at neutral pH and one was performed on W318C mutant-natural substrate complex at acidic pH. We performed the simulations at the neutral pH because the general working mechanism of the chaperone therapy is defined as their binding to the misfolded or partially folded proteins just after being synthesized in ER, and their stabilization to help the protein pass the ER control mechanism before its segregation or intracellular transport to other cell compartments like lysosome. They are expected to act as reversible inhibitors, but also to accompany the protein through its journey to the corresponding compartment [28–33,114,115].

#### 4.7.1. MD Simulations on Apo and Holo Forms of W318C, and the Selected *de novo* Synthesized Molecule (D28) Complexes

Three simulations on D28 bound (two IF docked and one rigidly docked with XP) forms of W318C monomers were performed at neutral pH. The results of the analyses, and visual examination showed that D28 cannot preserve its location in the active site and changes its conformation severely. The snapshots from the simulations on IF docked forms, and rigidly docked form were taken and shown from Figure 4.64 to Figure 4.69 respectively. Since these ligands cannot maintain their binding positions and conformations we did not follow all the previous steps to analyze the simulations.

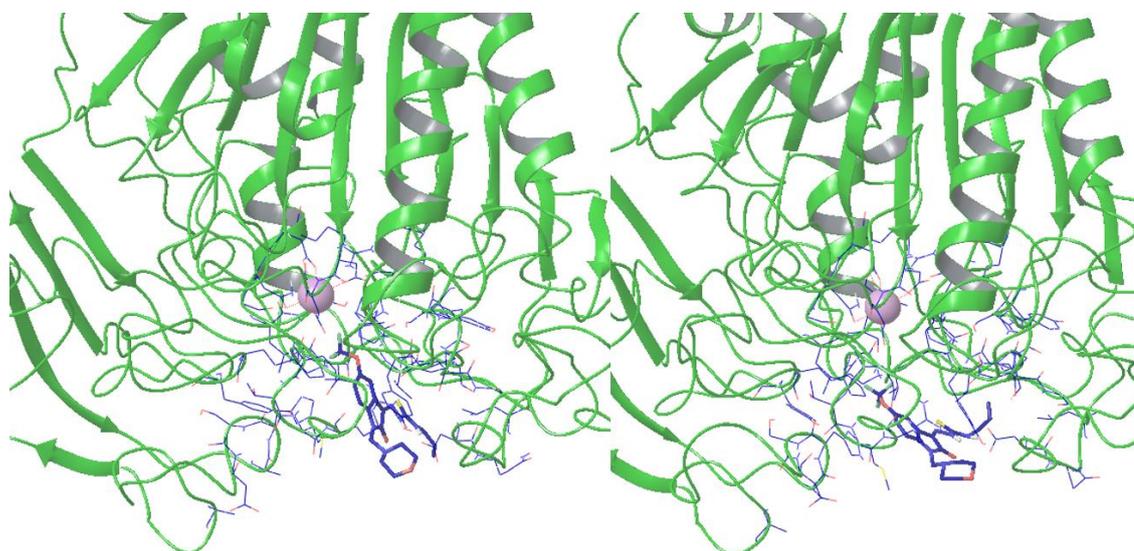


Figure 4.64. D28 docked W318C monomers at 0 ns and 22 ns of simulation at neutral pH.

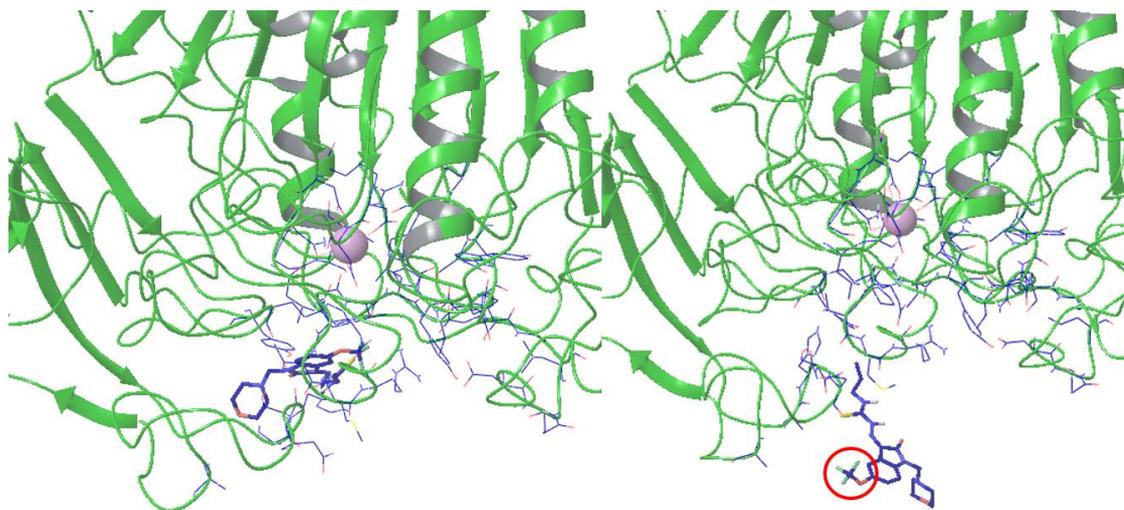


Figure 4.65. Snapshots at 36 ns (left) and 50 ns (right) of the first simulation on IF docked D28 + W318C monomer complex at neutral pH.

Red circle in Figure 4.65 shows the head group of D28 that were in the active site at 0 ns.

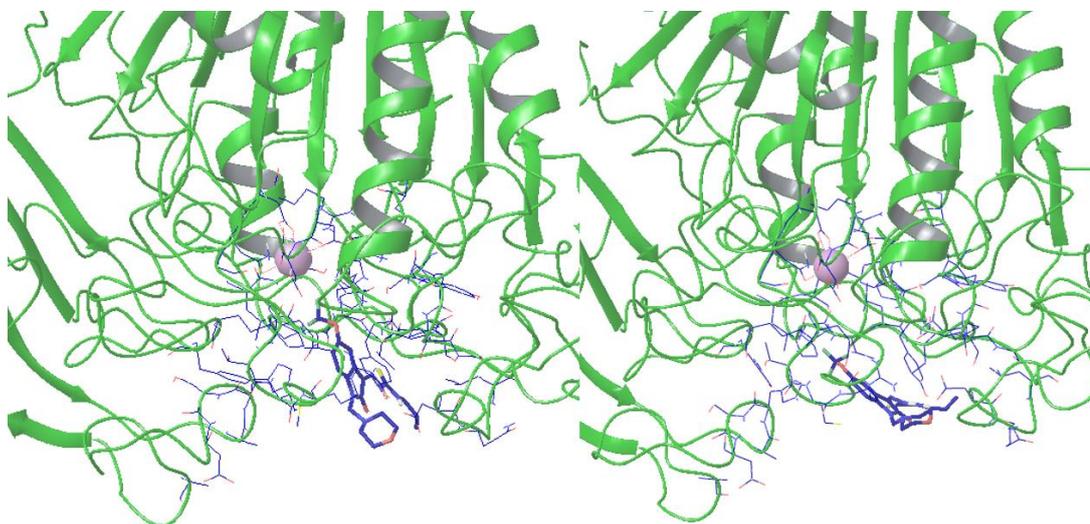


Figure 4.66. Snapshots at 0 ns (left) and 26 ns (right) of the second simulation on IF docked D28 + W318C monomer complex at neutral pH.

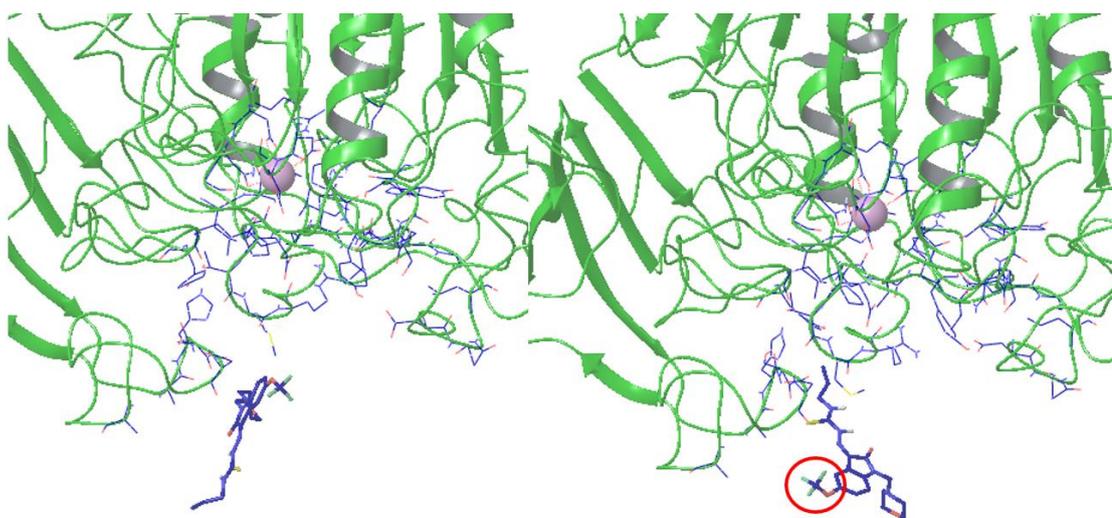


Figure 4.67. Snapshots at 46 (left) and 50 ns (right) of the second simulation on IF docked D28 + W318C monomer complex at neutral pH.

Red circles in Figure 4.67, Figure 4.68, and in Figure 4.69 show the head group of D28 that were in the active site at 0 ns.

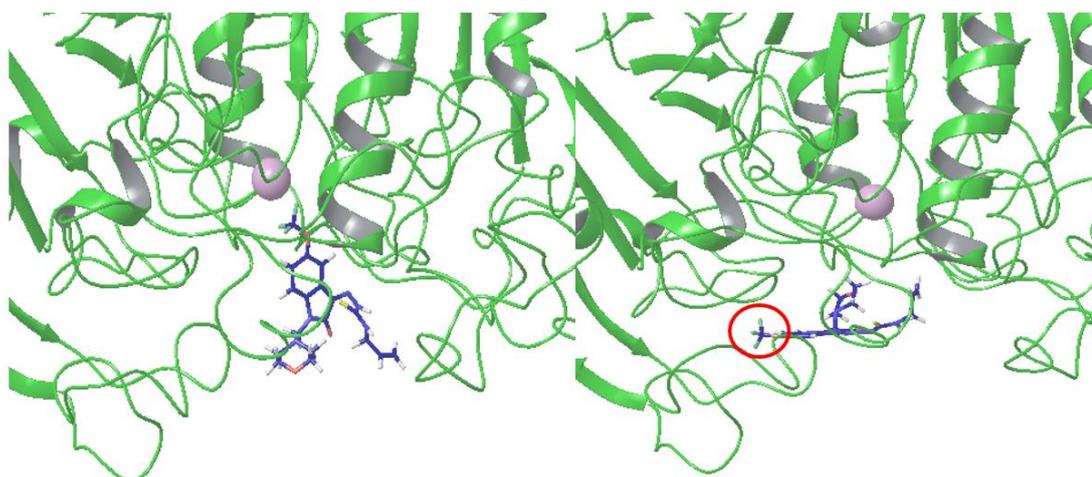


Figure 4.68. Snapshots at 0 ns (left) and 11 ns (right) of the second simulation on rigidly docked D28 + W318C monomer complex at neutral pH.

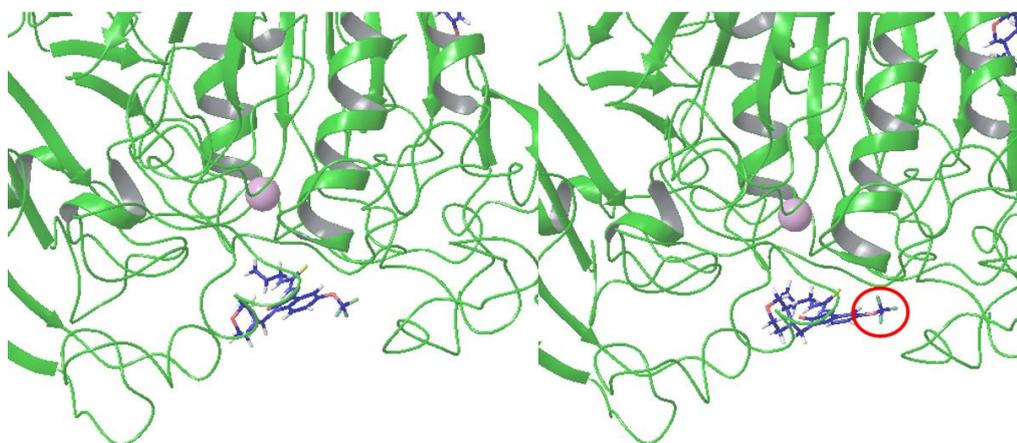


Figure 4.69. Snapshots at 43 ns (left) and 50 ns (right) of the second simulation on rigidly docked D28 + W318C monomer complex at neutral pH.

In addition to these, we calculated the RMSD profiles of one of the IF docked D28 + W318C mutant complex and the rigidly docked D28 + W318C monomer complex. The results are plotted along with the WT, apo form of W318C and other ligand bound forms of W318C monomer at neutral pH as follows. First, the dimerization1 site was analyzed as we concluded that it may have an importance to give clue about the stability of the enzyme. The results are given in Figure 4.70.

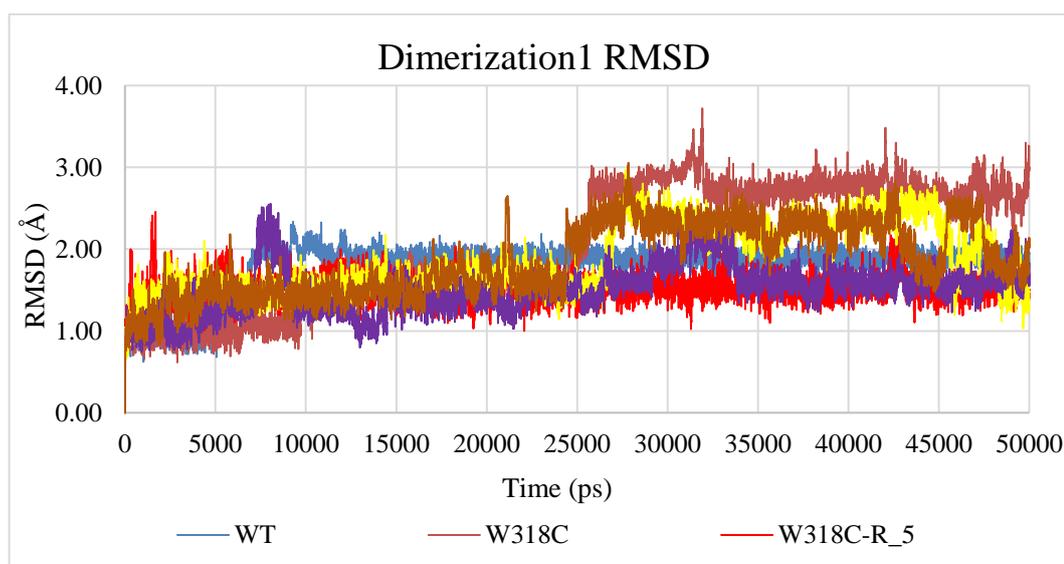


Figure 4.70. RMSD profiles of dimerization1 region for WT, apo W318C mutant and ligand bound W318C mutant monomers at neutral pH.

The dimerization1 RMSD results showed that binding of rigidly docked D28 and R\_5 molecules make dimerization1 region less deviate from the initial coordinates although D28 cannot preserve its position in the active site. On the other hand, IF docked D28 molecule cannot make dimerization1 region remain rigid and after 25 ns the region changes its conformation.

The dimerization2 region becomes more rigid when any of the ligands are bounded to W318C and their RMSD profiles show a lower trend as depicted in Figure 4.71.

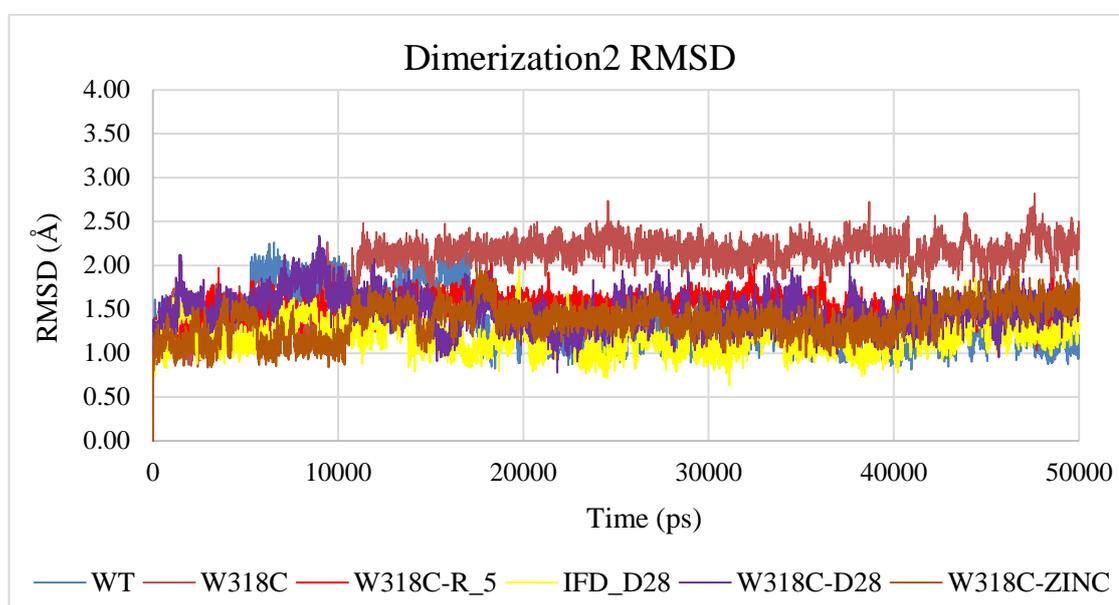


Figure 4.71. RMSD profiles of dimerization2 for WT, apo W318C mutant, and ligand bound W318C mutant monomers at neutral pH.

When we look at the RMSD results of the dimerization3 region we saw that IF docked D28 cannot enhance the stability of the region while rigidly docked D28 does, but since this is a very flexible region and moves a lot also in the simulation on WT forms, the stability of this region may not be an improvement for the mutant structure and dynamics.

Lastly, when we analyzed the RMSD results of the octamerization helices we saw that IF docked D28 makes the octamerization helix more stable while rigidly docked D28 cannot. We did not show the results, but the RMSD profile of the octamerization helix in the complex

of rigidly docked D28 + W318C mutant monomer becomes similar to one of apo form of W318C and rises to 3.0 Å while WT form reaches a plateau around 2.0 Å after 10 ns.

In conclusion, due to instable dynamics of the D28, we concluded that it cannot be a selective and should be improved for more tight binding to be a good PC candidate for ARSA.

#### **4.7.2. MD Simulations on R\_5 Bound WT and W318C Mutant Monomers**

We selected a ligand-protein complex after the docking results for further analysis of the dynamics and structure of ARSA. The first candidate is the top compound for all forms of ARSA (R\_5) among the docking results of known aSMase inhibitors. The inhibitors were docked to the active sites of the two types of ARSA (WT and W318C) at both neutral and acidic pH.

When we look at the docking results of the known aSMase inhibitors and the natural ligands we see that, at the neutral pH, the binding free energy of the natural ligand (SGC) and the selected inhibitors differs approximately by 0.83 kcal/mol. According to the MD simulation analysis we concluded that R\_5 can be a good PC candidate, the p(Ki) values of SGC and the selected ligand are 7.54 and 7.18, respectively. For neutral pH, we can conclude that the selected inhibitor (R\_5) can be a competitive inhibitor for ARSA. Some of the molecules cannot preserve their positions in the active site and their conformations as well, thus we eliminated them. These results are shown and discussed in more detail in the next sections.

At neutral pH, the docking calculations showed that the most tightly bound compound is R\_5. We selected this compound for MD simulations also for its binding pose is close to the natural substrate's. It interacts with many active site residues on both monomers. The docking calculations were performed on the coordinates of the last frame structures of the simulations after they were minimized. The analyses were performed on MD simulations by taking the first frame as the reference structure of these new simulations on ligand-protein complexes.

4.7.2.1. Overall Structure and Active Site. RMSD calculations for the backbones and active sites of the ligand bound WT and W318C mutant forms were performed and the results are given in Figure 4.72. The backbone consists of the coordinates of the alpha carbon atoms, C=O groups, and nitrogen atoms of the structures.

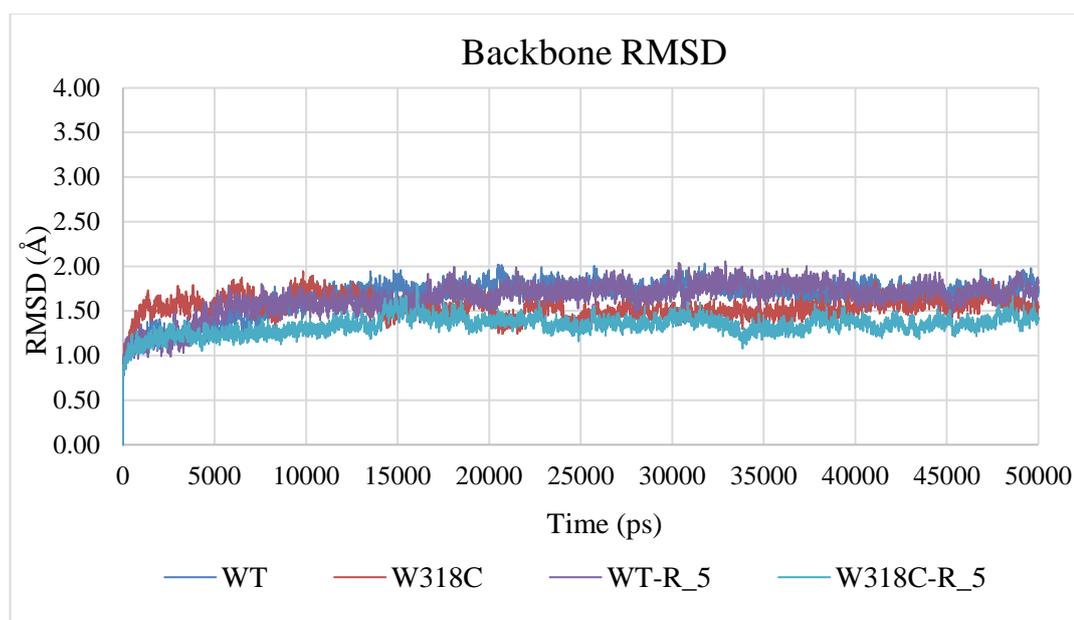


Figure 4.72. RMSD profiles of the backbones for the apo and ligand docked WT and W318C mutant monomers of ARSA are given.

The results show that both protein-ligand complexes reach equilibrium after 15 ns. W318C mutant seems to obtain a more rigid backbone structure due to ligand binding but the RMSD values when the profiles reach the plateau are not significantly different. The results in Figure 4.72 show that ligand binding to W318C mutant form makes the backbone structure less fluctuate while it does not affect the stability of the overall protein in WT form. On the other hand, apo W318C form has the lowest trend among the backbone RMSD profiles of the apo forms of ARSA in neutral pH (Figure 4.1) although their deviations from the initial structure are not significantly different in terms of RMSD calculations. The trends are close to each other and a significant change was not observed between the backbone stabilities of the structures. Thus, we checked the RMSD profiles of the specific regions on ARSA forms and compared their trends as well. The active site RMSD profiles are shown in Figure 4.73. The RMSD profiles of the active sites do not change significantly due to

ligand binding. The difference between the RMSD values of WT forms (apo and ligand bound) at the equilibrium, after 20 ns, is around 0.35 Å. On the other hand, the profiles of the W318C forms follow almost the same trend especially after equilibrium (after 20 ns).

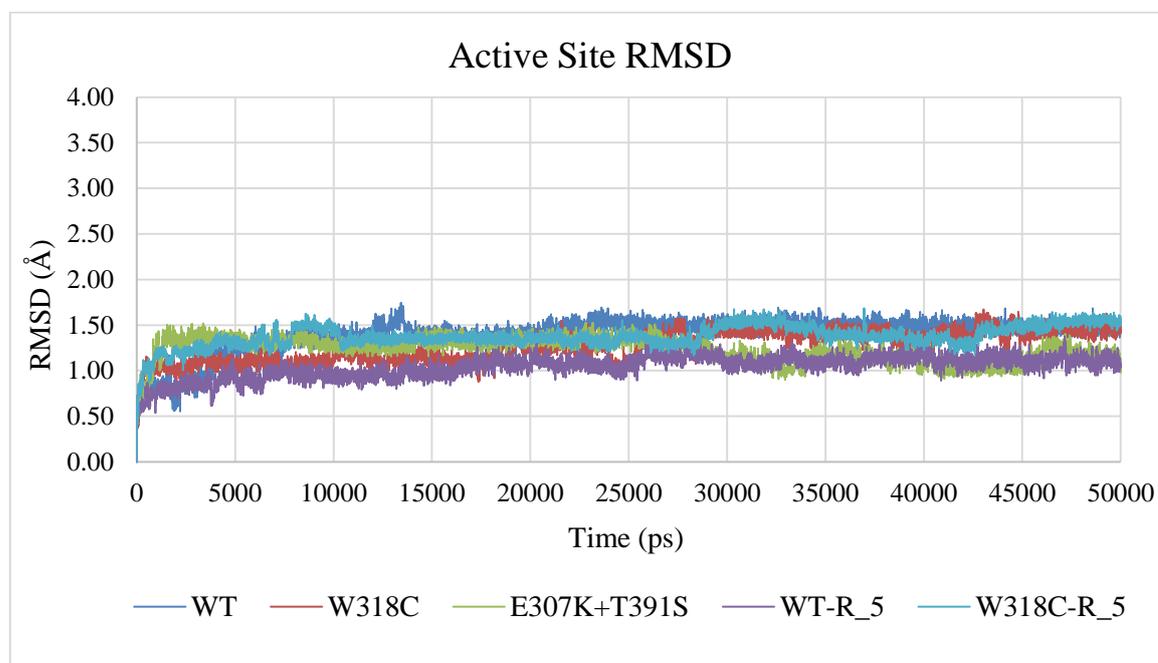


Figure 4.73. RMSD profiles of the active site carbon alpha atoms at neutral pH are given.

From the results in Figure 4.73, we can say that mutations do not affect the active site structure, and the dynamics for W318C mutant so that there is not any significant change on the dynamics of the active site cavity. In addition, the active site of WT form becomes more rigid due to ligand binding as expected.

We also investigated the effects of ligand binding on the residue movements and overall dynamics by calculating the RMSF values of the alpha carbons of the ARSA forms as shown in Figure 4.74 and Figure 4.75.

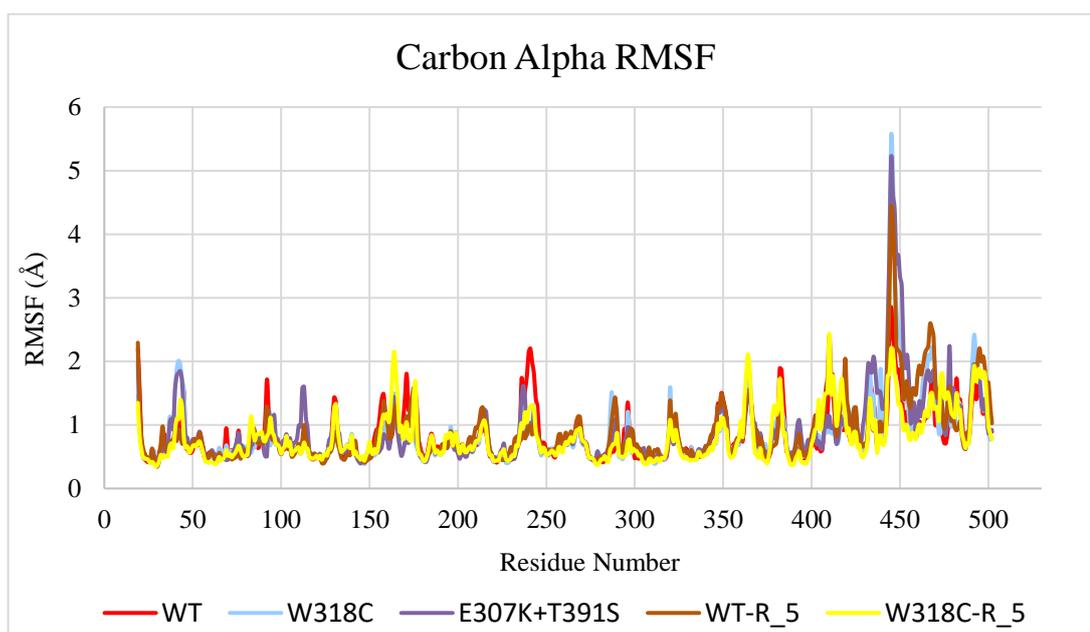


Figure 4.74. Residue based RMSF values for WT, W318C mutant, E307K+T391S mutant, WT + ligand, and W318C +ligand simulations.

Ligand binding has an enhancive effect on the RMSF result of the region that was missing in the crystal structure. This region is very flexible and shows various motion patterns in each simulation, thus the effect of ligand binding on dimerization<sup>3</sup> region may not be significant. On the other hand, the core region (residues 150-180 and around 240) shows a decreased RMSF trend with ligand binding. We know that aSMase inhibitors bind to ARSA tightly and the stabilizing effect of the binding is clearly seen from RMSF calculations.

When we analyze Figure 4.75 for comparison of the changed behavior of W318C mutant, we see that missing loop part, region between residues 280-330, C-terminus, and the core region between residues 30-60 show more stable profiles in terms of protein dynamics. On the other hand, two regions (between residues 160-180 and 400-420) have significantly higher RMSF results after ligand binding.

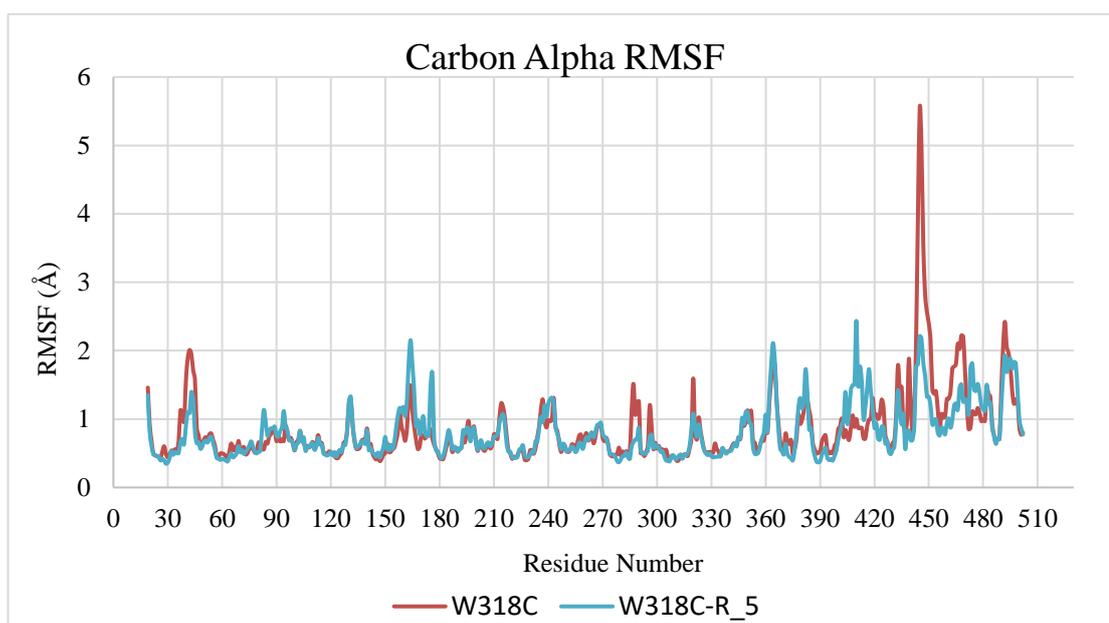


Figure 4.75. Residue based RMSF values for apo and ligand bound W318C mutant monomers.

Two regions (between residues 160-180 and 400-420) are located near to C-terminus and the bottom of the dimer interface respectively. The former has no known significance in terms of enzyme dynamics, folding and oligomerization, but the latter one may be crucial for dimerization of the enzyme. When we compare the RMSF results of WT and R\_5 bound W318C mutant, we see that residues between 400 and 420 still have higher RMSF values in the simulation on ligand bound W318C monomer. If the reason of inactivity due to W318C mutation is not a defect in oligomerization, this difference may not be significant since this region is a long loop but the further investigation should be made on the dimer forms of WT and mutant ARSA proteins.

The possible changes of the mutants were investigated further for the other specific regions on ARSA as follows.

4.7.2.2. Specific Regions on ARSA. The specific regions were indicated on ARSA as dimerization, octamerization and glycosylation regions. These were summarized before in Table 4.4. Accordingly, the RMSD profiles of the dimerization1, dimerization2, dimerization3, octamerization and glycosylation regions were calculated and shown below. The dimerization1 region RMSD profiles of the ARSA forms at neutral pH are given in Figure 4.76 and discussed as the following.

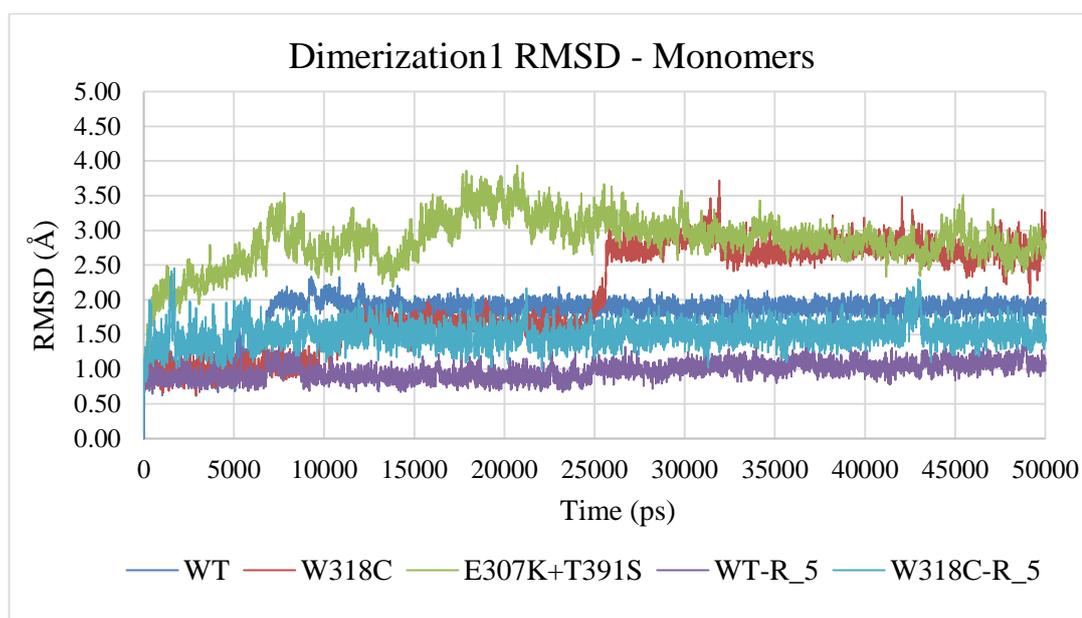


Figure 4.76. RMSD profiles of dimerization1 loops for apo and ligand bound ARSA monomers at neutral pH.

The results show that ligand binding significantly affects the dynamics of the dimerization1 region. It is the region that is the closest to the active site, and contains two active site residues, Asp29 and Asp30. It is expected to see this decrease and flat trend in dimerization1 region due to its closeness to the ligand binding site, the active site. The differences between the RMSD values of both apo forms of WT and W318C mutant and their ligand bound forms are significant for dimerization1 region. The differences occur due to the limitation on the movements of the residue Pro42 for W318C mutant form. In the simulation on the apo form of W318C mutant Pro42 turns its side chain upside but in the simulation on ligand bound W318C mutant form it stays rigid as shown in Figure 4.77 and Figure 4.78, respectively.

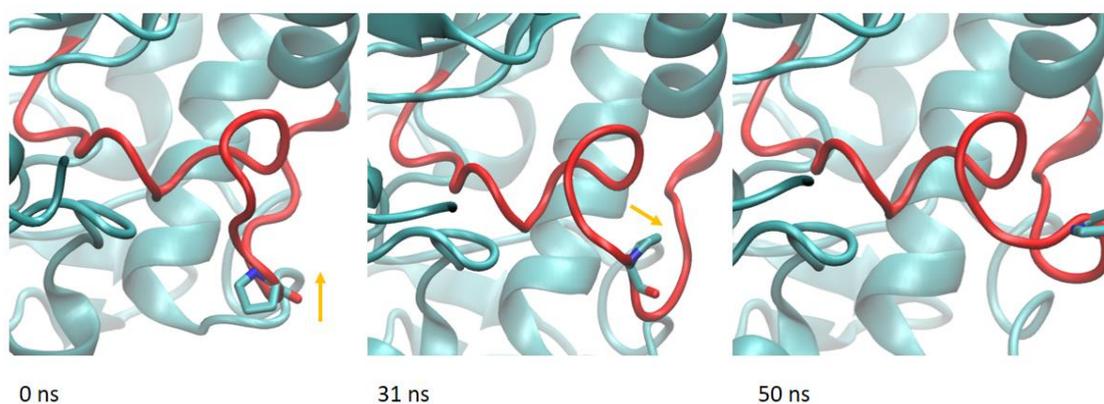


Figure 4.77. Pro42 on the dimerization1 loop (red) of apo W318C mutant form at 0 ns, 31 ns, and 50 ns.

In Figure 4.77 The yellow arrow indicates the movement of Pro42 which is shown by stick representation.

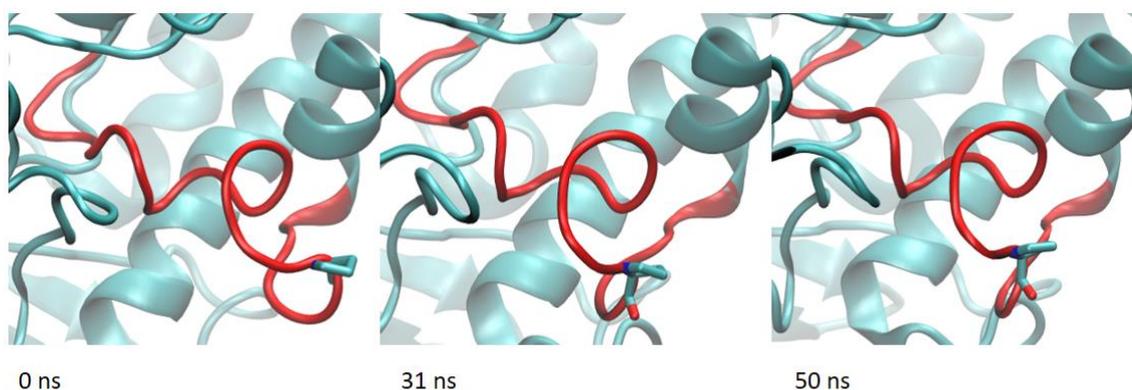


Figure 4.78. Pro42 (sticks) on the dimerization1 loop (red) of ligand bound W318C mutant form at 0 ns, 31 ns, and 50 ns.

From the figures, dimerization1 region fluctuates less in the simulation on ligand bound W318C mutant form. Dimerization1 region starts with the active site residue, Asp29, thus it is expected to see this rigid state of this region. The side chain of Pro42 looks towards the solvent in the simulation on ligand bound form because 5\_R ligand was docked to the structure which has the last coordinates of the simulation on apo W318C mutant form of

ARSA. After ligand docking the conformation of Pro42 as well as the dimerization1 region do not show major changes as in the simulation on apo W318C form.

The gap between the RMSD profiles of the apo and ligand bound WT forms can be linked to the very rigid standing of dimerization1 region in the simulation on ligand bound WT form. When we inspected the simulations on apo and ligand bound WT forms we see the differences occur on the dimerization1 region due to the movements of Thr45 as shown in Figure 4.79 and Figure 4.80.

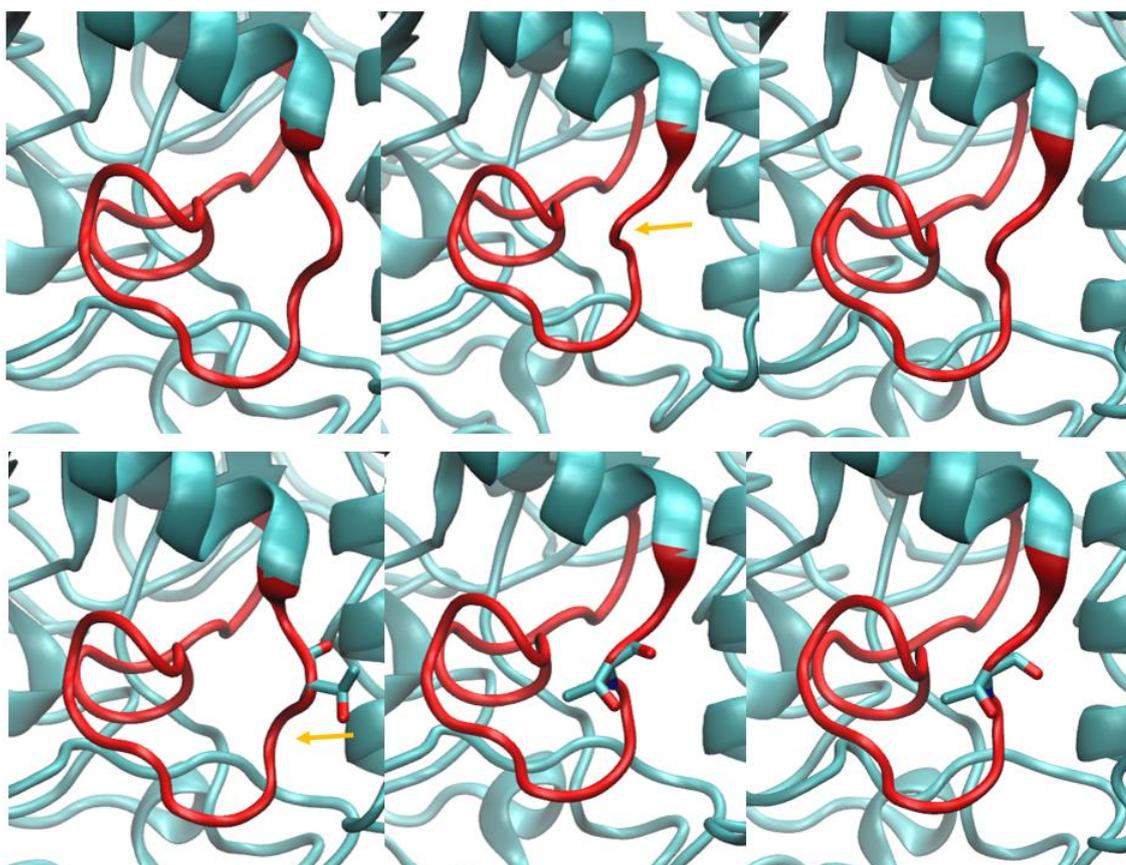


Figure 4.79. Thr45 at the 8th ns of the simulation on apo WT form of ARSA at 0ns, 8 ns and 50 ns.

In Figure 4.79, The yellow arrows indicate the structural change on the dimerization1 region (up), and the direction of the movement of Thr45 (down) which is shown with sticks representation. The dimerization1 region is the red part.

The docking was performed on the last frame structure of the simulation on apo WT form. The coordinates of the last frame were taken as the basis of the simulation on ligand bound structure thus, the conformation of dimerization1 region for the first frame of the simulation on ligand bound WT form is the same with the last frame of the simulation on apo WT form (last image in Figure 4.79). After ligand binding the conformation of this region almost never changes during the simulation as shown in Figure 4.80. The time steps were taken as the same in Figure 4.79.

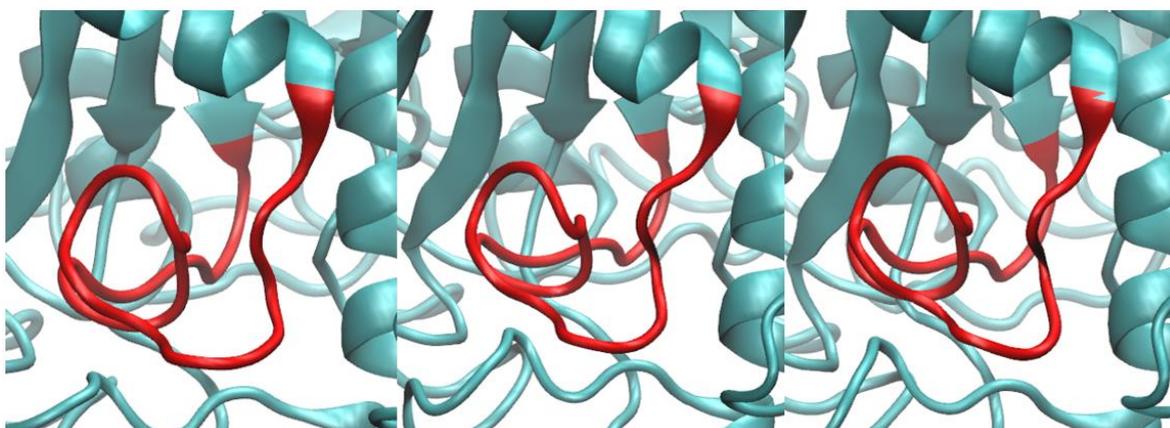


Figure 4.80. The conformations of the dimerization1 region (red) of ligand bound WT ARSA form at neutral pH at 0ns, 8ns and 50ns.

The movement of Thr45 may be necessary for dimerization since it happens in the simulation on WT ARSA form. On the other hand, it may be happening because the simulation was performed of the monomer structure but not dimer form. The movement of Thr45 should be analyzed in the simulation on WT dimer form of ARSA. This movement of Thr45 was also observed in the simulations on W318C and E307K+T391S mutants in neutral pH, WT-diol, W318C-diol, E307K+T391S-diol, E307K+T391S-nonhydrated, and E307K-nonhydrated forms. In the acidic pH, this movement was observed in simulations on three diol forms and two nonhydrated forms where E307K mutation is present. In the neutral pH, the ARSA forms exist in their nonhydrated states, and the movement of Thr45 was observed in all. Due to these results, we can conclude that this movement of Thr45 may not be significant for dimerization or be a result of flexible behavior of monomeric structure. Making rigid of the dimerization1 region is questionable and should be discussed further in the light of the results of simulations on dimer structures.

After analyzing the results of the RMSD calculations for the dimerization1 region, dimerization2 region was also examined and RMSD calculations were performed as shown in Figure 4.81.

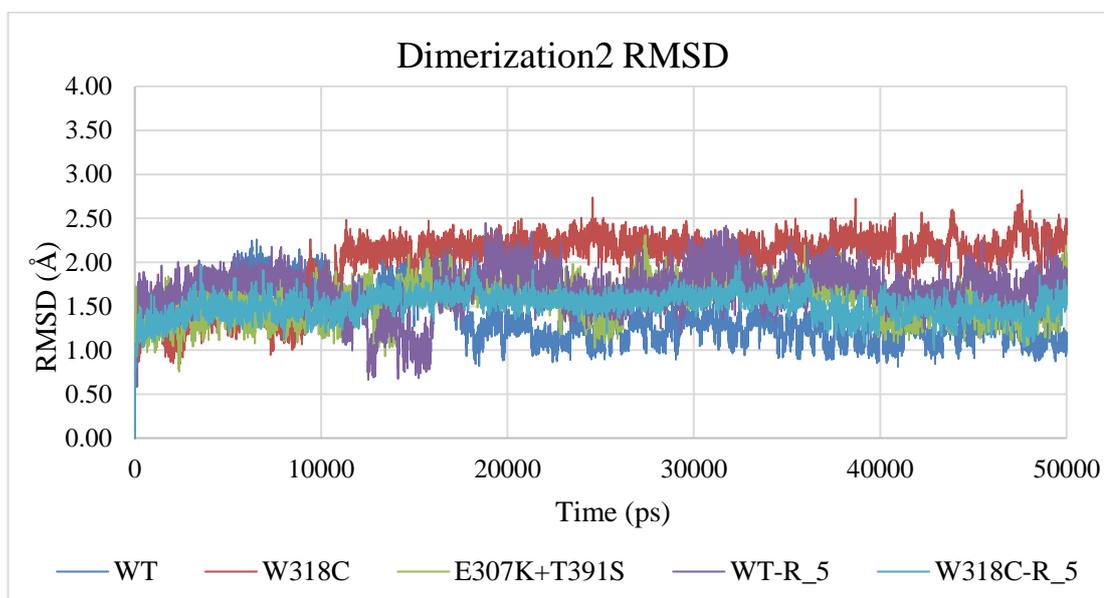


Figure 4.81. Dimerization2 region RMSD profiles of ARSA at neutral pH.

The highest trend among the RMSD profiles of the Dimerization2 regions belongs to the W318C mutant form therewithal the lowest trend after 20 ns belongs to the ligand bound W318C mutant form. There is an almost 1.0 Å difference between the trends reach their plateaus. It shows that ligand binding to the W318C mutant form makes its dimerization2 region also rigid as well as its dimerization1 region. On the other hand, the situation is on the contrary for WT forms. Ligand bound WT form shows a more fluctuating RMSD profile for the dimerization2 region first 20 ns of the simulation. Average RMSD values of the trends after 20 ns are not significantly different but the behavior of the trend of the ligand bound WT form is unexpected. The gap between the RMSD profiles of apo W318C mutant form and ligand bound W318C mutant form can be linked to the movement of Met289 on the dimerization2 loop as shown in Figure 4.82. Met289 turns its side chain upwards in the simulation on apo W318C mutant form as shown in the figure while it preserves its position during the simulation on ligand bound form of W318C mutant monomer. The position of Met289 in the ligand bound form stays as the horizontal position which it gained in the simulation on apo W318C mutant. The docking calculations were performed on the protein

coordinates that were taken from the last frame of the simulation on apo W318C monomer. The important thing is that the residue does not change its conformation during 50 ns and this stability may be a consequence of the ligand binding.

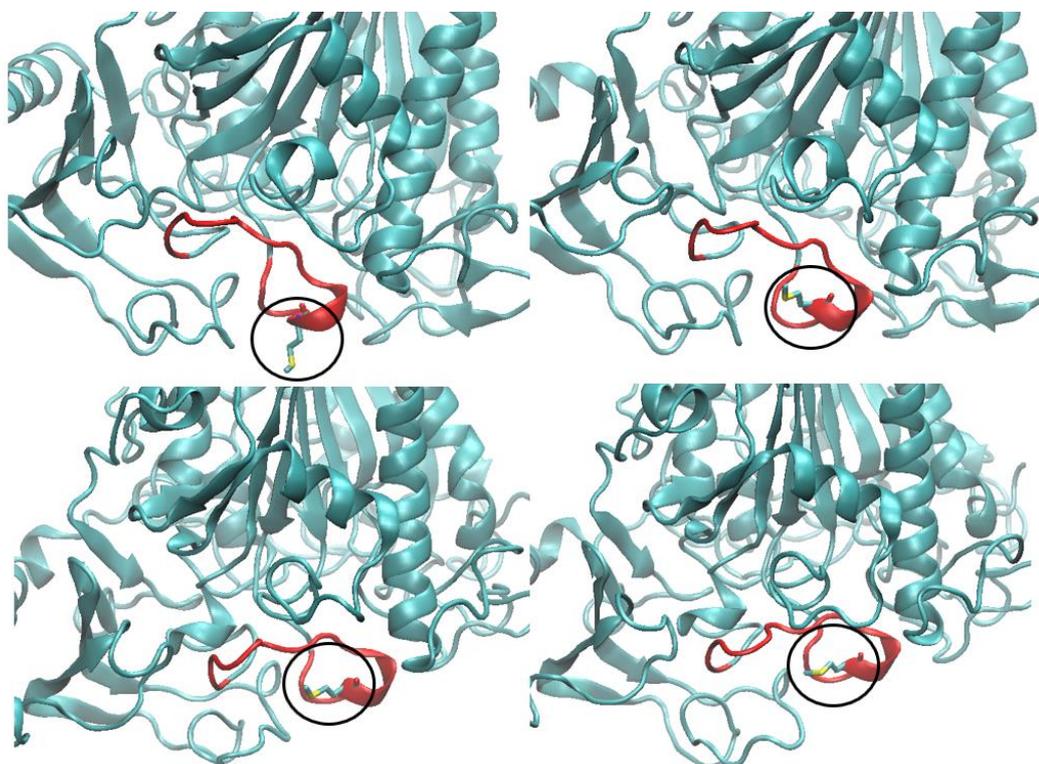


Figure 4.82. Met289 (shown with stick representation) in apo W318C monomer (top panel) and ligand bound W318C mutant form (bottom panel).

The snapshots in Figure 4.82 were taken from the first and the last frames of the simulations.

Methionine is a hydrophobic Sulphur containing residue. The conformational change of this residue at the dimer interface can cause a disruption in this region. This conformational change was also observed in the simulation on W318C-nonhydrated mutant form at acidic pH and W318C dimer form at neutral pH but not W318C dimer at acidic pH.

The distance between the Sulphur atom of Met289 and beta carbon atom of the Val445 in the other monomer in the crystal structure of the dimer is 3.38 Å. The movement of Met289 towards another hydrophobic residue (Val445) at the dimer interface can prevent

ARSA from dimerization. The movement of Met289 should also be observed in the dimer simulations but it should be kept in mind that the dimer simulations were performed on already dimerized mature ARSA forms and a repulsion due to this movement may not be observed. Even the movement of Met289 may not be observed in the dimer simulations due to reduced degree of freedom.

We also calculated the RMSD profiles of the dimerization3 regions for five forms of ARSA at neutral pH. The results are shown in Figure 4.83. The trends are similar to the ones for dimerization2 region.

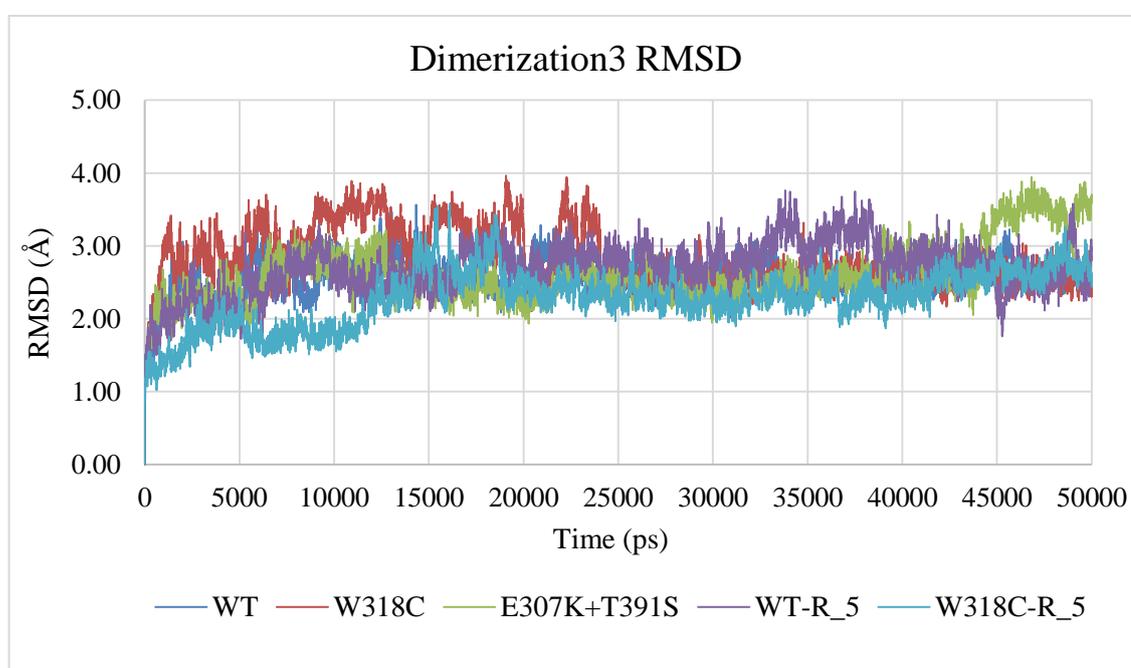


Figure 4.83. RMSD profiles of the dimerization3 region at neutral pH.

Ligand binding affects the way of reaching of dimerization3 region to equilibrium in W318C mutant form, but after 25 ns, the trends become similar for apo and ligand bound W318C mutant forms. Two forms of W318C monomers do not show significant deviations from their initial structures when compared to each other. The biggest differences after equilibrium belong to the trends of E307K+T391S mutant and ligand bound WT ARSA monomers. The jump in the trend of the profile of E307K+T391S mutant was linked to the movement of Val445 as explained previously. The ligand bound form of WT shows a more

fluctuating profile when compared to its apo form but these are not significant and we mostly focus on the differences between the mutant form, W318C, since we aim to target the mutant ARSA forms for PC candidates.

After analyzing the dynamics of the dimerization region, we also calculated the RMSD profile of the octamerization helix as shown in Figure 4.84.

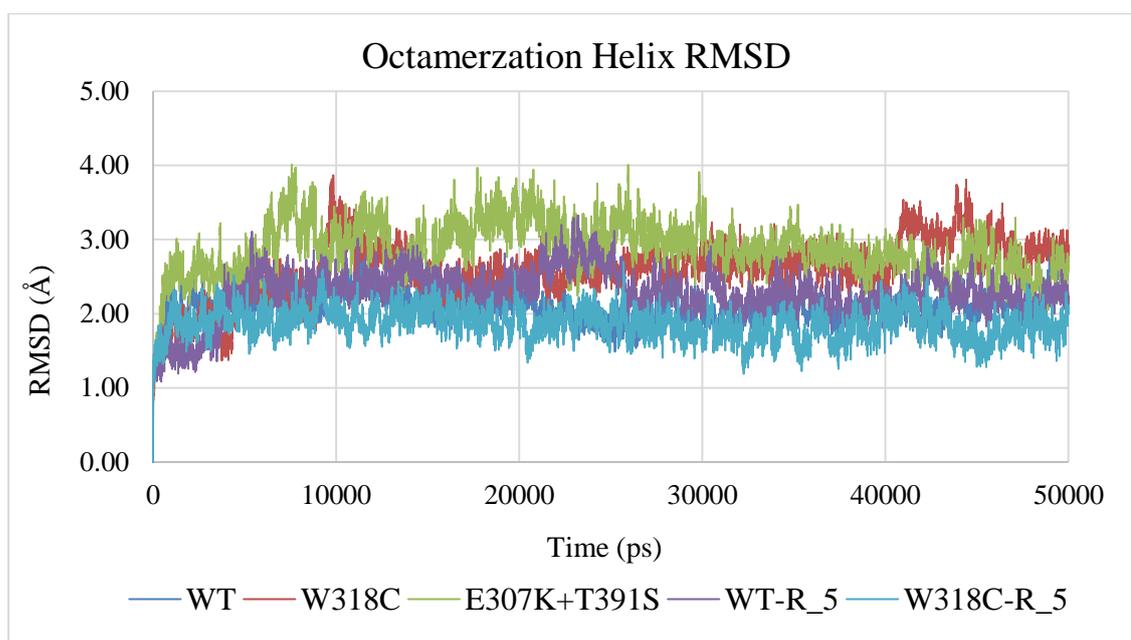


Figure 4.84. RMSD profiles of the octamerization helices at neutral pH.

The results show that ligand binding makes WT and W318C ARSA monomers deviate less from its initial coordinates when compared to the apo forms of mutant monomers. Especially after 40 ns, apo form of WT ARSA shows an increasing trend in Figure 4.84, although ligand bound W318C monomers preserves its flat trend almost for the whole simulation time after reaching equilibrium very quickly.

The RMSD results of the specific regions and RMSF calculations for the overall structures showed that ligand binding (R\_5 binding tightly in the active site) makes the dimerization regions and consequently the octamerization region less flexible. It may be a desired status for protein folding in the case of misfolding of ARSA due to mutation W318C.

4.7.2.3. Salt Bridges and Hydrogen Bond Occupancies. Possible salt bridges and hydrogen bond occupancies were calculated as previously described. The results of the ligand bound forms were compared to the apo forms of ARSA monomers at neutral pH. The results for all possible salt bridges are given in Table A.5 under the corresponding titles of the simulations. In addition, the possible salt bridges which are missing in any other simulation are also summarized in Table 4.38 with the average distances between the residues that were calculated for the whole trajectories and written.

Table 4.38. The possible salt bridges in the simulations on the corresponding forms of ARSA forms.

<b>Residue Pairs</b>	<b>WT</b>	<b>W318C</b>	<b>WT-R_5</b>	<b>W318C-R_5</b>
ARG19-ASP216	-	10.45 Å	10.97 Å	4.12 Å
ARG19-GLU272	-	-	12.23 Å	-
ARG19-GLU273	-	-	5.60 Å	7.81 Å
ARG97-GLU131	-	6.17 Å*	4.99 Å	6.50 Å
ASP152-ARG288	-	-	-	4.82 Å
ASP152-HIS229	-	-	-	4.64 Å
ASP169-ARG288	-	-	7.90 Å	-
ASP29-HIS229	-	-	-	4.23 Å
ASP29-LYS123	-	-	3.44 Å	-
ASP30-ARG288	-	-	-	6.62 Å
ASP30-HIS229	5.04 Å	4.70 Å	-	4.05 Å
ASP467-LYS463	5.80 Å	5.50 Å	-	-
ASP61-LYS393	-	-	10.83 Å	9.95 Å
GLU103-ARG479	-	-	-	9.10 Å
GLU131-ARG496	9.17 Å	-	-	-
GLU285-HIS405	4.64 Å	4.67 Å	8.26 Å**	-
GLU382-ARG384	-	-	9.37 Å	-
GLU424-LYS463	-	10.63 Å	3.42 Å	3.36 Å
LYS123-ASP152	-	-	5.82 Å	-

\*: After 11 ns, the distance becomes approximately 4.5 Å between the residues.

\*\* : The last 10 ns the distance becomes approximately 4.0 Å between the residues.

From the results, it is clear that the ligand binding makes the structure more rigid in terms of enhancing the salt bridge formation. When compared to WT simulation the number of possible salt bridges in the simulation on ligand bound forms may seem high but in the case of misfolding it may be the desired situation in terms of having more stable and rigid forms to enhance the folding of the enzyme.

4.7.2.4. State of NAG+NDG Sugar Moiety. When the conformational changes of sugar moiety were observed during the simulation it can be concluded that ligand binding do not affect the dynamics of the NAG+NDG sugar complex. Even on the contrary, it may be make the sugar moiety less mobile as the specific regions of ARSA. The first and last frames of the simulations on apo form of W318C monomer and R\_5 docked W318C mutant monomer were taken as snapshots for comparison as shown in Figure 4.85 and Figure 4.86, respectively.

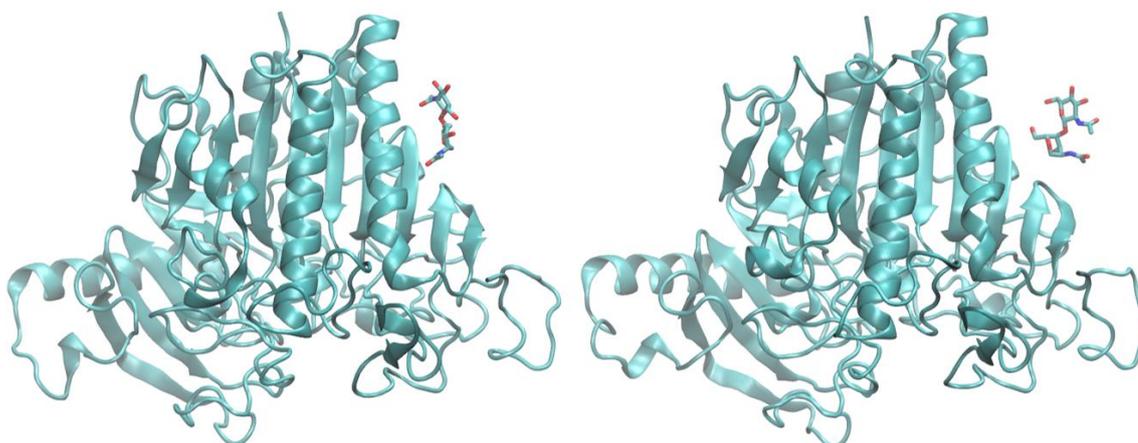


Figure 4.85. Snapshots from at 0 ns (left) and 50 ns (right) for apo form of W318C mutant monomer at neutral pH.

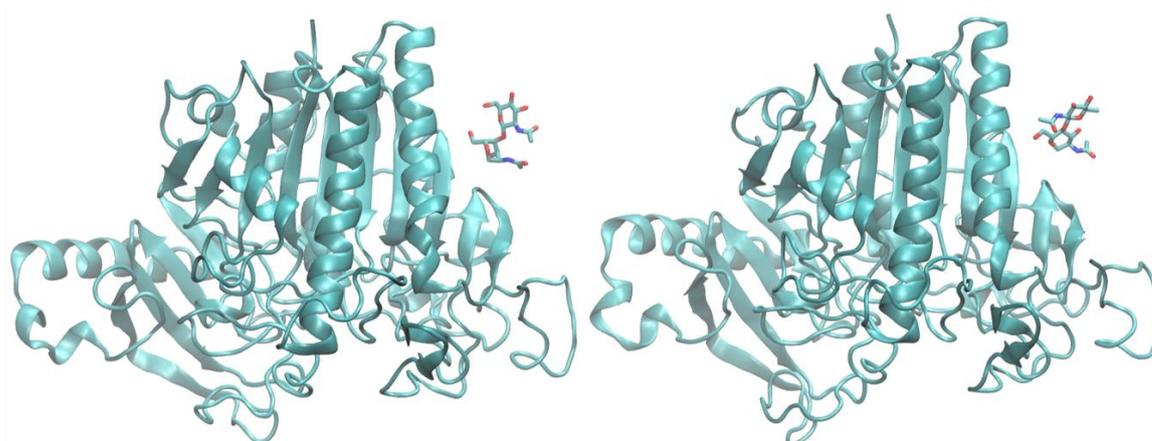


Figure 4.86. Snapshots at 0 ns (left) and 50 ns (right) for R\_5 bound form of W318C mutant monomer at neutral pH.

4.7.2.5. Solvent Accessible Surface Area (SASA) Calculations. SASA calculations were done by using a tcl script in VMD. We calculated the monomer forms of WT, W318C, E037K+T391S, and R\_5 bound forms of W31C mutant monomers. The results were plotted as in Figure 4.87.

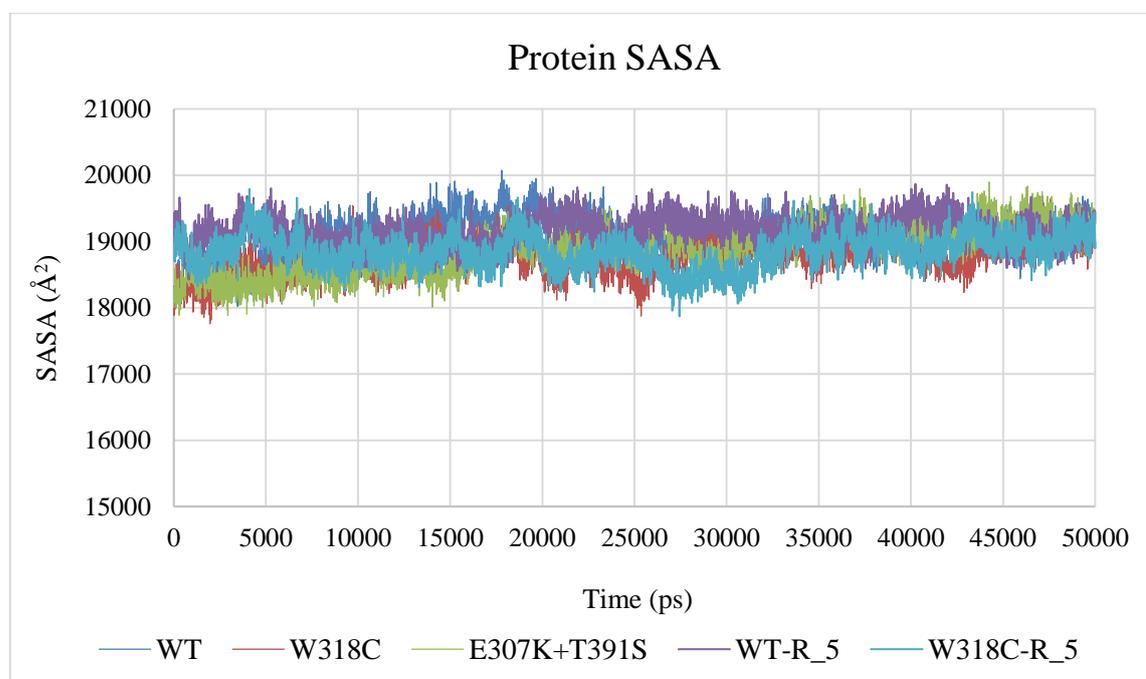


Figure 4.87. SASA profiles of the monomeric ARSA forms at neutral pH.

The SASA results of the monomers forms show that ligand binding to W318C mutant or the mutant's itself do not have significant effect on the overall protein dynamics.

4.7.2.6. Secondary Structure Analyses. The secondary structure analysis was performed in SID of Maestro GUI. The heat map of the secondary structure of R\_5 bound W318C monomer at neutral pH is given in Figure 4.88.

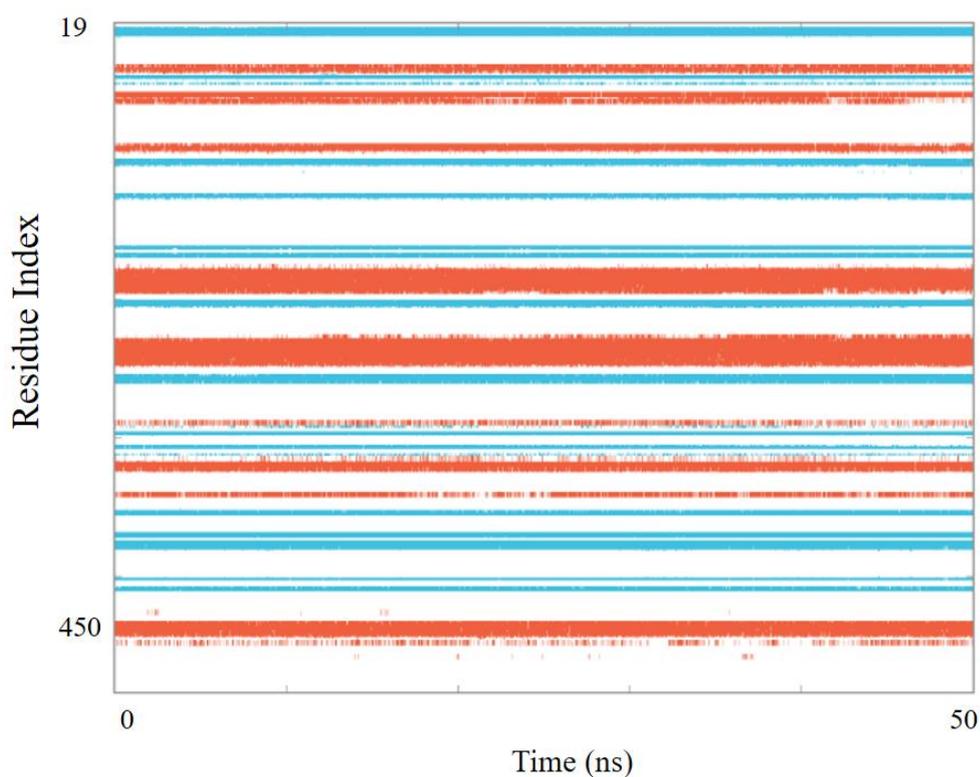


Figure 4.88. Heatmap of the secondary structure analysis for R\_5 bound W318C monomer.

The results of the secondary structure analysis show that R\_5 binding to W318C monomer enhances the monomer for preserving its secondary structure. For instance, even at the Val470 end of the octamerization helix does not show discontinuity as much as we see in the simulation on apo form of W318C and E307K+T391S mutants.

### 4.7.3. MD Simulation on ZINC90709065 compound Bound W318C Mutant Monomer

In section 4.4.2, we stated that the hypotheses that were found were not satisfactory in terms of their scores, nevertheless we decided to use the APD hypothesis which was generated by using the CNS library of OTAVA for finding matches in ZINC library (4 million compounds). After finding matches there came out 1000 molecules which satisfy the APD hypothesis. After that we again performed a VSW on W318C monomer at neutral pH by using these 1000 matches. After the VSW calculations, only one pose came out as a result with an XP GScore of -7.42 kcal/mol. Although the ligand does not fit on the APD hypothesis very accurately, the binding pose of the ligand was favorable since it entered into the active site quite well as shown in Figure 4.89. The interaction map of the ligand is also shown in Figure 4.90.

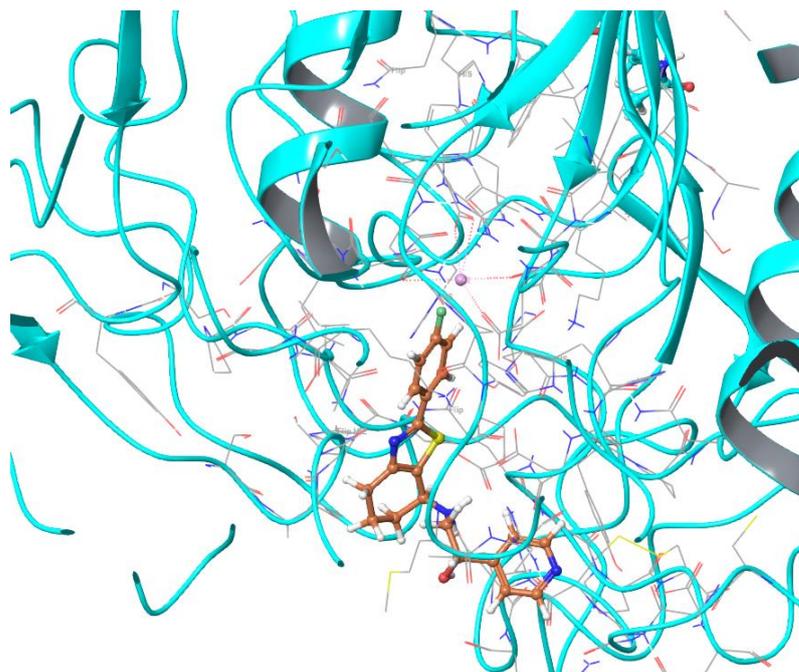


Figure 4.89. ZINC90709065 (orange) in the active site cavity of W318C monomer (blue) at neutral pH.

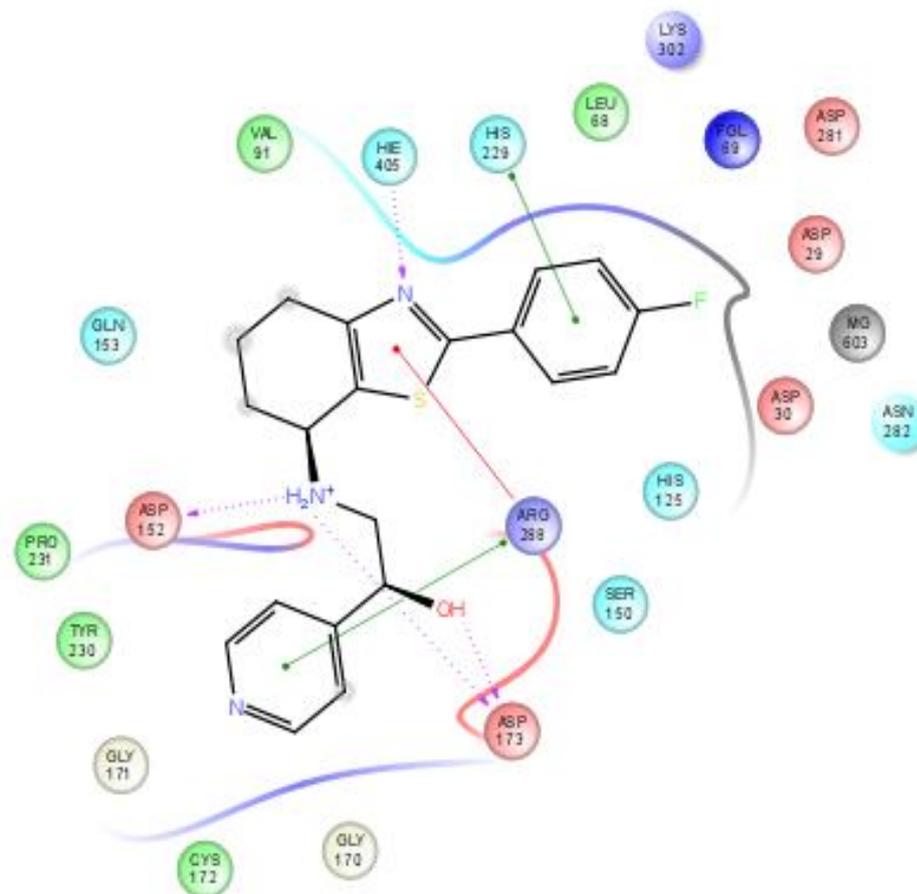


Figure 4.90. Interaction map between ZINC90709065 and W318C mutant monomer.

Figure 4.89 and Figure 4.90 show the first frame which is the system after docking. We see that the molecule entered into the active site quite well and interacts with the active site residues (Arg288, Ser150, His229, Asp29, Asp30, and His125), and the  $Mg^{2+}$  cation. Due to its binding free energy, good binding pose and the dataset that the original hypothesis comes from, we decided to perform an MD simulation on ZINC90709065+W318C monomer complex at neutral pH. The analyses of the simulation is discussed as follows.

**4.7.3.1. RMSD and RMSF Calculations.** The RMSD and RMSF calculations for the overall structure and the specific regions of this ligand bound ARSA form were performed as previously described. The results of the analyses were compared to those of WT, apo form of W318C, and R\_5 bound form of W318C monomer at neutral pH. Firstly, the backbone RMSD results are given in Figure 4.91.

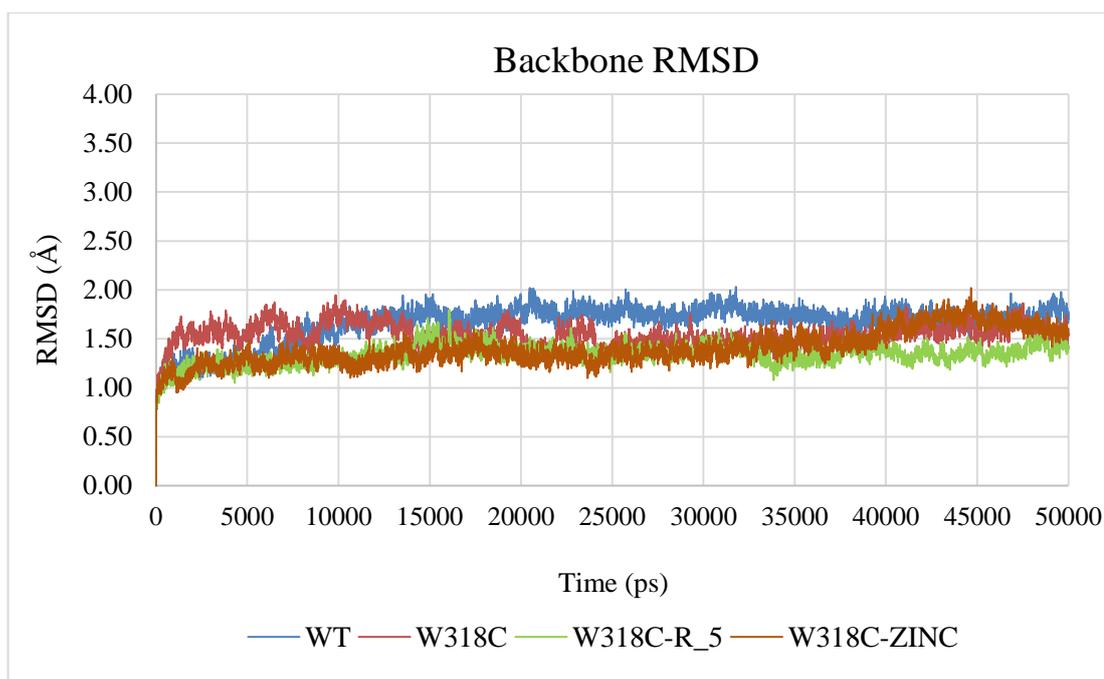


Figure 4.91. RMSD profiles of the backbones for WT and three forms of W318C mutant monomers at neutral pH.

The backbone RMSD profiles in Figure 4.91 do not show significant differences but it is clear that ZINC90709065 cannot provide the stability for W318C to preserve the rigid overall backbone dynamics as does the R\_5 compound, but since the most deviating profile belongs to WT monomer this difference cannot be considered as significant in terms of overall structural stability. We continued our analyses with the specific regions of ARSA as follows. First the active site RMSD profiles were calculated as shown in Figure 4.92.

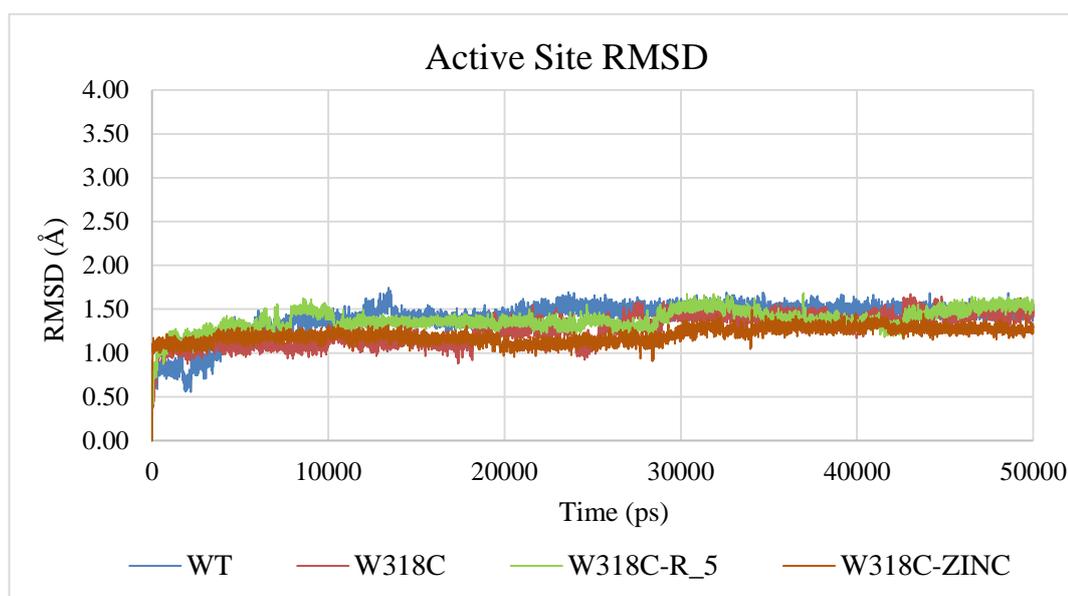


Figure 4.92. RMSD profiles of the active site residues for WT and three forms of W318C mutant monomers at neutral pH.

Three forms of W318c monomers have active site RMSD profiles that increase slightly before 30<sup>th</sup> ns of the simulations as in figure 4.80. Apo form of the mutant and ZINC molecule bound form have RMSD trends lower than WT between 10 ns and 30 ns but after a while (last 10 ns) all the profiles reach a plateau around 1.40 Å and they do not show significant differences. The slight increase around 30<sup>th</sup> ns in the profiles can be linked to the conformational changes of the ligands in the active site as shown in Figure 4.93.

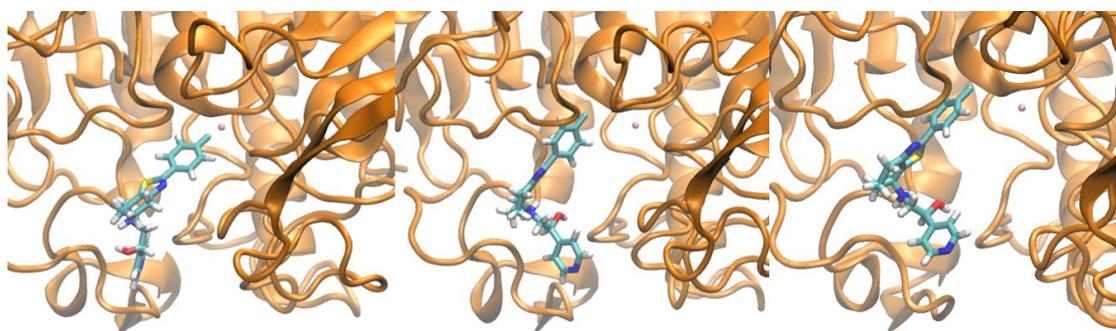


Figure 4.93. ZINC90709065 compound at 0 ns, 29 ns and 50 ns for the simulation on W318C monomer.

The compound, ZINC90709065, gains a bent conformation and shifts its position towards left as shown in Figure 4.93. These cause changes in the interactions between the ligands and the active site residues and consequently changes in the active site conformation although we found in this case insignificant. After examining the overall structure and active site dynamics we also performed RMSD calculations for the specific sites as follows. Dimerization1 site results are shown in Figure 4.94.

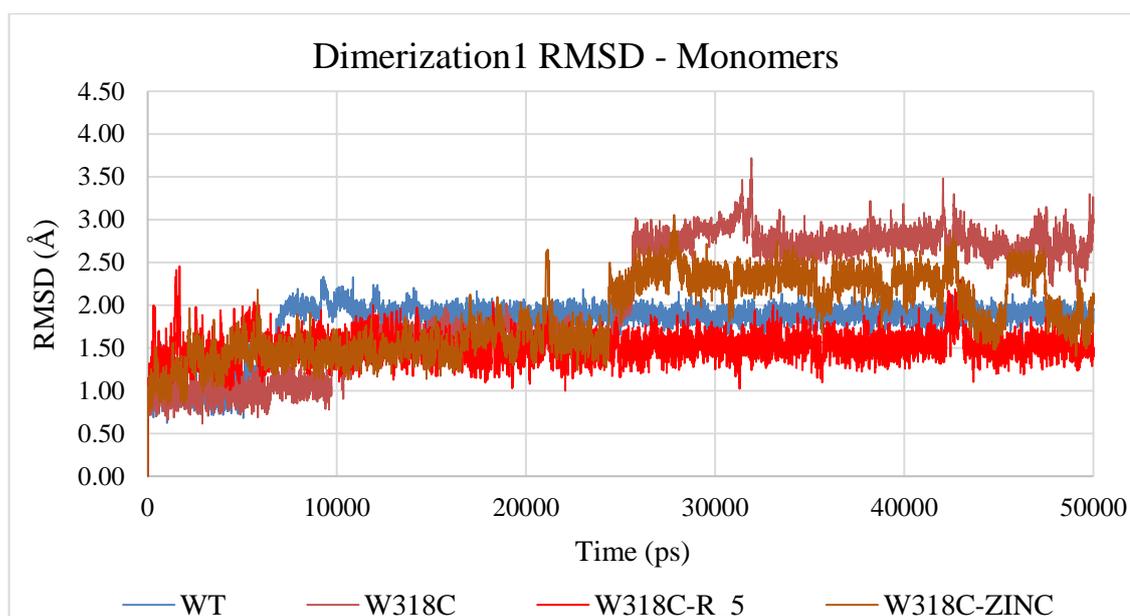


Figure 4.94. RMSD profiles of dimerization1 regions for WT and W318C mutant monomers at neutral pH.

The RMSD results of the dimerization1 regions of the monomer forms show that binding of ZINC90709065 cannot make this region preserve its stability in the case of binding of R\_5. After 25 ns, the RMSD profile of ZINC molecule bound W318C mutant form becomes fluctuating. This behavior can be explained by the motions and conformational changes of Thr45 as shown in Figure 4.95.

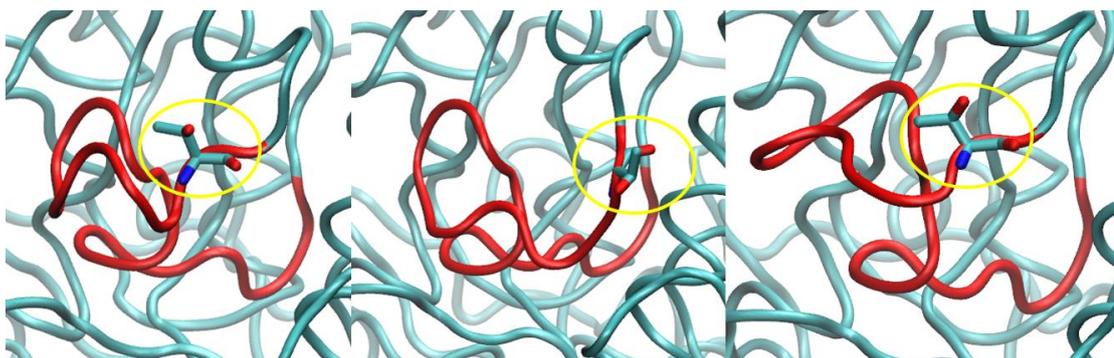


Figure 4.95. Thr45 (shown with sticks representation) and the dimerization1 region (red) at 0 ns, 25 ns and 50 ns of the simulation on ZINC90709065 bound W318C mutant monomer at neutral pH.

The conformational changes on this region and Thr45 on the mature crystal structure may be a clue for the dimer stability or the misfolding of mutant ARSA forms.

Secondly, the dimerization2 region RMSD profiles were calculated and plotted as in Figure 4.96.

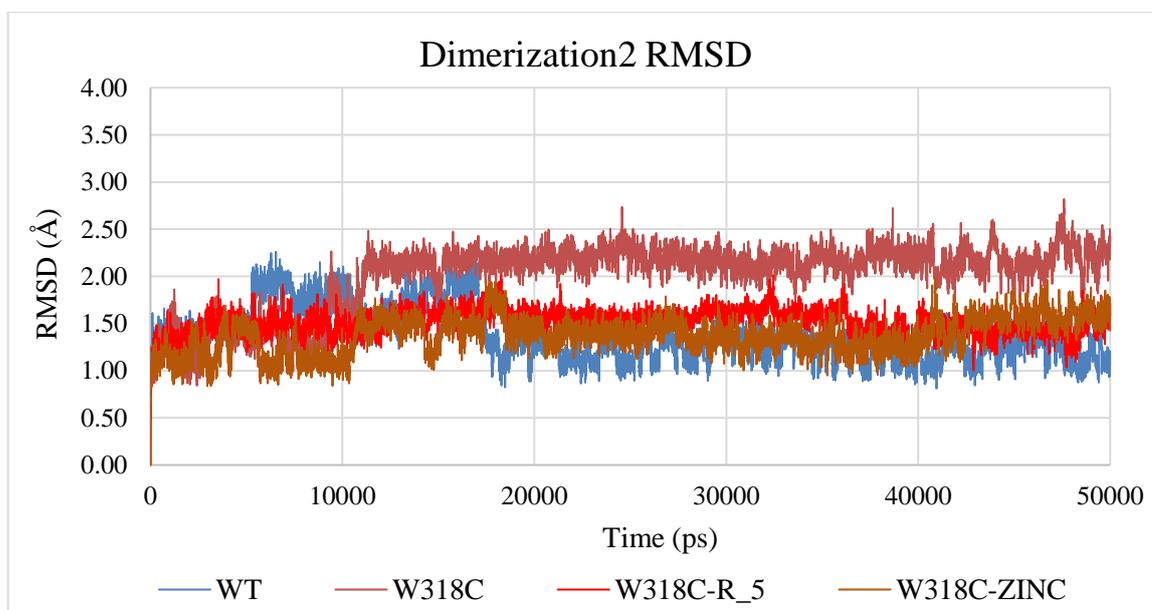


Figure 4.96. RMSD profiles of dimerization2 regions for WT and three forms of W318C mutant monomers at neutral pH.

The RMSD results for dimerization2 region show that binding of both R\_5 and ZINC90709065 molecules makes the mutant form's dimerization2 region rigid again. From the results, we concluded that mostly affected region from E307K+T391S and W318C mutants on ARSA is dimerization1 region and the deviations of the regions become less when we continue on from 1 to 3 of the dimerization regions. If the ligand binding makes these regions, especially the dimerization1 region, rigid again this may be an indicator of a potential PC for W318C. When we compare R\_5 and ZINC90659070, we see that R\_5 binds more tightly and makes dimerization1 region more rigid but as we continue our analyses we see that ZINC90659070 can also be a good candidate for PC improvement for ARSA. We continue our analyses on specific regions with the dimerization3 region as shown in Figure 4.97.

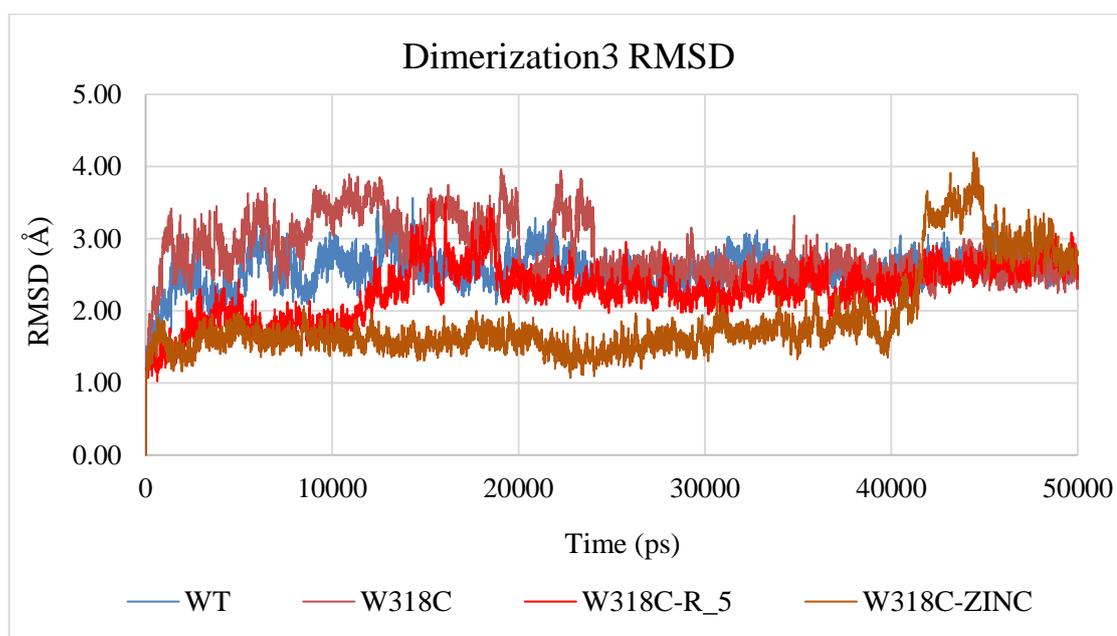


Figure 4.97. Dimerization3 RMSD profiles of WT and three forms of W318C mutant monomers at neutral pH

The profile of the ZINC90709065 bound W318C mutant form shows until the last 10 ns of the simulation this region deviates the least among the other forms. Since the WT form has a very flexible end loop in this region (the missing region in the crystal structure) this rigidity may not be desired. After 40 ns the region fluctuates significantly, especially the residue Val445 changes its conformation, and this is reflected to the RMSD trend of the

ZINC bound W318C form. Due to the rigid behavior in the first 40 ns we can say that this ZINC molecule may be helpful during folding if the inactivity mechanism of W318C is misfolding. In addition, at the end of the simulation we see that the monomers acquire similarly deviated conformations for dimerization3 region. Since this region is very flexible by nature [1] it may not be a determinant for the severity of the mutation or its inactivity mechanism. Lastly, we compared the RMSD results of the octamerization helix as shown in Figure 4.98.

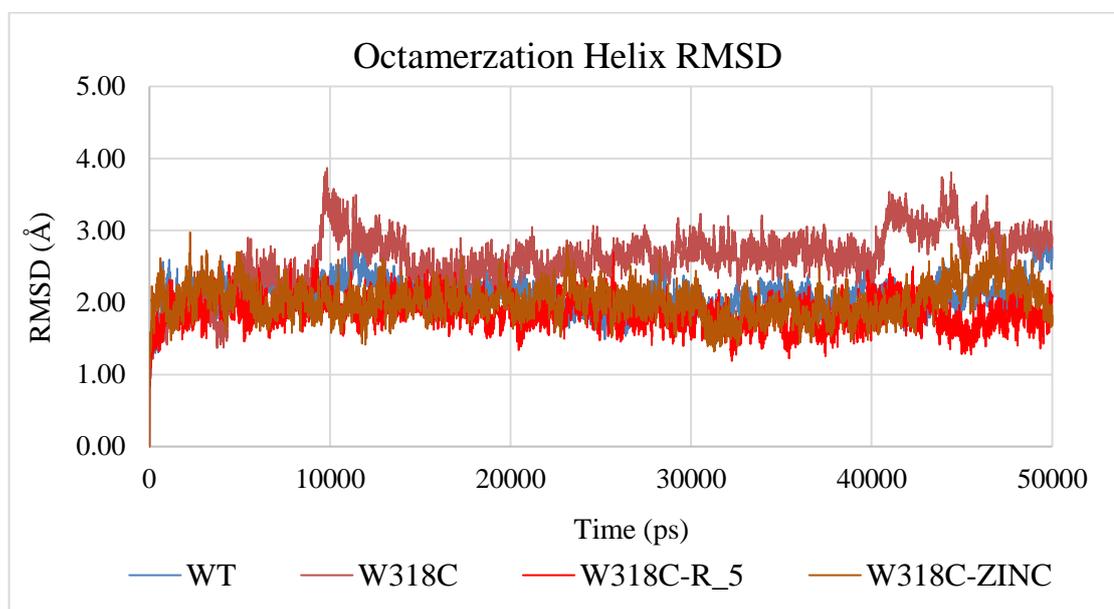


Figure 4.98. Octamerization helix RMSD profiles of WT and three forms of W318C mutant monomers at neutral pH.

The deviations from the first frame coordinates of the octamerization helices are not significantly different according to the results in Figure 4.98, but it is clear that ligand binding makes octamerization helix of W318C more rigid again. We learned that the conformational changes at the monomer-monomer interface may affect the octamerization indirectly [61,62], thus we can say that the stabilized behavior of the octamerization helix can be linked to the more rigid dynamics of the dimerization1 and dimerization2 regions.

Lastly, when we calculated the RMSD profile of the region, Asp335-Arg370, we did not see any significant differences or deviations. The dynamics of this region, in fact, is not affected even by the misfolding causing mutations T274M and P377L.

**4.7.3.2. Possible Salt Bridges and Hydrogen Bond Occupancies.** The possible salt bridges were calculated as previously described. This time we calculated the distances between oxygen-nitrogen pairs of the basic and acidic residues for the last 20 ns of the simulations on WT, apo W318C and ligand bound forms of W318C mutant monomers at neutral pH that were listed in Table 4.39.

Table 4.39. Distances between the oxygen-nitrogen pairs of the basic and acidic residues for the last 20 ns.

<b>Residue Pairs</b>	<b>WT</b>	<b>W318C</b>	<b>WT-R_5</b>	<b>W318C-R_5</b>	<b>W318C-ZINC</b>
ARG19-ASP216	4.66 Å	10.60 Å	10.77 Å	3.66 Å	8.78 Å
ARG19-GLU272	16.88 Å	12.80 Å	11.30 Å	14.68 Å	11.14 Å
ARG19-GLU273	9.53 Å	8.92 Å	4.96 Å	9.25 Å	4.64 Å
ARG97-GLU131	3.80 Å	6.72 Å	4.86 Å	7.76 Å	4.80 Å
ASP152-ARG288	5.84 Å	6.86 Å	10.54 Å	5.49 Å	6.63 Å
ASP152-HIS229	5.03 Å	4.38 Å	4.48 Å	4.64 Å	4.50 Å
ASP169-ARG288	8.94 Å	7.96 Å	7.02 Å	8.60 Å	8.88 Å
ASP29-HIS229	4.69 Å	5.49 Å	2.62 Å	4.60 Å	4.70 Å
ASP29-LYS123	4.52 Å	3.75 Å	3.41 Å	2.94 Å	3.41 Å
ASP30-ARG288	7.51 Å	8.38 Å	10.34 Å	6.95 Å	5.85 Å
ASP30-HIS229	4.66 Å	4.32 Å	3.22 Å	4.15 Å	4.81 Å
ASP467-LYS463	8.43 Å	13.20 Å	12.76 Å	13.04 Å	13.80 Å
ASP61-LYS393	12.46 Å	7.90 Å	13.21 Å	7.86 Å	10.11 Å
GLU103-ARG479	9.20 Å	8.16 Å	8.56 Å	9.50 Å	9.30 Å
GLU131-ARG496	11.95 Å	12.17 Å	10.46 Å	14.92 Å	13.34 Å
GLU285-HIS405	3.45 Å	3.77 Å	6.16 Å	7.20 Å	7.46 Å
GLU382-ARG384	9.41 Å	9.20 Å	8.66 Å	7.37 Å	7.58 Å
GLU424-LYS463	5.36 Å	4.83 Å	4.10 Å	3.05 Å	3.07 Å
LYS123-ASP152	5.46 Å	7.01 Å	5.80 Å	6.54 Å	5.38 Å
ARG241-ASP255	11.23 Å	11.54 Å	10.39 Å	10.70 Å	10.83 Å

The most significant difference between the distances of the residue pairs of the forms is the distance between Arg19 and Asp216. In the simulations on WT and R\_5 bound W318C mutant monomer forms, the distances between these two residues are similar. They preserve their positions close to each other while in the simulation on the other forms they get away from each other. It is interesting to see this separation between Arg19 and Asp216 in the simulation on ligand bound WT monomer but, apart from that, we can say that binding of R\_5 to W318C mutant makes the mutant structure more stable. In addition, these two residues are located at the top of the enzyme and they probably play critical roles in the sequential protein folding of ARSA [4,7]. When we look at the distance between Arg19 and Asp216 in the simulation on ZINC molecule bound W318C we do not see they get closer. If this is critical for ARSA stability, we can say that ZINC90709065 should be improved or cannot be accepted to be a good PC candidate. Other significant differences are the distances between the pairs Arg19-Glu273 and Arg97-Glu131 among the ARSA forms. R\_5 binding to WT ARSA monomer and ZINC90709065 binding to W318C make the distances between Arg19 and Glu273 shorter throughout the production runs. When compared to the other forms in Table 4.40, this might not be the desired situation for protein folding. When we look at the distances between Arg97 and Glu131 pairs, we see that ligand binding makes the length close to the WT's, but these two residues become further when R\_5 is bound to W318C mutant monomer. This makes R\_5 binding or the importance of the salt bridge between Arg97 and Glu131 questionable.

#### **4.8. MD Simulations on Dimer ARSA Forms at Neutral pH**

ARSA forms its dimer form in neutral pH after the monomers are synthesized and fold correctly. This is the first and the crucial step of the oligomerization of ARSA. We analyzed the trajectories of the simulations on dimer forms as previously done for monomer ARSA forms. At the first glance, the biggest difference is that the conformation and stability of the octamerization helix becomes affected by the mutation C300F at pH=5. It is the mutation that prevents ARSA dimers from octamerization. Although it is known that the intramolecular cysteine knot between Cys300 and Cys414 is crucial for dimer stability, the mutation C300F was observed to hinder the octamerization of ARSA due to the change in the association angle change between the monomers [61,62].

After analyzing the trajectories, we can say that there are two main differences among the dimer forms at neutral pH namely, the secondary structure of one of the octamerization helices and the angle between the monomers. Monomers of WT become closer after 50 ns but this tilting movement is different for the mutant dimers. The chains of E307K+T391S dimer seem to lose their contact more while the monomers of W318C mutant becomes closer than those of WT dimer. For a better understanding, the last frame coordinates of the three forms were aligned and these changes are shown in Figure 4.99. The RMSD result after alignment was found as 2.20 Å. The last and the first frames of the simulations on the dimers of W318C and E307K+T391S mutants were overlapped and shown in Figure 4.100 to see the structural changes clearly.

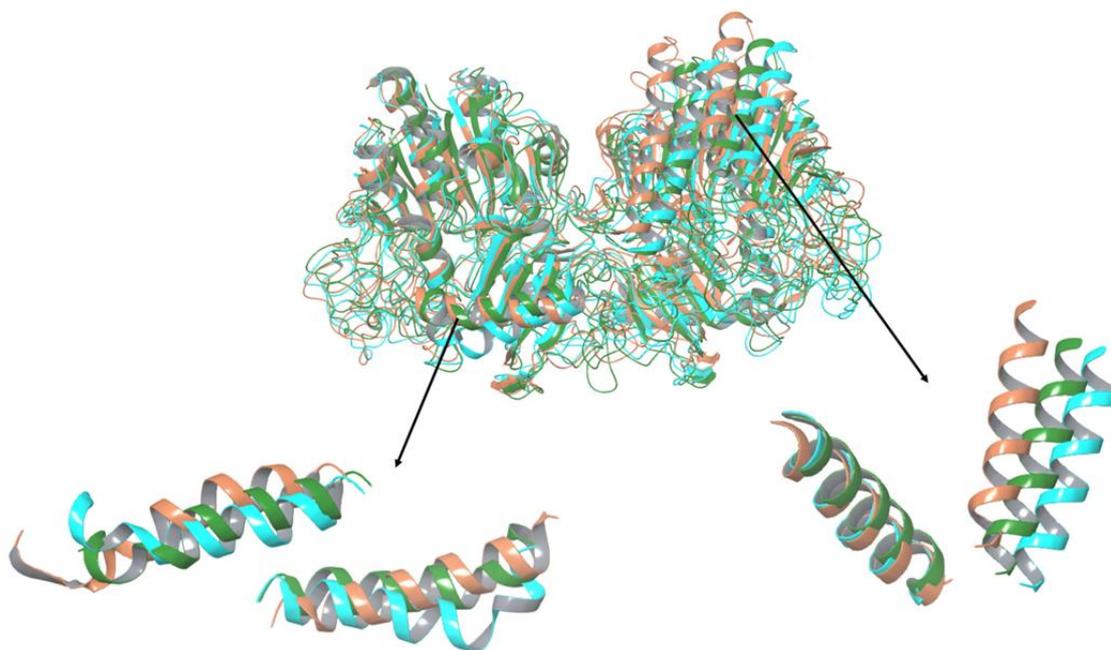


Figure 4.99. Last frame coordinates of E307K+T391S mutant (blue), WT (green) and W318C mutant (orange) dimers.

Three forms of ARSA are shown at neutral pH as aligned in Figure 4.99. To show the main differences, secondary structure changes on octamerization helices of both chains (left) and the different angles between the monomers (right), two helices of both chains are shown as indicated with arrows.

In Figure 4.99, two differences can be seen. One is the secondary structural change on the Val470 and of the octamerization helix. This change occurs in both chains for E307K+T391S mutant whereas only one chain of W318C mutant dimer shows a significant structural change on the Val470 end of the octamerization helix. The other difference is the tilted position of the chain on the right. During the simulations, monomers change their angles. These changes are also observed in the WT form such that one of the chains moves closer to the other one as in the simulation on W318C mutant form, but in the simulation on E307K+T391S mutant, the chain on the right moves far away from the chain on the left in contrast to the other two forms of ARSA.

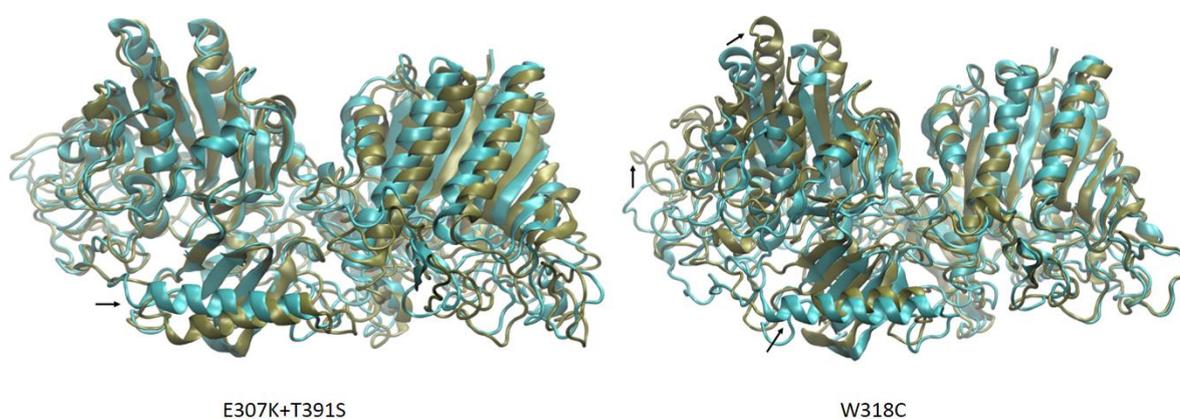


Figure 4.100. Simulation snapshots at 0 ns (blue) and 50 ns (brown) for E307K+T391S and W318C mutant dimers.

The arrows in Figure 4.100 indicate the biggest differences between the two frames from the simulation regarding the secondary structure of the protein.

The secondary structure change on the octamerization helix of W318C mutant is more significant. The angle change between the monomers is present in the simulation on W318C mutant as well at neutral pH. The structural change on the C-terminus loop part, and the changed position of the longest alpha helices are shown with arrows in Figure 4.100. In addition to the structural changes observed at the first view we also calculated the RMSD and RMSF values of the specific regions of ARSA and the all residues for dimer forms as well.

Finally, average structure of each form was calculated and generated as previously described. These average structures of dimers were aligned based on the protein backbone, and the RMSD values were calculated for the mutant forms, based on the coordinated of the average structure of WT dimer. The results are summarized in Table 4.40.

Table 4.40. RMSD values between the simulation average structures of ARSA dimers at neutral pH.

Dimer Form	RMSD (Å)
WT	0.00
W318C mutant	2.02
E307K+T391S mutant	2.56

As seen from Table 4.40, the overall protein structures do not show significant differences after reaching equilibrium, but the dimerization regions of the mutant monomer forms show structural changes. Due to these results, we calculated the dimer RMSD values based on their average structures in the same manner by hypothesizing that the mutations may be affecting the dimer structure. The RMSD value of WT dimer is zero because it was taken as the reference structure for RMSD calculations. When the same procedure was applied, we saw that mutant dimer structures show higher RMSD results as expected since the size of the protein becomes much larger as shown in Table 4.40. To support our results, we also performed RMSD calculations for the overall structure under the text title.

The aligned average structures of the dimers show the most significant change on octamerization helix of one of the monomers in the dimer structures. Val470 ends of the octamerization helices of the mutant dimers tend to unwind more than WT as shown in Figure 4.101, and the association angles are different than the WT structure. We aligned the structures also based on both octamerization helices, each time we see that the association angle between the two chains in the mutant dimers are different than the WT structure.

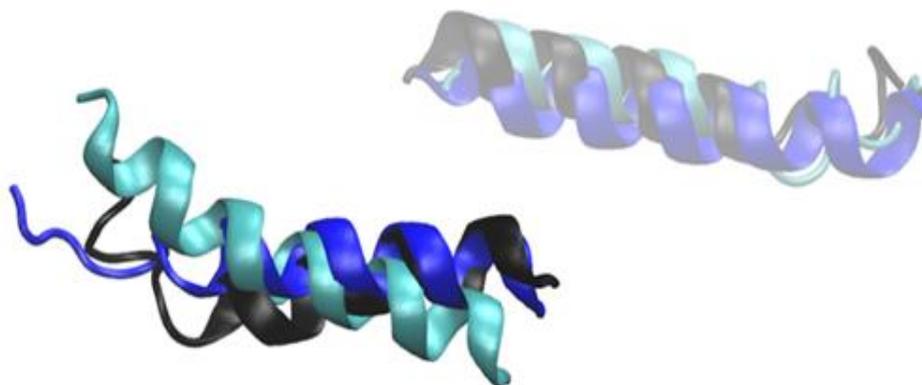


Figure 4.101. Two octamerization helices of the dimers when the average dimer structures after alignment on the backbone structure.

The color codes in Figure 4.101 are as follows: light blue = WT, dark blue = W318C mutant, black = E307K+T391S mutant. One of the octamerization helices in the average dimer structures of mutant ARSA forms cannot preserve its secondary helical structure, as shown in Figure 4.101. The conformational changes on the Val470 ends of the helices of the mutants become significant when compared to WT structure. It means that after reaching the equilibrium, the mutant forms still cannot acquire a helical form on the Val470 end of the octamerization helix. This can be due to the indirect effect of the more mobile dimerization regions in the monomer forms.

#### 4.8.1. RMSD and RMSF Calculations – Overall Structure

First, the whole backbones of the dimer forms were considered to calculate their RMSD trends throughout the trajectories. The results of the dimer backbone RMSD profiles are given in Figure 4.102.

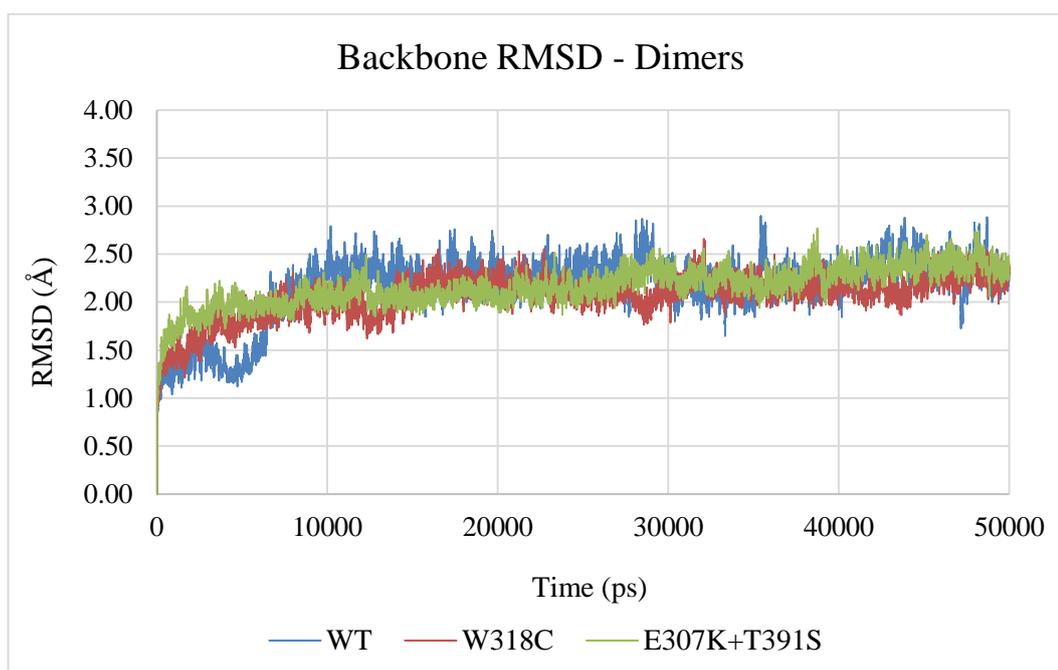


Figure 4.102. Dimer backbone RMSD profiles of three forms of ARSA dimers.

The results in Figure 4.102 show that general backbone structures of the dimers reach equilibrium after 20 ns around 2.2 Å meaning that the general dimer integrity is not affected by the mutations at neutral pH. Since ARSA dimers are big in size we did not expect to see a direct effect of the mutations on the dimer stability even as the monomers are not affected.

Secondly, we selected the chains separately for their backbone RMSD calculations. The backbone RMSD calculations (Figure 4.103) showed that chain RMSD profiles of the WT form is less than the mutant forms. The main outcome is that one of the chains is always more flexible and prone to change its secondary structure except for E307K+T391S dimer. The general trends of the profiles show that WT chains are more stable than the mutant ones. The RMSD differences between the chain1 of WT and the other chains of all forms are significant, such that the average RMSD difference between the two chains of WT form is 0.65 Å and between the chain1 of WT and the others around 1.00 Å.

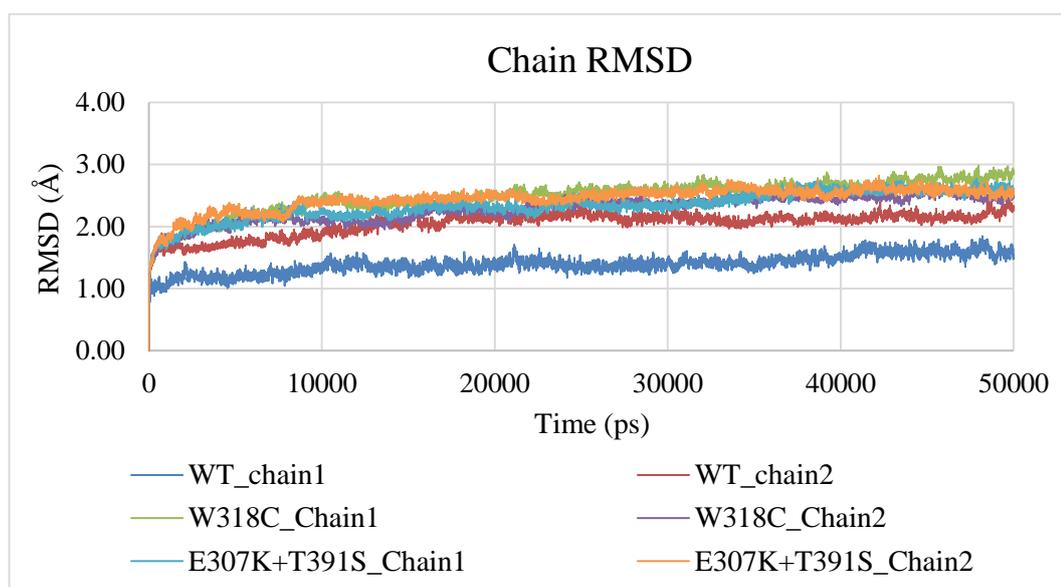


Figure 4.103. RMSD profiles of the chains of the dimer forms of ARSA at neutral pH.

In the simulations on dimer forms, we generally see that the RMSD profiles of some regions (dimerization1, octamerization and active site) show apart trends for each chain. This situation is most significant for the backbone RMSD profiles of the chains in the dimer forms at neutral pH, but we see that the lowest trend among the RMSD profiles of the chains belong to WT ARSA dimer. Both chains gain flat trends at lower RMSD values than the mutant dimer forms. The active site RMSD profiles, however, showed a different trend that WT chains have higher RMSD profiles for each than the those of E307K+T391S mutant form. The active site of ARSA forms the bottom part of the core structure of the enzyme and the presence of  $Mg^{2+}$  cation makes the active site architecture very stable by coordination. The increase in the profile of one of the chains of W318C is more significant than that of E307K+T391S mutant chain. That is why the stable profiles of the active site are expected (Figure 4.104), thus the increasing RMSD profile of the one of the active sites of W318C dimer form implies an instability of the protein due to mutation.

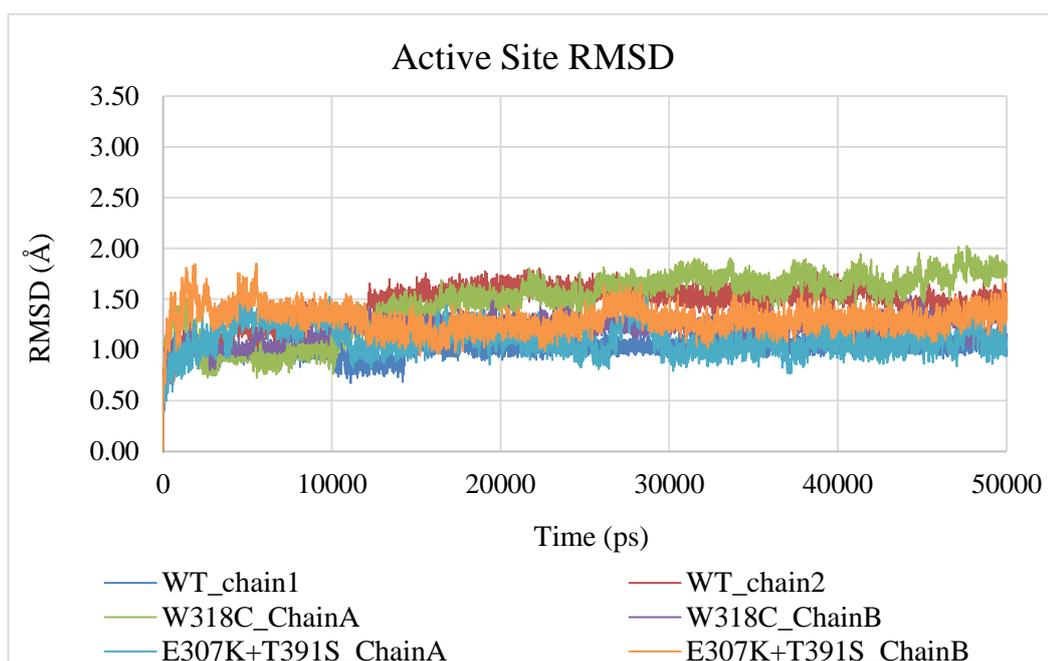


Figure 4.104. The RMSD profiles of the chains of the dimer forms of ARSA at neutral pH

We calculated the RMSF values of the alpha carbons of all residues and of all atoms of all residues for each chain in the simulations. The RMSF calculations showed that one of the chains of W318C mutant displays extreme points at some regions as in Figure 4.105.

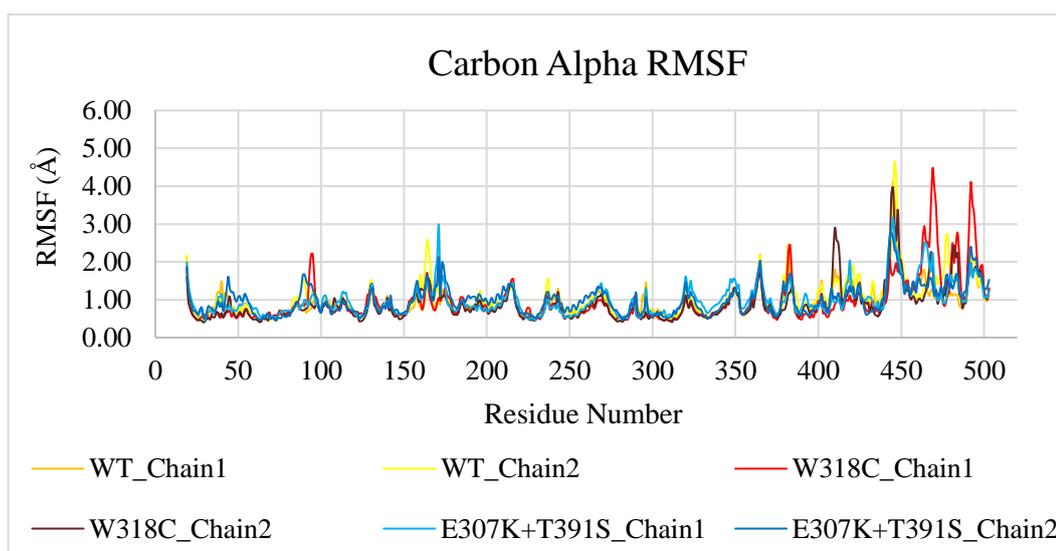


Figure 4.105. Residue based RMSF values of all atoms of the chains in dimers at pH=7.

The later built missing part of the crystal structure shows the highest RMSF values for all chains. The significant differences for the mutant chains occur around Gly170 and Gly171 of the chain1 of E307K+T391S mutant, Pro163-Ala165 of the chain2 of WT dimer, and C-terminus of chain1 of W318C mutant form. These regions are all far away from the specific regions and are mainly loop regions. The interesting thing is that these fluctuations or highly mobile profiles of the residues occur in only one of the chains of the dimers. In addition, Thr409-Ala413 region of chain2 of W318C mutant dimer has a significantly higher RMSF results than the other chains. This region may be playing a crucial role for dimerization of ARSA since it is located at the bottom part of the dimer interface, but during the simulation dimer form of W318C mutant do not become disrupted. This brings the other outcome into question: misfolding. Since the simulations were run on mature protein structures and ARSA protein is rather large, it is hard to disrupt the whole protein after mutating a residue on it, thus we cannot observe the effects of the mutations on the folding of ARSA. The mobility of the region, Thr409-Ala413 in the simulation on W318C dimer may be a sign of misfolding due to W318C mutation.

When we analyzed also the RMSF results (Figure 4.106) of all atoms in the dimer ARSA forms, we see that chain2 of WT and chain1 of W318C mutant dimers show similar and higher RMSF profiles. This shows that side chains of the residues in these two chains of WT and W318C mutant are very mobile.

The RMSF profiles of all atoms of the chains follow the similar trends of their carbon alpha results, but the WT and W318C profiles are interestingly different for the region between residues Thr60-Phe220. This region is shown in Figure 4.107 on the dimer form of ARSA. Loosening of this region during the simulations on ARSA forms is common. The high RMSF values of the residues can be linked to this behavior of the region. Residues act like they repel each other and the loop part in Figure 4.107 loosens. This movement on one chain in its dimer form may be necessary for ARSA to adjust its association angle before forming dimers. The dimer simulations in acidic pH should be observed for any conclusion but this difference between RMSF values maybe an indicator of a lack of octamerization in the case of E307K+T391S mutant.

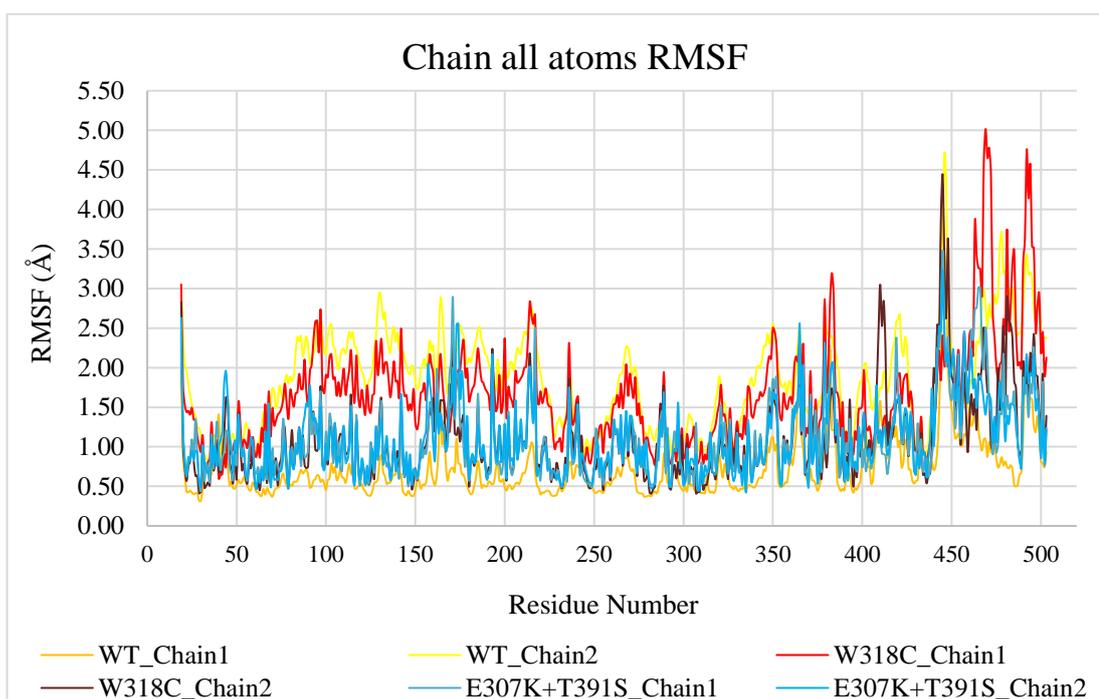


Figure 4.106. Residue based RMSF values of all dimer forms of ARSA.

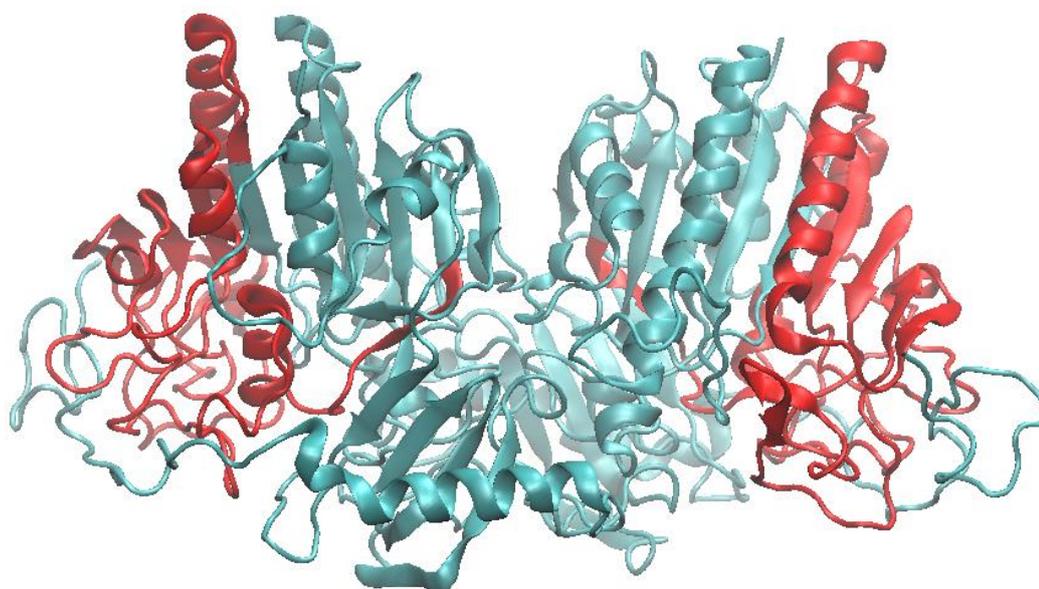


Figure 4.107. The red region (Thr60-Phe220) shows high RMSF values for chain2 of WT and chain1 of W318C mutant dimers.

The differences between the movements of the residues and the structures are clear when compared the RMSD and RMSF results. From these results, we suggest that mutations cause an instability on ARSA dimers and one chain of the dimer can act more flexibly than the other. When we compare all simulations on WT and various mutant dimers and monomers at different pH values, it seems that W318C probably causes misfolding, and E307K+T391S causes a hindrance in the octamerization, but these outcomes are hard to grasp from the simulations on mature structure of the enzyme, thus further investigation should be made.

#### **4.8.2. RMSD Calculations – Key Structural Elements**

RMSD calculations were done for the specific regions of each chain of the dimers. According to the results, each chain in the dimers shows higher RMSD profiles than the other for octamerization helix and two of the dimerization regions.

Dimerization1 region is one of the most mobile regions on ARSA monomers, but when the simulations were run on dimer forms, this region becomes more stable when compared to monomers at neutral pH due to dimer contacts. RMSD profiles of dimerization1 region of the chains reach the highest point at 2.5 Å whereas it becomes 3.5 Å in the simulations on monomer forms.

When we compare the RMSD profiles of dimerization1 regions in every chain as shown in Figure 4.108, we see that the chains of the WT dimer show significantly different trends after 30 ns. RMSD profile of one of the chains starts to increase from 1.3 Å to 2.0 Å while the other stays at equilibrium around 1.0 Å. This increase can be linked to the movement of little helical structure on the dimerization1 region and the turning movement of Thr45 on one of the dimerization1 regions of WT dimer (Figure 4.109). The movements of Thr45 that has no direct contact with the other chain is present also in the simulations on WT monomer forms at acidic pH, diol monomers of mutants at acidic pH, and mutant forms of monomers at neutral pH.

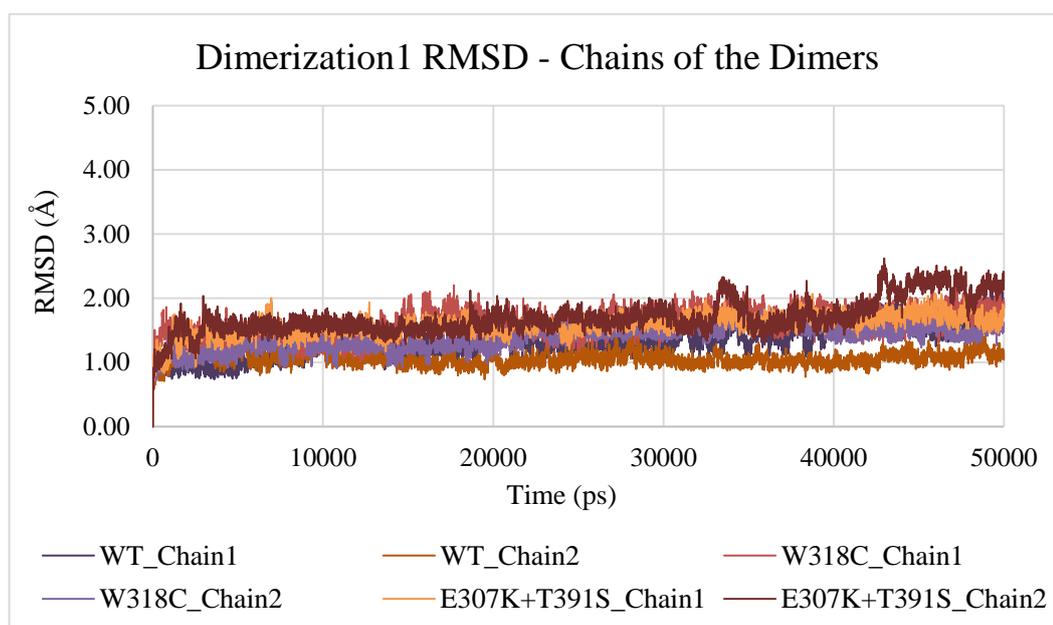


Figure 4.108. RMSD profiles of the dimerization1 regions of the chains in the dimer forms of ARSA at neutral pH.

As in Figure 4.108 the difference between one of the chains of E307K+T391S mutant and WT dimers becomes more significant after 40 ns. Even the lowest trend (green: WT) shows a very slight increase after 40 ns. Due to the trends of the chains maybe it would be better to extend the simulations to see whether the RMSD profiles of these chains continue to increase or reach another equilibrium.

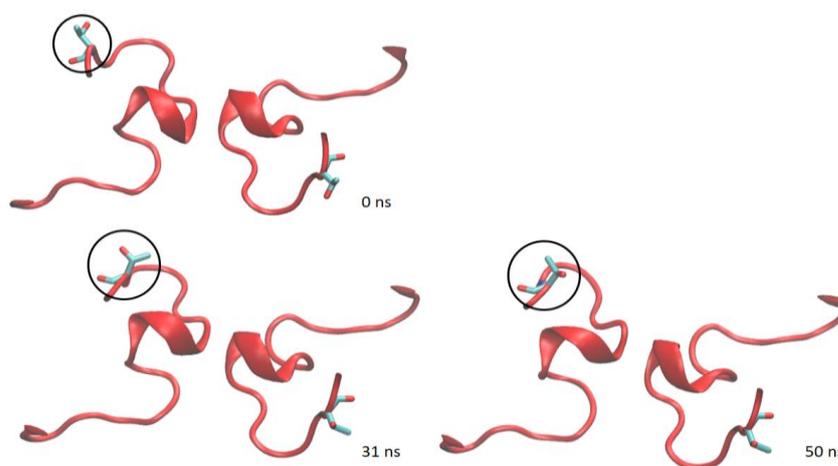


Figure 4.109. Dimerization1 regions (red) of both chains of WT ARSA dimer.

The movements of Thr45 (circled) that is represented with sticks on the dimerization1 region is summarized with three snapshots from the simulation in Figure 4.109. Only the one chain shows high RMSD values in WT dimer, and it is shown that only one Thr45 residue (circled) changes its conformation significantly at the dimer interface of WT ARSA form.

The RMSD profiles of the chains of the W318C mutant dimer show similar trends and reach equilibrium after 25 ns around 1.7 Å. Both chains show very stable dynamics on the contrary to its corresponding monomer forms at neutral pH. When we look at the RMSD profiles of E307K+T391S dimer we see that one of the chains increases its RMSD trend after 40 ns from approximately 1.6 Å to 2.3 Å. The dimerization1 region on the same chain also shows a jump during the simulation around 34 ns. These changes on the RMSD trends of E307K+T391S dimer can be related to the movements of Ser44 in one of the chains. Ser44 is not in direct contact with the other chain but its movements can have an indirect effect on the secondary structure of oligomerization sites. The changes on the dimerization1 region of E307K+T391S mutant are shown in Figure 4.110.

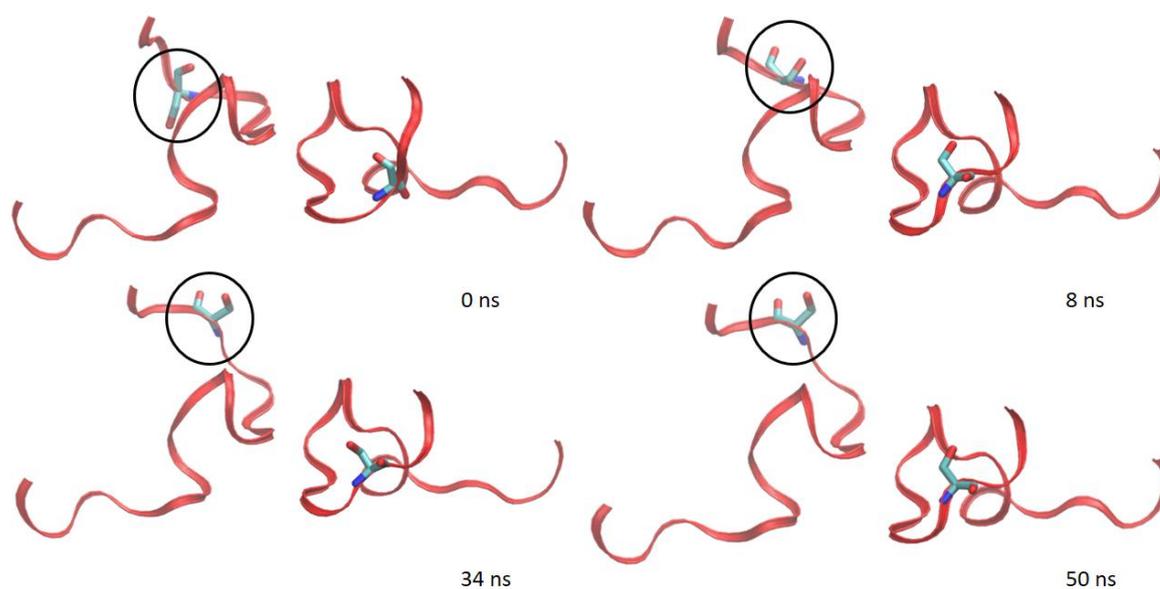


Figure 4.110. Ser44 (stick) at 0 ns, 8 ns, 34 ns, and 50 ns, in the simulation on E307K+T391S mutant dimer.

In the simulation of the mutant dimer, E307K+T391S the dimerization1 region seems to change its conformation but after reaching equilibrium the conformation of the loops in this region does not change more as also shown in Figure 4.110. Ser44 is shown with stick representation and it has high RMSD values in chain2 (in circle).

The aligned average structures of the dimers show the most significant change on octamerization helix of one of the monomers in the dimer structures. Val470 ends of the octamerization helices of the mutant dimers tend to unwind more than WT and the association angles are different than the WT structure. We aligned the structures also based on both octamerization helices, each time we see that the association angle between the two chains in the mutant dimers are different than the WT structure. The significant differences on the average structures were observed on the dimerization regions of the monomers whereas in the dimer forms the octamerization helices and the association angles between monomers seemed to be affected mostly. The screenshots after alignment of the monomer and dimer ARSA forms based on their backbones and octamerization helices are shown in Figure 4.111 to clarify the structural differences.

In Figure 4.111, the color codes for the structures are as follows: light blue = WT, dark blue = W318C mutant, black = E307K+T391S mutant, orange = P377L mutant, and red = T274M mutant.

Figure A shows the average conformation of Leu92 (with sticks representation) which is close to the active site and C-terminus loop of ARSA. Figure B shows the difference between the average conformations of Pro42 which is in the dimer interface and on the dimerization1 region of ARSA. Figure C shows the same Pro42 residue for all forms except W318C mutant. Only the Pro42 residue of E307K+T391S mutant is black, WT and the other mutants have their Pro42 residues colored according to atom name. Three dimer forms (WT, W318C mutant and E307K+T391S mutant) are shown as aligned based on one of their octamerization helices in Figure D. In Figure E, they are shown as aligned based on the backbone structure, thus two images show different alignments of the same dimers. Figure F shows two octamerization helices of the dimers when the average dimer structures are aligned based on the backbone structure.

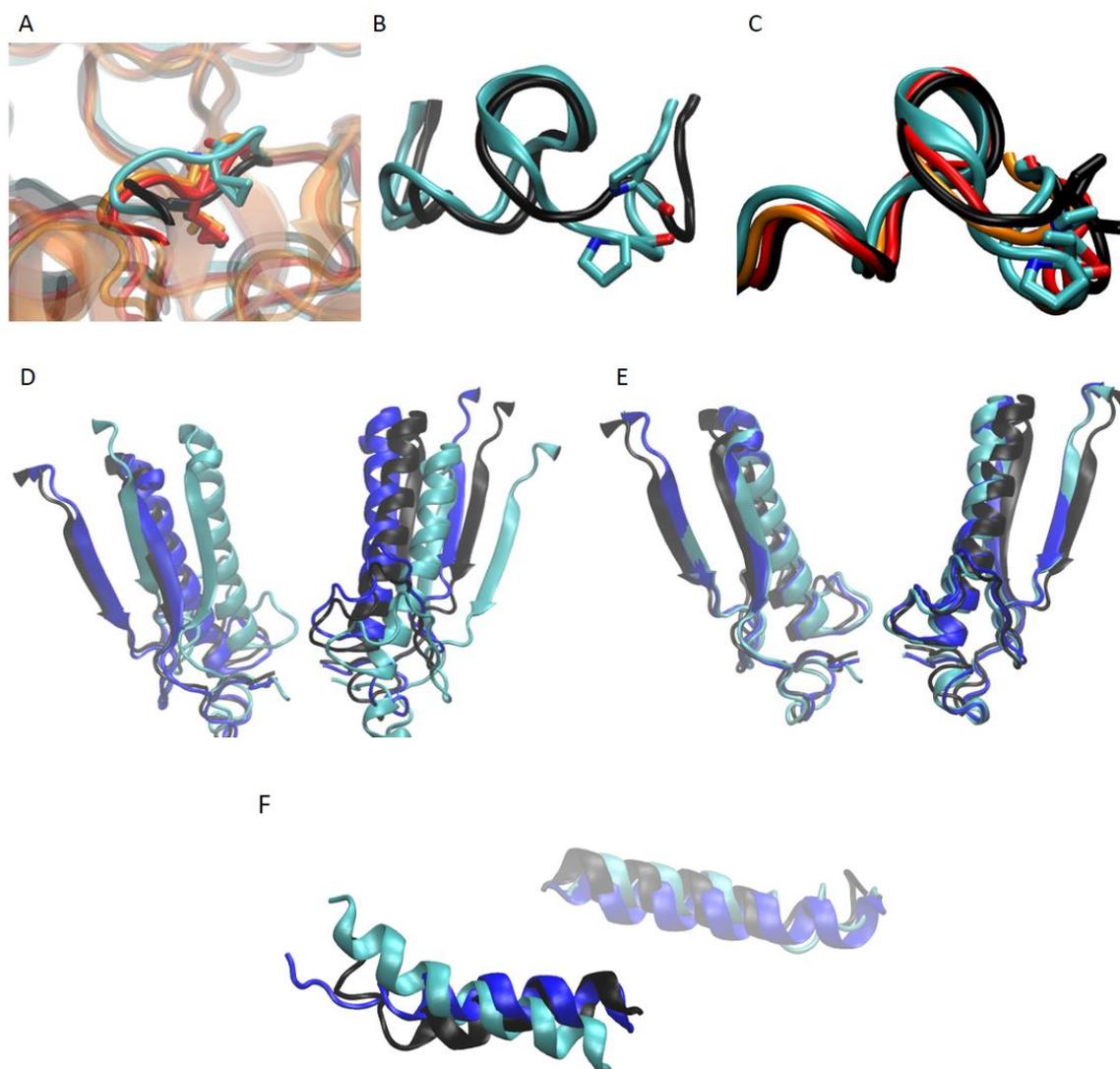


Figure 4.111. Aligned structures for the monomer forms including WT, E307K+T391S, T274M, and P377L forms (upper panel) and specific parts of ARSA dimers aligned both based on one of the octamerization helices and the backbone (below).

As seen in Figure 4.111, in the dimerization region of E307K+T391S mutant Pro42 preserves a structure of an upward conformation whereas the other forms have similar Pro42 conformations, sidechain looking downwards. We know that P377L and T274M mutant forms cannot fold properly and become degraded in ER [4,7] although their overall and regional conformations are much similar to the one of WT ARSA. E307K+T391S shows a different trend in its dimerization region which leads us to consider that this mutation may cause a hindrance in the dimerization of ARSA. Also, the conformational difference of

Leu92 can be a hint of an indirect effect of the E307K mutation, in addition to the polymorphism T391S, on the dimerization of ARSA. There are examples of unexpected indirect effects of the mutations of the specific residues on the oligomerization of the protein for ARSA [61,62]. These small conformational changes on the shown residues may lead to major conformational changes which affect the association angle between the monomers in the dimer structure. In the lower panel of Figure 4.111, this situation can be seen. The alignment of the structures based on an octamerization helix shows that the association angle change between the mutant monomers is significant when compared to WT. On the other hand, overall alignment result in Figure E shows that the difference of the angle between the monomers of WT and W318C mutant becomes almost zero while E307K+T391S mutant dimer still has a conformation of a slightly expanded structure. In addition to these, one of the octamerization helices in the average dimer structures of mutant ARSA forms, the conformational changes on the Val470 ends of the helices of the mutants become significant when compared to WT structure. It means that after reaching the equilibrium, the mutant forms still cannot acquire a helical form on the Val470 end of the octamerization helix. This can be due to the indirect effect of the more mobile dimerization regions in the monomer forms.

#### **4.8.3. Salt Bridges and Hydrogen Bond Occupancy Calculations**

The possible intrachain and interchain salt bridges of the dimer forms of ARSA were also calculated. The only intermolecular possible salt bridge is between Asp50 and Arg58 in the simulations. This salt bridge between the monomers may be crucial for dimer formation and stability in ER. The residue pairs that form possible salt bridges during the simulation and that are absent in any other chain of the dimer forms, were listed with their corresponding distances as in Table 4.41.

Table 4.41. The possible salt bridges in the ARSA chains in dimer forms.

Possible salt bridges	WT		W318C		E307K+T391S	
	Chain1	Chain2	Chain1	Chain2	Chain1	Chain2
Residue Pairs	Chain1	Chain2	Chain1	Chain2	Chain1	Chain2
ASP152-ARG288	7.50 Å	7.27 Å	-	-	-	-
GLU307-ARG311	4.08 Å	4.14 Å	-	-	-	-
GLU307-LYS395	3.63 Å	3.64 Å	3.64 Å	3.62 Å	-	-
ASP152-HIS229	4.71 Å	-	5.00 Å	4.70 Å	-	-
ASP29-HIS229	4.20 Å	4.85 Å	4.80 Å	4.16 Å	5.04 Å	-
GLU481-ARG97	-	-	9.99 Å	6.61 Å*	-	-
ASP50-ARG58 (interchain)	2.15 Å	2.38 Å	3.51 Å	3.42 Å	7.00 Å	7.43 Å
GLU437-ARG299	-	-	7.00 Å	6.60 Å	-	-

\*: The distance between Arg97-Chain2 and Glu481-Chain2 in the simulation on W318C dimer decreases to 4.2 Å after 13 ns and preserves its average distance around 4.0 Å for 26 ns. After 40 ns, these two residues become apart again and the distance between them increases to approximately 11.0 Å.

In Table 4.41, average distances between the center of mass of the oxygen in the acidic side chain and center of mass of the nitrogen in the basic side chain of the pairs were calculated for the whole simulation time.

Table 4.41 shows that there are some changes in the possible salt bridge formations between the monomer and dimer forms at neutral pH. It is expected not to see any possible salt chain between Lys307 mutant and any residue in the simulations on this mutant, but also in the simulation on the dimer form of W318C mutant the distance between Glu307 and Arg311 never comes close as in the simulation on its monomer form, thus the possible salt bridge in the dimer form of W318C mutant disappears. For this residue pair in the simulation on WT dimer, the average distance does not deviate significantly from the average distance in the simulation on the monomer form of WT ARSA. Another significant difference between the monomers and dimers is the possible formation of a salt bridge between two active site residues, Asp29 and His229, when the WT and W318C ARSA forms become dimer. They do not come close in the simulation on the monomer form but on E307K+T391S mutant. Here, one of the chains of E307K+T391S mutant dimer has a close average distance between Asp29 and His229, but in the other chain these residues stay away from each other. The dimer form is expected to be more stable and close to the active form of ARSA. The

occurrence of this possible salt chain in the dimer form may be expected but the situation does not change for E307K+T391S mutant. His229 also comes close to Asp152 which is located near to the active site on the loop at the active site entrance in the simulations on WT and W318C mutant dimers. Although it is observed in one of the chains of the WT dimer it does not occur in the simulation on E307K+T391S mutant dimer. This difference can be linked to the extended overall structure of E307K+T391S mutant dimer which will be discussed in the following section. Lastly, both chains of W318C have possible salt bridge formations between Glu437 and Arg299. Although the average distances between these residues are not significantly small for a conclusion. The position of these residues is at the center of the dimerization region of ARSA.

Hydrogen occupancy percentages were calculated for dimer forms as well. The chains in the dimers were taken separately for the calculations. Since the list is too long we only share the specific results in Table A.4 in the Appendix section.

The first notable change among the hydrogen bond occupancies is the one that Arg19 forms with Asp216 and Glu273. These three residues are on the top of the enzyme and are the first residues at the beginning of the three core beta sheets of ARSA. Arg19 is on the top of the middle beta sheet and is also the first residue after the signal peptide as well as one of the early forming epitopes. The selectivity of Arg19 for hydrogen bond forming towards Asp216 or Glu273 can be associated with a minor stretching movement at the core of the protein, but since these residues are at the top of the protein their interactions probably do not cause a significant overall conformational change, also as seen in RMSD calculations. The central beta sheets continue to the bottom of the protein and at the end of these beta sheets, loops come to the stage with the active site residues on them. As mentioned above the possible salt bridge between His229 and Asp152 may be affected by this hydrogen bond selection of Arg19 (right of left) leading to a change in the association angle between the monomers of E307K+T391S mutant dimer. The other significant change is in the active site due to hydrogen bond forming of Arg73 with the active site residues Asp29, FGL69, and Asp281. In the simulations on dimers each chain shows different bonding occupancies between Arg73 and the active site residues. When we analyzed the simulation on W318C dimer we observed the same situation. There is not a unique sequence for the hydrogen bond forming of Arg73 with active site residues when compared three simulations on dimers. The

different behaviors of the mutant may be due to the mutations but the data is not sufficient to conclude and result. Another notable difference is the hydrogen bond occupancies between Asp169 and corresponding pairs in Table A.4. Asp169 is at the entrance of the active site residue as Asp152 but it is more solvent exposed. The distance between the alpha carbons of Asp169 and Asp152 in the average structure of WT monomer at equilibrium is 9.3 Å. Arg291, which is the opposite neighbor residue of Asp169, gains a farther position from Asp169 in the simulation on E307K+T391S mutant dimer. The average distances between the NH2 and CG atoms of Arg291 and Asp169 respectively are 5.21 Å for WT-chain1, 4.8 Å for WT-chain2, 7.47 Å for E307K+T391S-chain1 and 3.4 Å for E307K+T391S-chain2 in the average dimer structures at equilibrium. The hydrogen bond difference between WT and mutants for the residues Asp169 and Arg291 is also seen in the simulations on monomers forms as indicated above. The indirect effects of the further mutations on the conformational changes affect the interactions between the residues and they also affect the conformation of dimer structures. The extended conformation of E307K+T391S dimer can be linked to these changes. Lastly, Gly171 is close to the C-terminus of ARSA which contains only loop structures. The conformational changes on the average structure become more significant on this region and significant increase in interaction between Gly171 and Asp152 in one of the chains of E307K+T391S mutant dimer can be also linked to these conformational changes. These all can be summed up to conclude that the association angle change between the monomers of E307K+T391S dimer may be occurring because of some minor conformational changes on the dimerization regions and their consecutive effects on the overall structure. In the simulation on W318C dimer, the difference of the hydrogen bond occupancies when compared to WT dimer becomes significant when we look at the results for Asp29. It is one of the active site residues and it prefers Arg73 and Lys123 for forming hydrogen bond more than His227, on the contrary of the simulations on WT and E307K+T391S dimers. We saw that active site structure becomes more rigid in the simulation on W318C monomer too, when looked at the RMSD profiles. Another significant difference is that Asp61 in one of the chains of W318C dimer prefers to form hydrogen bond with Arg390 rather than Thr60. In other chain, it is the contrary and the whole situation is the contrary to those of in the simulations on WT and E307K+T391S dimers. These three residues are located at the dimer interface and their hydrogen bond occupancies may affect the rigidity and consequently the conformation of this area. This may be a hindrance on the oligomerization of W318C mutant ARSA forms. Asp169 and

Thr230 also show different H-bond occupancy results for W318C dimer chains. They are located at the bottom part near to the active site entrance of ARSA. In the case of mutations, we saw different measurements like distance or hydrogen bond occupancies for the residues at the active site entrance of the enzyme when compared to WT. This situation may be linked to misfolding of W318C and this part of the protein may be crucial to understand the effects of the mutation. Finally, as expected, Cys318 in one of the chains of W318C mutant dimer cannot form hydrogen bonds with other residues but water molecules in a low occupancy. This region is crucial since Trp318 is close to the first forming epitope, Arg19 such that in the crystal structure the distance between Arg19 and Trp318 is around 14 Å. During folding of ARSA the distinctive change in this region may cause the protein to be misfolded.

#### 4.8.4. Monomer-Monomer Clashes and SASA Calculations

The coordinates of the WT monomer from the last frame of its simulation were aligned on the coordinates on the dimer form from the last frame of dimer form's simulation. Only clashes were observed between the residues Leu297 of both monomers. The aligned structures are shown in Figure 4.112.

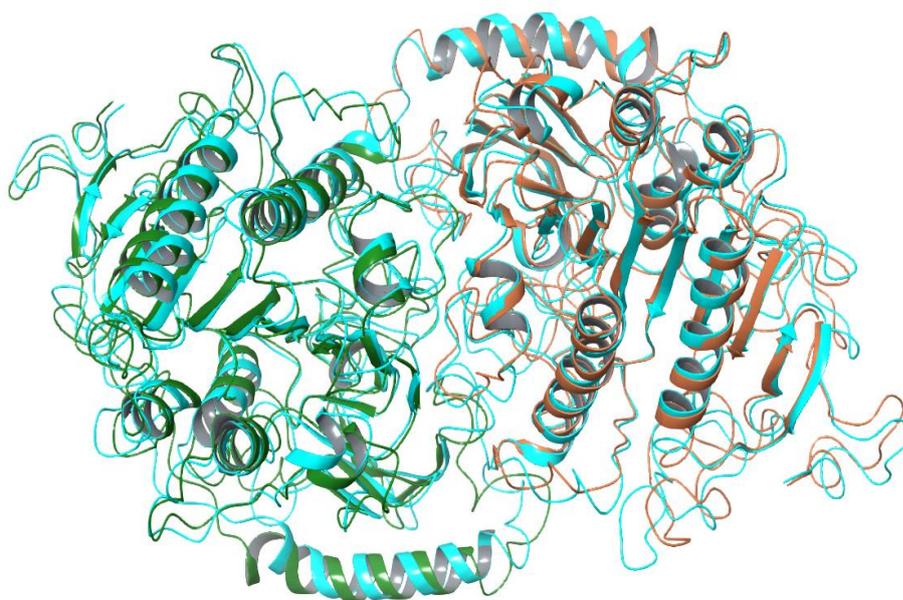


Figure 4.112. WT monomers in orange and green as aligned on the WT dimer (blue).

When the same procedure was applied to the mutant forms, the clashes in the dimer interface become more. For the E307K+T391S mutant, Ty439 clashes with Gly34 and Ser290, Lue297 clashes with Gly296 which are on the dimerization2 and dimerization3 regions. The clashes between the W318C mutant monomers are between Pro42-Arg58 and Ser43-Thr60 which are close to dimerization1 region. The dimerization regions become closer to the octamerization helix in the increasing order of their sorting. This shows that in the simulation on mutant E307K+T391S, the mutation affects the dynamics more of the back side of the protein. These results may imply that the octamerization of E307K+T391S mutant is indirectly affected by the conformational changes in the dimerization2 and dimerization3 regions, but the clashes may not be significant since the dimer surface of the monomers are free to move in the solvent box and the dimerization mechanism of ARSA is not known in terms of conformational changes that occur on the residues on the dimer interface for dimerization. To understand how the area is different between the dimer and the monomer structures we calculated the Surface Accessible Surface Area (SASA) values for the whole trajectories and took the average of them for the monomer and dimer ARSA forms at neutral pH. The results are listed in Table 4.42.

Table 4.42. The average SASA values of the monomers and dimers with their corresponding standard deviations in Å<sup>2</sup>.

	<b>WT monomer</b>	<b>W318C monomer</b>	<b>E307K+T391S monomer</b>
average	19077.2	18775.4	18886.2
std. dev.	287.2	276.0	352.7
	<b>2x (WT mono)</b>	<b>2x (W318C mono)</b>	<b>2x (E307K+T391S mono)</b>
average	38154.3	37550.8	37772.5
	<b>WT dimer</b>	<b>W318C dimer</b>	<b>E307K+T391S dimer</b>
average	34657.5	33728.7	34234.7
std. dev.	455.0	375.0	468.1
difference	3496.9	3822.1	3537.8

The results of the monomers in Table 4.42 were multiplied by 2 and the total SASA of dimers were subtracted from these values to find the enclosed dimerization area for each

form of ARSA. They show that the difference between the embedded SASA of the dimer interface of WT and E307K+T391S mutant is  $41 \text{ \AA}^2$ . This difference can be linked to the conformational differences between dimer forms of ARSA. The dimer interface becomes more open and solvent exposed in the simulation on E307K+T391S mutant. This can cause a change in the association angle between chains leading to hindrance in the octamerization of ARSA in lysosome as reported before for the mutation C300F [61,62].

#### **4.9. MD Simulations on Dimer ARSA Forms at Acidic pH**

We tried to explain the effects of the mutations on the ARSA monomer, now we analyze the simulations on dimers at acidic pH. After being transferred into the lysosome ARSA dimers form octamers and become active. After the ligand enters the active site cavity FGL69 becomes hydrated and form a diol for the nucleophilic attack part of the reaction. These steps happen too quickly to catch with experimental methods [1]. For this reason, although we performed the monomer ARSA forms with their FGL69 residues in its hydrated state, we only discuss the results of the forms of nonhydrated state FGL69 containing ARSA dimers because the stability in the lysosome is a prerequisite for the activity of the enzyme.

##### **4.9.1. Visual Comparison of the Simulation Snapshots**

The coordinates of the last frames were taken from the dimer simulations of two WT, W318C, E307K+T391S, and C300F mutant forms. They were aligned based on their backbones and the structural changes were compared accordingly. After the backbone alignment, their RMSD values were calculated and tabulated as in Table 4.43. The backbone alignment and the RMSD calculations were done by taking the coordinates from the first simulation on WT dimer, that is why the RMSD result in the first row is zero due to self-alignment.

Table 4.43. RMSD values between last simulation snapshots of dimer ARSA forms at acidic pH.

Dimer Form	RMSD (Å)
WT #1	0.00
WT #2	2.61
W318C	2.77
E307K+T391S	2.39
C300F	2.62

The results in Table 4.43 show that overall structure does not affected from the mutations when compared to the RMSD result of WT #2. The results are also twice of the monomer forms which are in accordance with the size and RMSD relationship as well as the preserved overall stability of the enzyme.

Overall structure seem preserved but the octamerization helices and the association angles in the dimers are seen to be affected by the mutations. For better understanding we took the snapshots of the aligned structures of the last frames as shown in Figure 4.113.

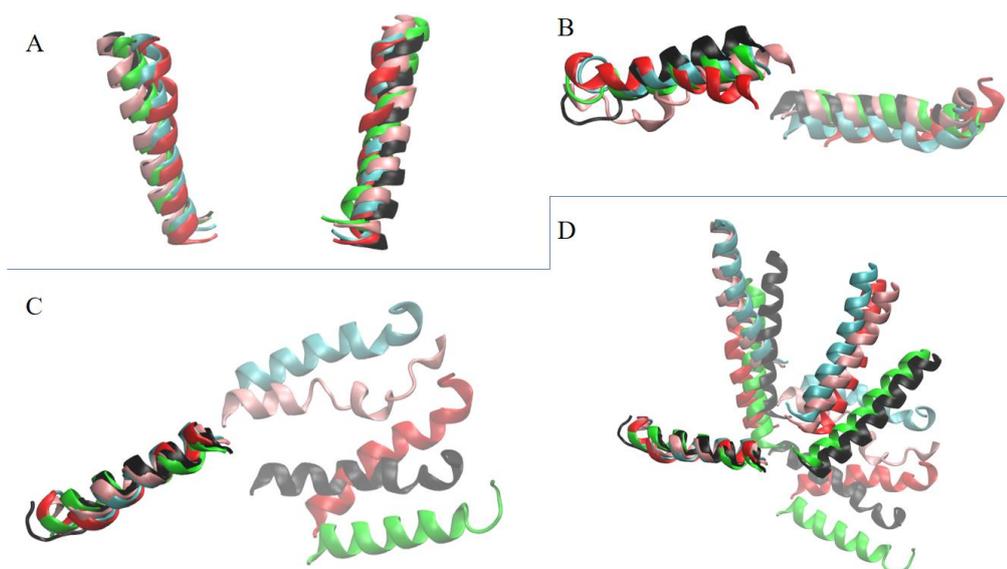


Figure 4.113. Dimer forms of ARSA aligned based on the backbone (upper panel). Lower panel: the same forms as aligned on one of the octamerization helices.

In Figure 4.113, figure A shows the vertical longest helix in the structure and figure B shows the octamerization helices of the dimers. Figure C shows the octamerization helices, while figure D shows also the vertical longest alpha helices ( $\alpha$ D and  $\alpha$ E) along with the octamerization helices and the color codes are as follows: WT#1=light blue, WT#2= pink, W318C=green, E307K+T391S=red, and C300F=black.

The last frames show that the association angle change is hard to see when the backbone alignment is performed, but if we take the light blue structure as the reference we can say that, C300F monomers become slightly apart while E307K+T391S monomers become slightly closer to each other according to the upper panel of Figure 4.113. When we look at figure B in the upper panel it becomes clear that the mutations have some conformational affects on the octamerization helix structure, but even between the two WT forms show differences in terms of the secondary structure of their octamerization helices.

The differences become significant in terms of association angle, in the cases of C300F and E307K+T391S mutant dimers. Before coming to any conclusion, we should note that the conformational changes should be examined with the dynamics of the dimerization region if the mutations cause any indirect hindrance in the octamerization of ARSA dimers as C300F does.

Lower panel shows the differences more significantly. When the forms are aligned based on one of the octamerization helices the angle differences between monomers become more significant for the mutants: C300F, E307K+T391S and W318C. From figure D, we clearly see that the monomers get away from each other in the simulation on W318C mutant when compared to WT, but the position of the octamerization helix of W318C at back bottom in Figure D can be an outcome of this separation of monomers. On the other hand, C300F and E307K+T391S mutants have more closely positioned monomers while their octamerization helices have significantly different conformations. Accordingly, we can say that the inactivity mechanism of E307K+T31S may be similar to one of C300F mutant [61,62].

After examining the last frames, we also calculated the average structures from the production runs of the simulations and compared them with the same procedure as follows.

#### 4.9.2. Structural Changes and Average Structure

The average structure of each form was calculated and generated after their backbone RMSD profiles reach a plateau for the forms at acidic pH as previously described. After the backbones aligned the RMSD values were calculated and the results are summarized in Table 4.44. The backbone alignment and after RMSD calculation were performed based on WT structure, thus WT has 0.00 Å RMSD.

Table 4.44. RMSD values of the simulation average structures of ARSA dimers at pH=5.

Dimer Form	RMSD (Å)
WT	0.00
WT #2	2.38
W318C mutant	2.46
C300F mutant	2.51
E307K+T391S mutant	1.97

The results in Table 4.44 show that also at the acidic milieu, the overall protein structures do not show significant differences after reaching equilibrium. The specific regions also do not show significant conformational differences in the monomers.

The average structures of the dimers that were aligned based on their backbones do not show significant change on octamerization helices, but on the Val470 end of the helices some deviations occur that are shown in Figure 4.114. When the average dimer structures were aligned based on one of the octamerization helices a chaos occurs. In this situation, below panel in Figure 4.114, the shifted positions of the chains in the dimer forms can be clearly seen, for better understanding only the longest alpha helices of the dimers are shown. This depicts how the association angle changes between chains.

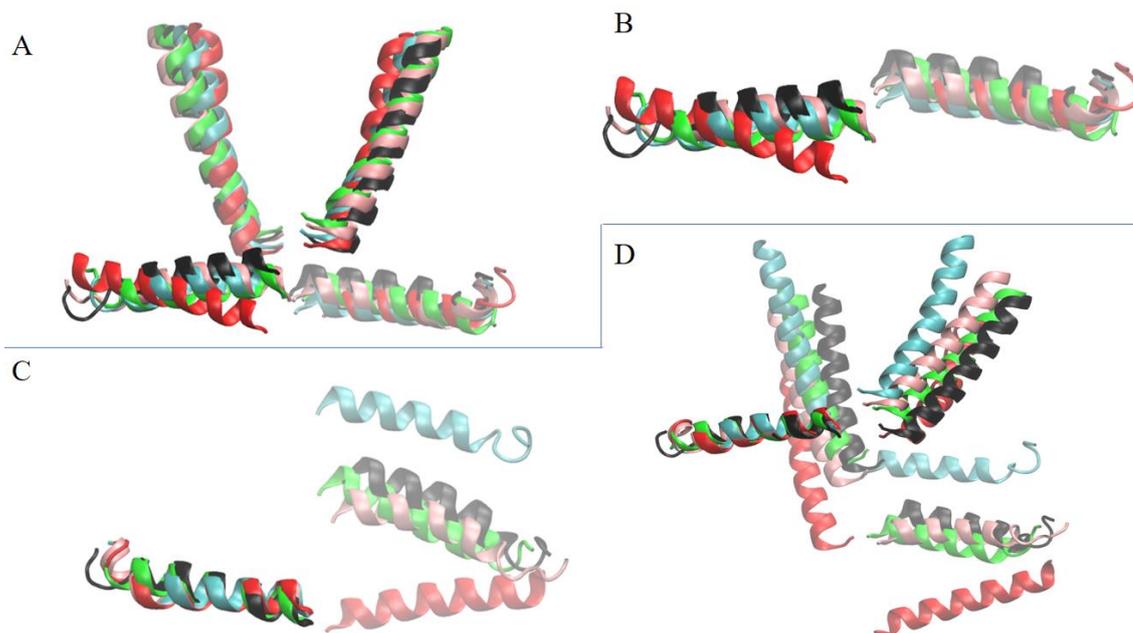


Figure 4.114. Octamerization helices ( $\alpha I$ , horizontal ones) and longest helices ( $\alpha D$  and  $\alpha E$ ) of the averagere structures after backbone alignment.

The average structures were generated from the frames of the production runs. The lower panel in Figure 4.114 shows the same secondary structures after the average structures were aligned based on one of the octamerization helices. Figure C shows only the octamerization helices while figure D also shows the longest alpha helices of ARSA dimers.

The figures show that E307K+T391S mutant dimer has a significant association angle change between its monomers when compared to WT and C300F dimers at acidic pH. The difference is more remarkable when the alignment is done on backbone of the average structures because when the alignment is done on one of the octamerization helices whole dimer shifts its position and it becomes harder to understand the angle changes when compared to the WT structure's coordinates. To understand the differences better, we again performed RMSD and RMSF calculations as follows as well as the possible slat bridge and hydrogen bond occupancy calculations.

### 4.9.3. RMSD and RMSF Calculations – Overall Structure

The whole protein RMSD results were calculated as described before and the results for the dimers at acidic pH are given in Figure 4.115.

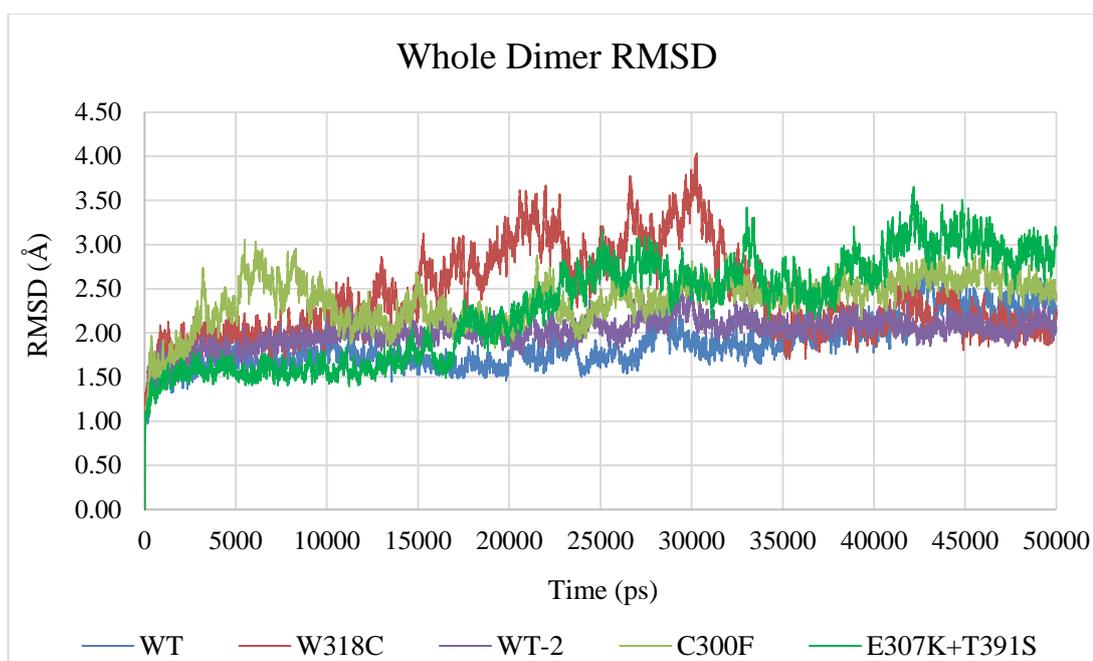


Figure 4.115. The RMSD profiles of the whole dimer backbones at acidic pH.

The results in Figure 4.115 show that the backbone RMSD profiles of the two mutants (E307K+T391S) cannot reach equilibrium until the last 10 ns, and the backbone profile of W318C dimer shows the most deviation from its initial coordinates for almost 15 ns and then reaches equilibrium at 35<sup>th</sup> ns of the simulation. In addition to the Overall structure RMSD calculations we also calculated the RMSD profiles of each chain in the dimers and RMSF profiles of the carbon alpha atoms in those chains as shown in Figure 4.116 and Figure 4.117.

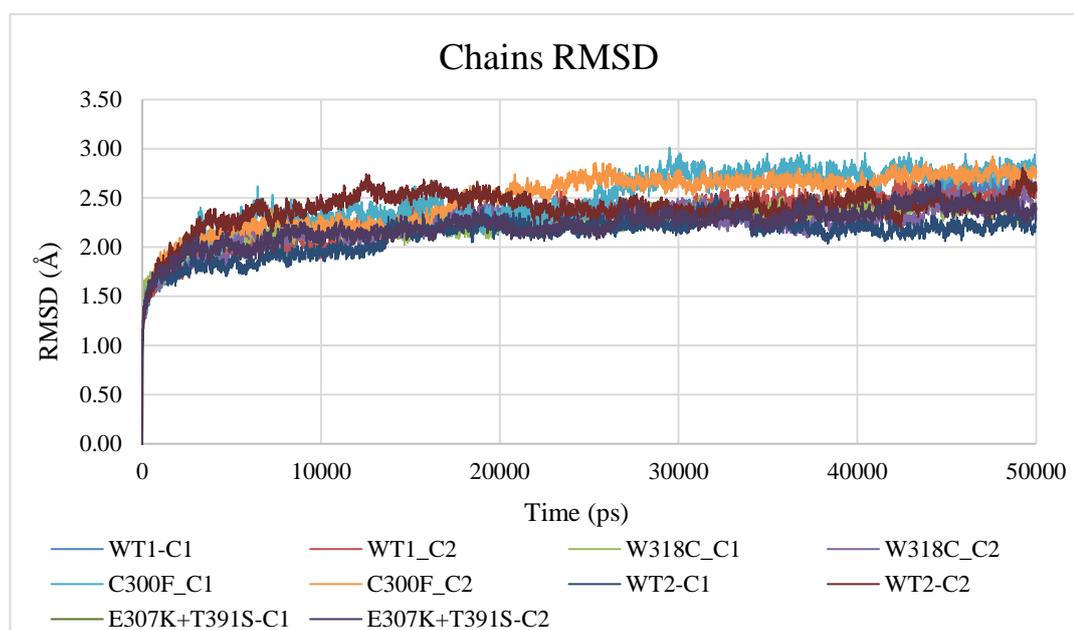


Figure 4.116. RMSD profiles of each chain in the dimer forms at acidic pH.

The RMSD profiles of the chains do not show significant differences but the chains of C300F dimer show the highest RMSD trends. These changes should be occurring due to conformational changes in the specific regions. The results of the specific regions are also discussed in the next section.

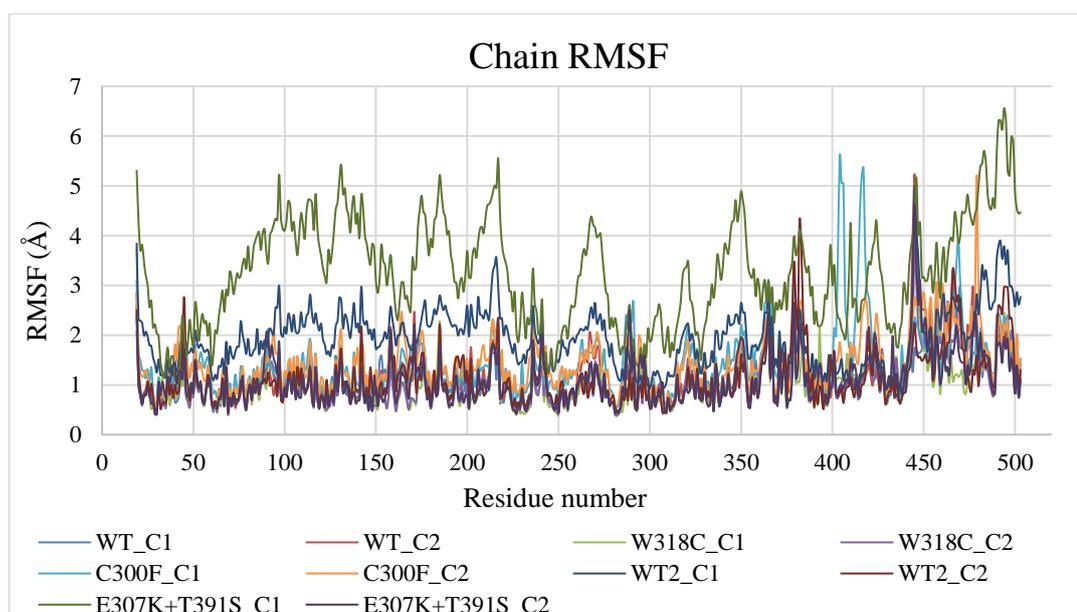


Figure 4.117. RMSF profiles of the residues on each chain in dimer forms at acidic pH.

The most significant RMSF fluctuations occur on one of the chains of both mutant dimers, C300F and E307K+T391S. In addition, the missing part of the crystal structure has the highest RMSF values in the profile of the other chain of C300F (light blue in Figure 4.117). According to these results E307K mutation may be causing a severe change in the association angle between monomers or it may not be preserving its dimer stability in lysosome. For better understanding we also calculated the RMSD profiles of the specific regions as follows.

#### 4.9.4. RMSD Calculations – Key Structural Elements

Specific region RMSD results are given as follow from Figure 4.118 to Figure 4.122.

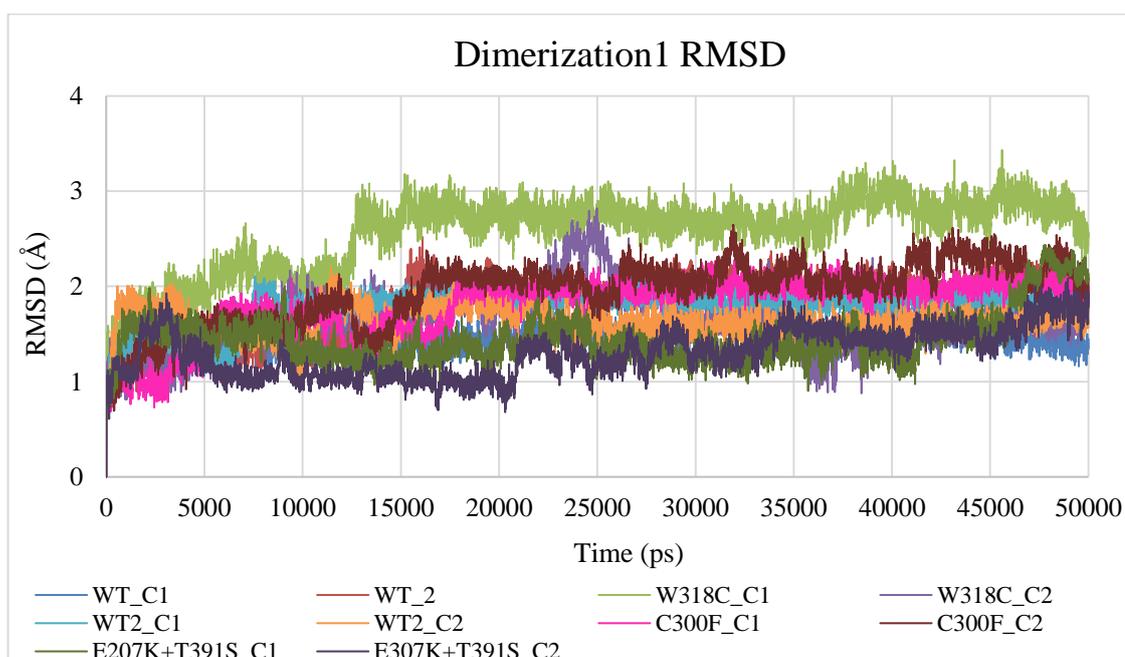


Figure 4.118. RMSD profiles of the dimerization1 regions of the chains in the dimers at acidic pH.

Dimerization1 RMSD profiles of W318C mutant's Chain1, C300F mutant's Chain2 and both chains of E307K+T391S cannot reach equilibrium during the simulations. These instabilities can be linked to the conformational changes of specific residues at dimer

interface. The changes are shown for certain time frames from the simulations after all the RMSD profiles of the specific regions are given as follows.

Secondly, all the RMSD profiles of WT dimers for both chains reach equilibrium after 15 ns as shown in Figure 4.119, and all of them show a decrease in the last 2 ns. This can be interpreted as that dimerization1 regions of WT chains are more stable than those of mutant dimers.

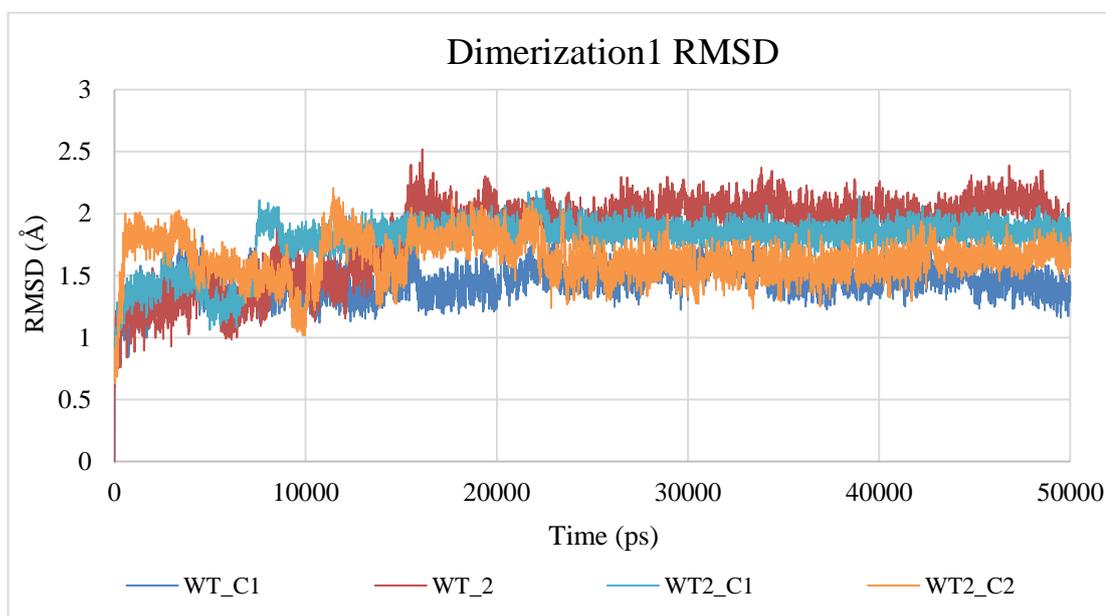


Figure 4.119. Dimerization1 region RMSD profiles of WT dimer chains at acidic pH.

Dimerization2 region RMSD profiles are given in Figure 4.120. One of the chains of C300F and E307K+T391S mutant dimers show the highest deviations from the initial coordinates during the simulations. The region of C300F reaches equilibrium after 40 ns around 3.2 Å while the RMSD trend of E307K+T391D (dark purple) starts to decrease again after 45 ns. A sharp decrease around 45 ns is also observed for the other chain of C300F RMSD profile (yellow in Figure 4.120). Other dimerization2 region RMSD profiles of the chains do not show significant changes and differences, thus a similar inactivity mechanism is considered to occur for the mutations C300F and E307K+T391S.

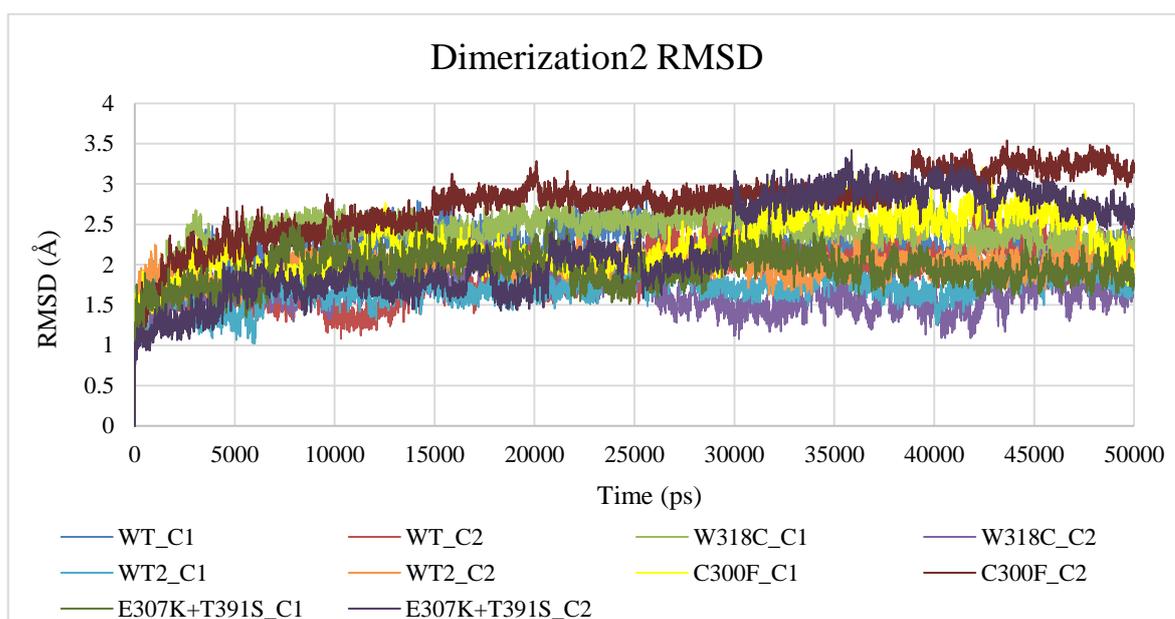


Figure 4.120. RMSD profiles of the dimerization2 regions of the chains in the dimers at acidic pH.

Thirdly, dimerization3 region RMSD profiles are shown in Figure 4.121.

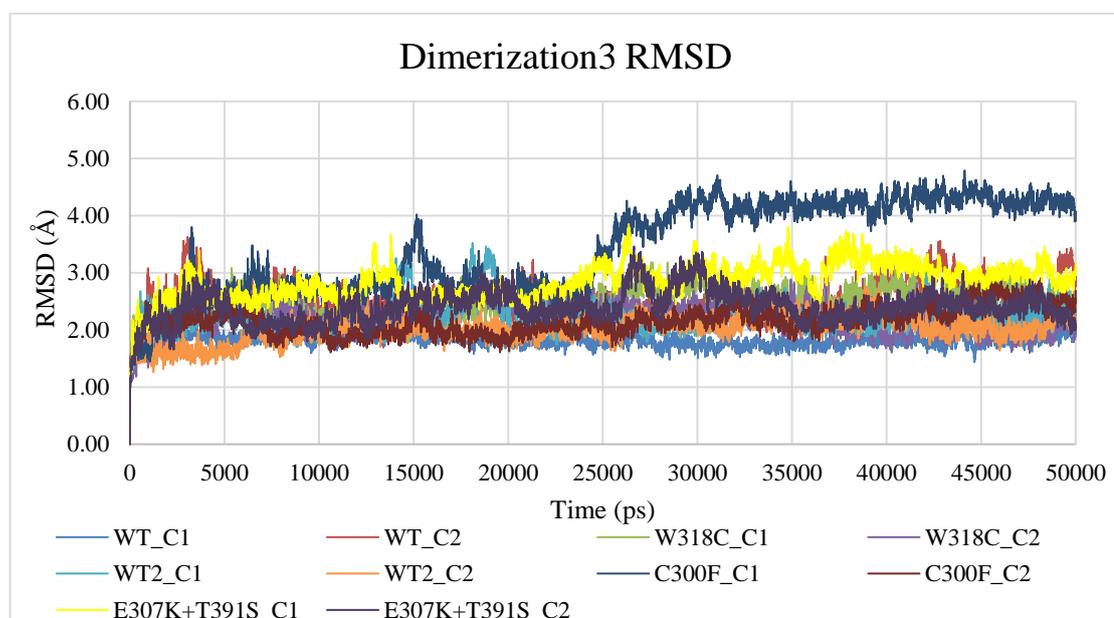


Figure 4.121. RMSD profiles of the dimerization3 regions of the chains at acidic pH.

The most significantly deviated RMSD profile belongs to one of the chains of C300F. The loop which also includes Val445 changes its conformation severely during the simulation as shown in the following figures which also explain the conformational changes on dimerization1 regions. When we look at the dimerization3 region RMSD profiles of WT dimers we see that no significant changes occur and the RMSD profiles of the chains follow a flat trend after 10 ns around 2.20 Å. For dimerization3 region this result can be considered as very stable or rigid, and the RMSD profiles of W318C and E307K+T391S chains follow the similar trends of WT profiles.

Lastly, the octamerization helices' RMSD profiles are depicted in Figure 4.122.

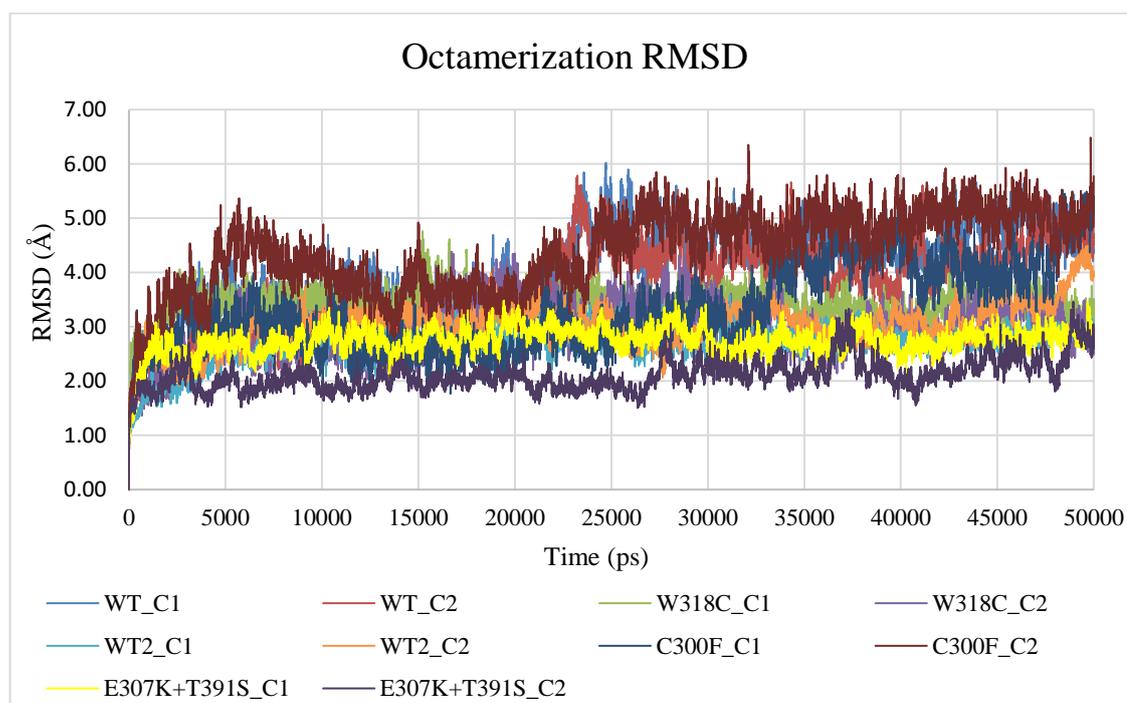


Figure 4.122. RMSD profiles of the octamerization helices of the chains at acidic pH.

The highest deviations among the octamerization helix RMSD profiles belong to C300F chains and one of the WT chains. The lowest RMSD trends, on the other hand, belong to the chains of E307K+T391S dimer. It is surprising that the last frame and average structure alignment results showed the contrary. W318C chains follow similar trends as the ones in the first simulation on WT dimer. Most significant changes were observed for the octamerization helices of the mutants, and the association angles between monomers

especially the ones in the dimers at acidic pH, but the RMSD calculations of octamerization helices show that the secondary structure change on the helix may not be related to the general stability of the dimer and the association angle between monomers. As stated before, Val470 end of the octamerization helices are very flexible and may not be a determinant of the effects of the mutations.

The conformational changes of the specific regions are shown for the mutant dimers for the certain time frame of the simulations from

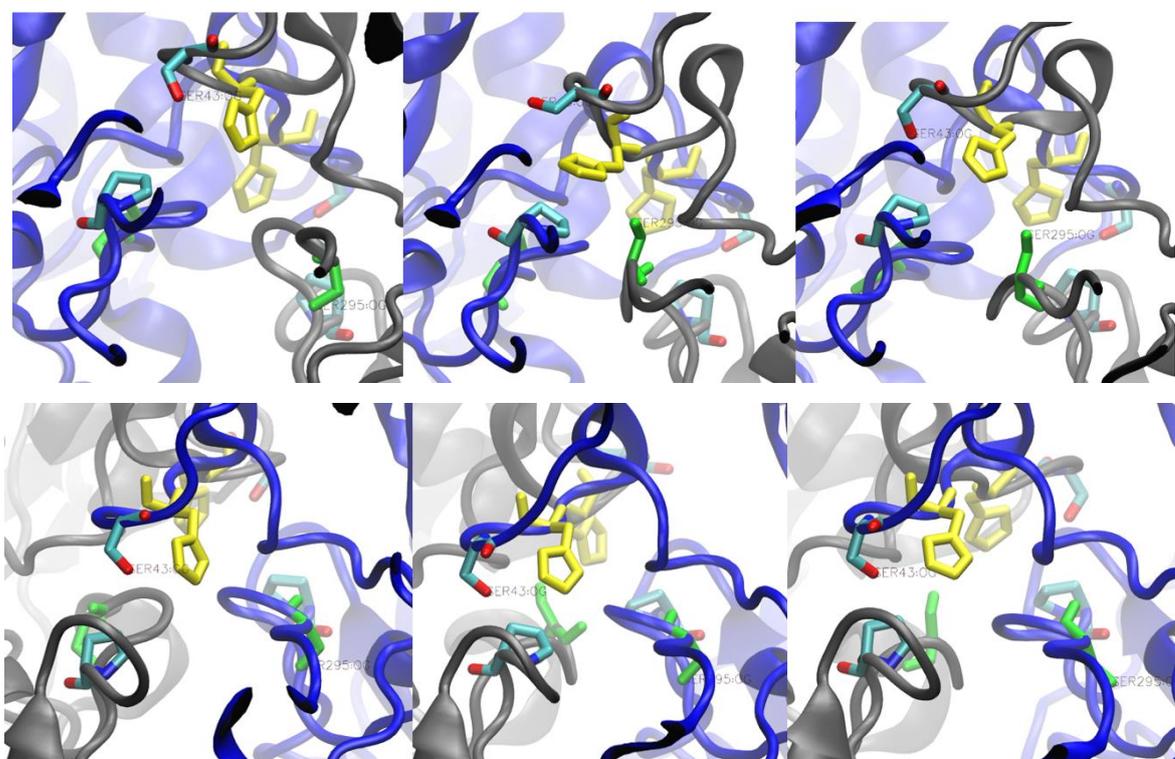


Figure 4.123. His41 (yellow), Ser43 (light blue above yellow residue), Ser295 (green), and Pro435 (light blue on the dark blue chain) in both chains of C300F dimer, at 0 ns, 21 ns, and 50 ns.

His41 is in the dimerization1 region and as it is seen from the figure the conformational changes of the residue also changes the secondary structure from beta sheet to loop and again to beta sheet. Also, another dimerization region residue, Ser295 which is on the dimerization2 loop, changes its conformation after 20 ns as shown in Figure 4.123. Ser43 also turns its side chain towards left and then its former position. These movements can be

linked to the high RMSD results in the dimerization regions, but they are not symmetric for the chains as shown. Additionally, Ser295 (green) gets close to His41 (yellow) as shown in the middle panel of the above figure in Figure 4.123. Accordingly, the distance between NE2 atom of His41 and OG atom of Ser295 was calculated and shown in Figure 4.124.

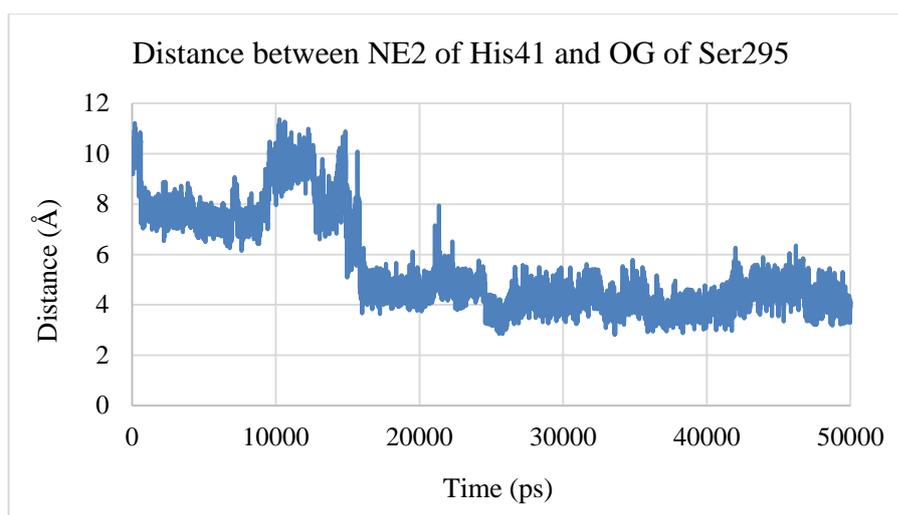


Figure 4.124. The distance between His41 and Ser295 in in C300F dimer.

The movements of Met289 at the dimer interface in the simulation on C300D dimer are also shown as follows, in Figure 4.125. The loop region changes its conformation during the simulation, that is why we analyzed the movements of the residues in this region.

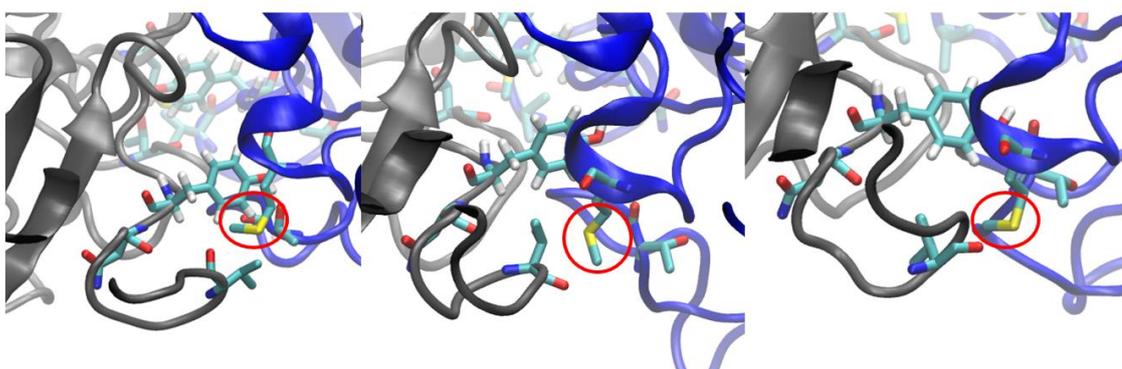


Figure 4.125. Met289 (in circle) at 25 ns, and 50 ns of the simulation in C300F dimer.

In Figure 4.125, the movements of Met289 are shown. It is in the dimerization2 region and Methionine is a hydrophobic residue. The conformational changes in the dimer interface may be causing C300F dimer to lose its proper association angle due to hydrophobic interactions. The secondary structure of the region changes as it is shown which can be linked to the instable RMSD profiles in Figure 4.120. The movements of Met289 on the other chain are also shown for the same time frames as follows in Figure 4.126.

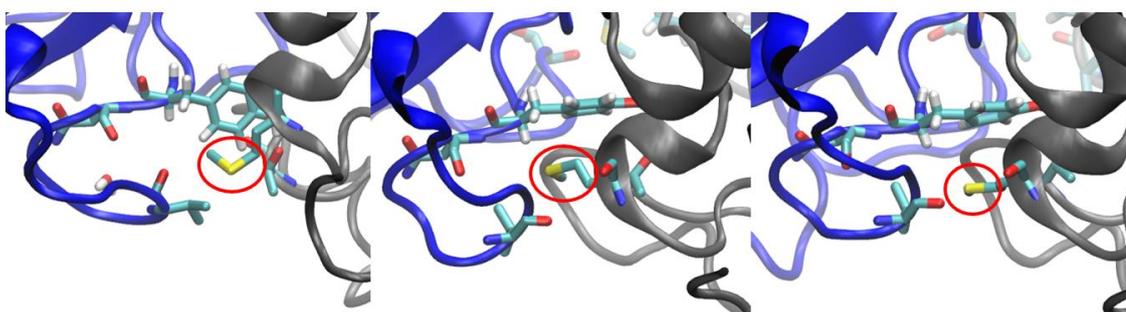


Figure 4.126. Met289 (circled) in the other chain at 25 ns, and 50 ns in C300F dimer.

Met289 on the other chain of C300F do not change its conformation as severely as the one in the other chains as depicted before. This one generally preserves its place between Asp407 and Val445. They are negatively charged and hydrophobic residues respectively, and the charged one is located on the solution side which may be playing a role of a shield for the hydrophobic residues inside, thus severe conformational changes are not observed in this side of C300F.

We also analyzed the movements of the residues in the specific regions of E307K+T391S mutant dimer at acidic pH and showed them as follows for the certain time frames from the simulation. Especially the movements of the mutant residue E307K were observed and the conformational changes of Lys307 of the mutant dimer are shown in Figure 4.127 as follows.

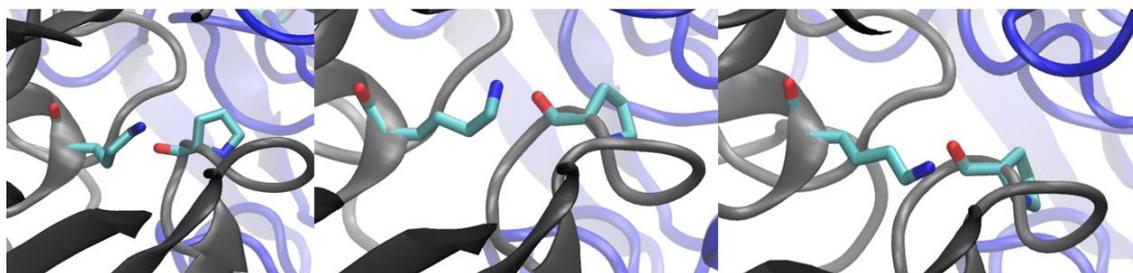


Figure 4.127. Lys307 (left) and Pro435 (right) in the same chain of E307K+T391S dimer at 8 ns and 50 ns.

Pro435 is a hydrophobic residue while Lys307 is polar and basic, thus no repulsion should occur between these two residues. As shown in Figure 4.127, in one chain of E307K+T391S mutant dimer, they move their side chains toward each other and preserve their positions as they are shown. On the other chain of the mutant, they cannot engage in a stabilized position together as shown in the following figure, Figure 4.128.

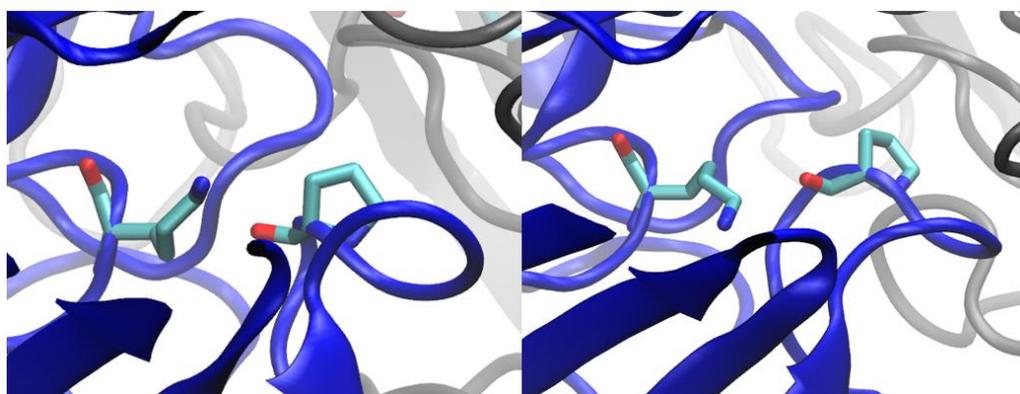


Figure 4.128. Lys3307 (left) and Pro435 (right) at 0 ns and 50 ns in E307K+T391S dimer.

In the other chain of E307K+T391S mutant dimer at acidic pH, Lys307 and Pro435 cannot preserve stable conformations and this causes the loop region (dimerization3 region) on which Pro435 is located, move flexibly during the whole simulation. This can be linked to the high RMSD trend of chain1 of the mutant dimer in Figure 4.121.

Second significant difference is that the intrachain salt bridge between Glu240 and Lys433. The residue pair of the mutant, E307K+T391S is shown in Figure 4.129 for the

certain time frames when interacting and not. This interaction is observed only on one side of the dimer, and it may be the reason of the changed angle between monomers.

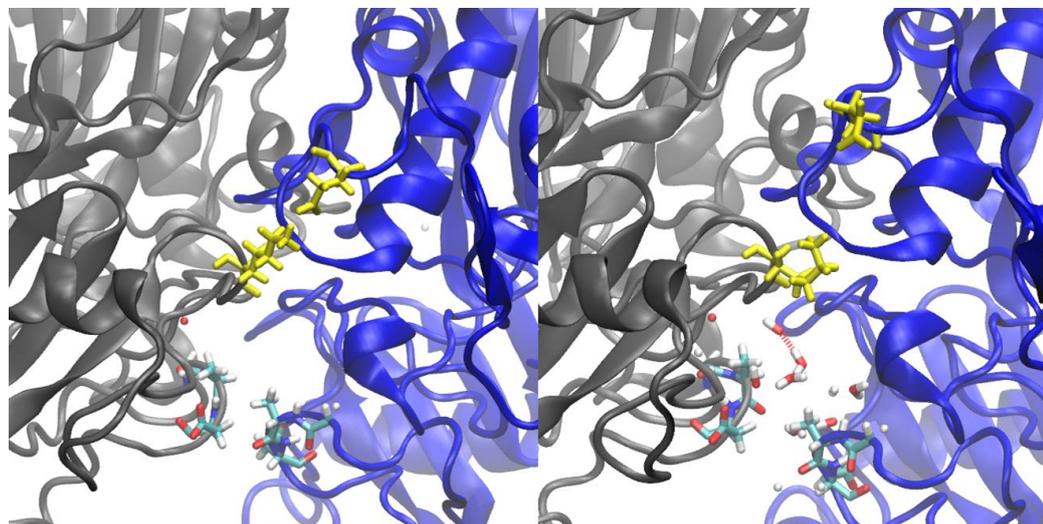


Figure 4.129. Glu240 and Lys433 (yellow on the right and left) and their interaction at 42 ns (left panel) and at 50 ns (right panel).

The the bottom part of the two panels in Figure 4.129 are shown two residues namely Ala410 and Ala413. Ala410 changes its conformation during the simulation and late in the simulation on the mutant dimer, waters molecules come in as shown in the right panel. This situation may affect the dimer stability and the angle between E307K+T391S monomers. The trajectory of W318C mutant dimer was analyzed and the only changes were observed in the dimerization1 region between residues as shown in Figure 4.130.

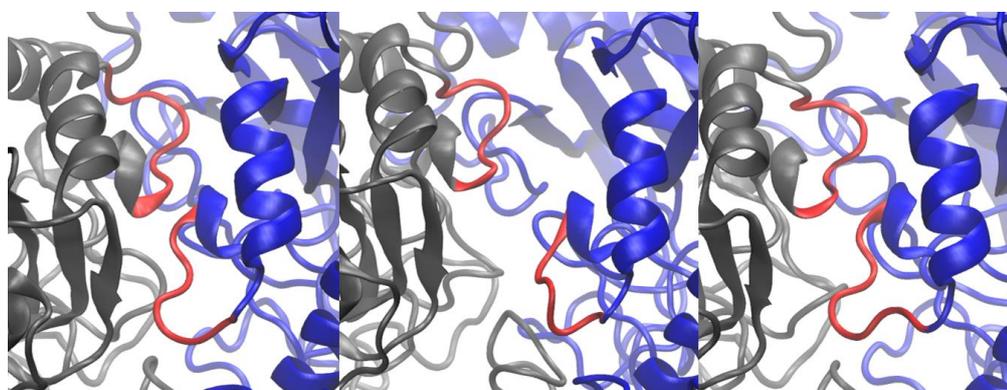


Figure 4.130. Dimerization1 region of W318C dimer at 25 ns and 50 ns.

The conformational changes shown in Figure 4.130 can be linked to the conformational changes of Pro42 in the region. The loops get away from each other but then acquire their initial positions. The changes on this region may not be crucial for dimer stability since the mutant dimer do not show significant changes at the last frame and as the average structure when compared to WT forms.

#### 4.9.5. Salt Bridges and Hydrogen Bond Occupancy Calculations

The possible salt bridge and hydrogen bond occupancy calculations were done for each chain in the dimer forms at acidic pH. The hydrogen bond occupancy results are given in Table A.6. The list is very long and since the secondary structure and association angle changes are more important for us at acidic pH, only the most significant differences are summarized briefly as follows.

The differences in hydrogen bond occupancies are observed between:

- Asp173 and Asn152 for E307K+T391S mutant chains and one of the chains in C300F dimer. It almost disappears in the first chain of the dimer and becomes one third of the first simulation on WT dimer for its second chain. On the other hand, the occupancy increases to 53 % and 37 % in the first and second chains of E307K+T391S mutant dimer.
- Glu272 and Trp/Cys318 for W318C, C300F mutant and WT#2 dimers. The occupancy between these two residues decreases in the simulations on these dimer forms. It is interesting that the dimer formation changes the significant difference of hydrogen bond preferences of Trp/Cys318 residue. It prefers Ile322 in the simulations on the dimers mentioned above in one of the chains and surrounding water molecules in the other chain of the dimers.
- Asp335 and Arg370 in the simulation on the second WT dimer form. The occupancy becomes almost half for this residue pair and Asp335 in one of the chains forms hydrogen bond mostly with Ser332. This shows that the hydrogen bond between

Asp335 and Arg370 may not be crucial in the dimer form for the stability of the glycosylation site.

- Glu/Lys307 and Glu437 in the simulation on E307K+T391S dimer as expected. The mutation on this region causes the hydrogen bond occupancy between these residues drop to less than the half of those in the other simulations on dimer forms. This change can affect the association angle between monomers due to instable monomer dynamics in each chain as in the case of C300F mutation.
- Asp430 and Ser432 in the simulations on C300F and E307K+T391S mutant dimers. Asp430 mostly forms hydrogen bonds with the surrounding water molecules approximately with 90 % occupancy in the simulations on these two mutants, while it forms hydrogen bonds with Ser432 and Lys393 in the simulations on WT dimers around 35 % occupancy. Asp430 is located at the bottom of the dimer interface and these changes in the hydrogen bond preferences of this residue may be one of the causes of the association angle change between the monomers.

All the possible salt bridges were listed in Table A.7. The differences in salt bridges were listed as in Table 4.46. The possible interchain salt bridges of the dimer forms of ARSA were also calculated. This time, in the simulation on E307K+T391S mutant dimer at acidic pH, there is an extra possible interchain salt bridge which is Glu240-Lys433. The only intermolecular possible salt bridge was between Asp50 and Arg58 in the simulations on dimer forms at neutral pH. This extra salt bridge can be linked to the significantly different conformations of the chains after the alignment as shown in Figure 4.114. Average distances between the center of mass of the oxygen in the acidic side chain and center of mass of the nitrogen in the basic side chain of the pairs were calculated for the whole simulation time.

Table 4.45. The possible salt bridges in the corresponding dimers of ARSA at acidic pH.

PSB	WT		WT2		W318C		C300F		E307K-T391S	
	C1	C2	C1	C2	C1	C2	C1	C2	C1	C2
ARG19- GLU272	-	-	-	-	-	-	-	-	9.70 Å	-
ARG19- GLU273	6.54 Å	4.35 Å	5.77 Å	4.22 Å	7.46 Å	6.87 Å	-	5.73 Å	4.20 Å	7.71 Å*
ARG299- GLU437	-	-	-	-	-	7.23 Å	-	-	-	-
ARG479- ASP482	-	-	-	-	-	-	-	13.25 Å	-	-
ARG479- GLU481	-	-	-	-	-	-	-	12.00 Å	-	9.82 Å
ASP169- ARG288	-	-	-	-	-	-	6.75 Å	5.05 Å	6.19 Å	8.52 Å
ASP29- HIS229	4.51 Å	Asp30- His229	Asp30- His229	Asp30- His229	Asp30- His229	Asp30- His229	-	4.87 Å	Asp30- His229	4.59 Å
ASP50(C1)- ARG58(C2)	5.17 Å	-	-	-	-	-	-	-	4.50 Å	-
ASP50(C2)- ARG58(C1)	-	5.29 Å	-	-	-	-	-	-	-	4.62 Å
ASP50- ARG58	-	-	-	-	-	-	-	9.12 Å	-	-
GLU103- ARG479	8.81 Å	10.46 Å	-	-	-	-	-	-	-	-
GLU240(C2)- LYS433	-	-	-	-	-	-	-	-	-	6.72 Å
GLU307- ARG311	4.23 Å	4.24 Å	4.29 Å	4.21 Å	4.24 Å	4.25 Å	4.23 Å	4.30 Å	-	-
GLU307- LYS395	3.62 Å	3.60 Å	3.61 Å	3.60 Å	3.61 Å	3.60 Å	3.62 Å	3.61 Å	-	-
LYS393- ASP430	3.14 Å	3.67 Å	2.90 Å	2.89 Å	5.83 Å	3.31 Å	3.05 Å	2.93 Å	-	5.98 Å
LYS433- GLU451	6.88 Å	9.28 Å	5.80 Å	4.62 Å	7.60 Å	8.26 Å	-	6.62 Å	-	-

\*: In the last 5 ns, the average distance becomes approximately 3.20 Å.

Results in Table 4.45 show that if there are differences between the chains of the mutant dimer at acidic pH its inactivity mechanism can be like the one of C300F mutant. In this case the angle change between the chains of the E307K+T391S mutant dimer can be explained by the differences between the possible salt bridges of its chains have. Additional salt bridge between Glu240 in chain1 and Lys433 in chain2 of E307K+T391S mutant dimer also enhances this assumption. On the other hand, in the simulation on W318C dimer at acidic pH, possible salt bridges do not show significant differences. The additional possible salt bridges consist of residue pairs that only get close to each other for short periods of time, thus the average distances are long in Table 4.46. We hypothesize that W318C causes misfolding due to the mutation's close position to Arg19, while E307K+T391S causes a hindrance in octamerization in the light of our results.

#### 4.10. Cross Correlation Calculations on Monomer Forms at Neutral pH

Cross correlation analysis was done for each structure to determine any intramolecular relationship between residues, in other words, to see whether the specific regions on the secondary structure moving in a correlated manner. A TCL script for VMD and R program were used to calculate the correlation matrices and to generate the cross correlation images. The codes are given in Appendix D. Correlated residues with a higher value than the threshold value, 0.5, were connected with red lines as shown in Figure 4.131 for each structure.

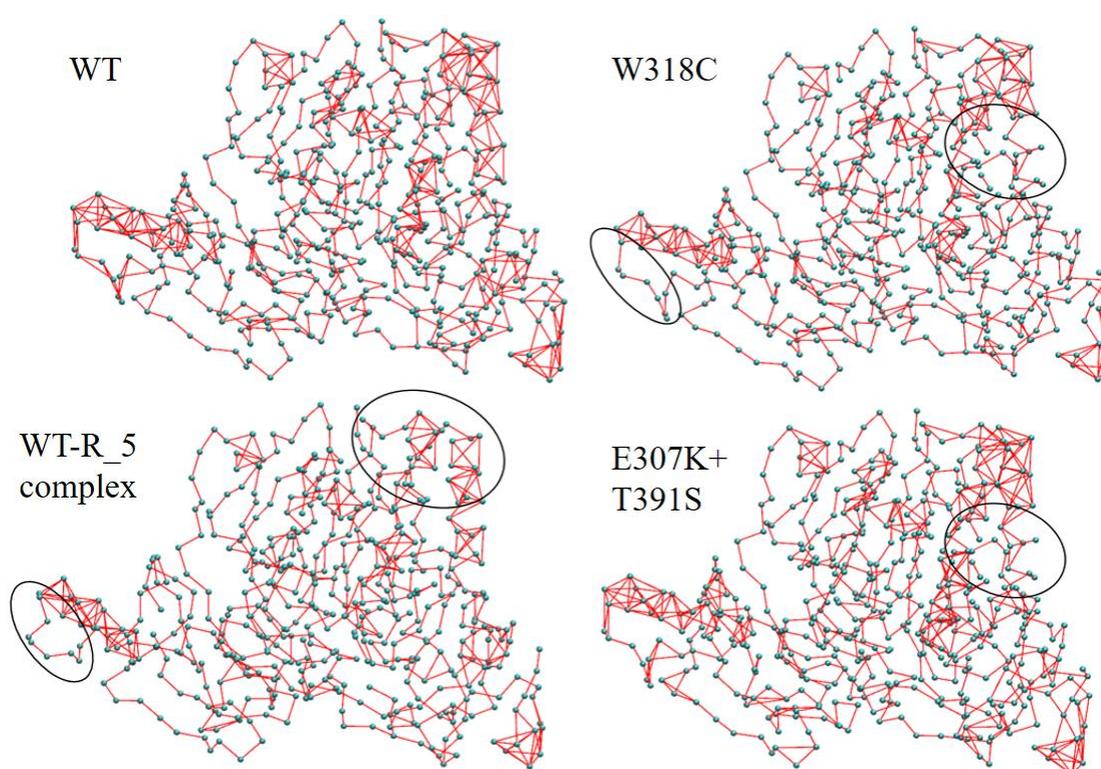


Figure 4.131. Cross correlation maps of the four structures.

In Figure 4.131, residues correlated with a score higher than the threshold, 0.5, were connected with red lines. Black circles show the regions with lower correlation coefficients than those in the WT structure. Highly correlated residues follow the backbone pattern of the structure as expected. Distant regions were not found to be correlated in all the

simulations, but some regions lost their communication in the mutant and ligand docked structures.

Figure 4.131 shows that mutants lose some correlation between residues located on the long alpha helix on the right of the protein. The circled regions on the left of the ligand docked and W318C mutant structures are in the dimerization3 loop region. Dimerization3 loop RMSD profile of E307K+T391S mutant being correlated may not imply being stable since the region shown in Figure 4.9 has more correlation lines than ligand docked and W318C mutant structures. On the other hand, even binding a molecule makes active site or overall structure more stable, the correlation can be lost as shown in Figure 4.131. Ligand docked structure has less correlated residues around the C-terminus of the structure which is the region on the right bottom corner of the structures.

To understand the regions better the correlation matrices of the three monomer forms are subtracted from the WT's at neutral pH. The correlation matrices and differences of these matrices are given in following figures from Figure 4.132 to Figure 4.134 . After subtracting the off-diagonal cells of the resulting matrices with values between 0.5 /-0.5 and 1/-1 were colored with red and color coded circled.

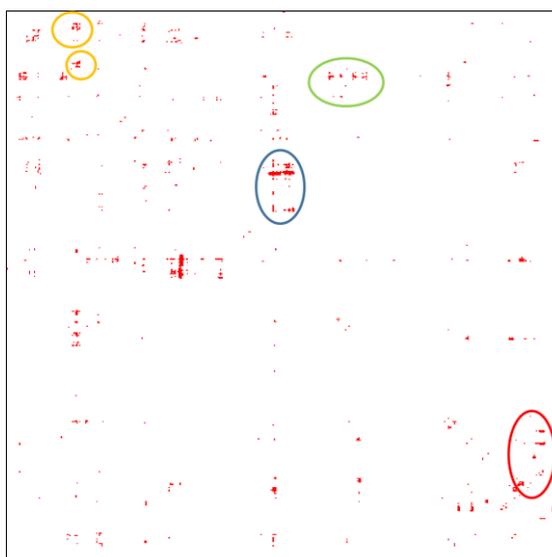


Figure 4.132. Difference between the cross correlation matrix of E307K+T391S mutant and that of WT.

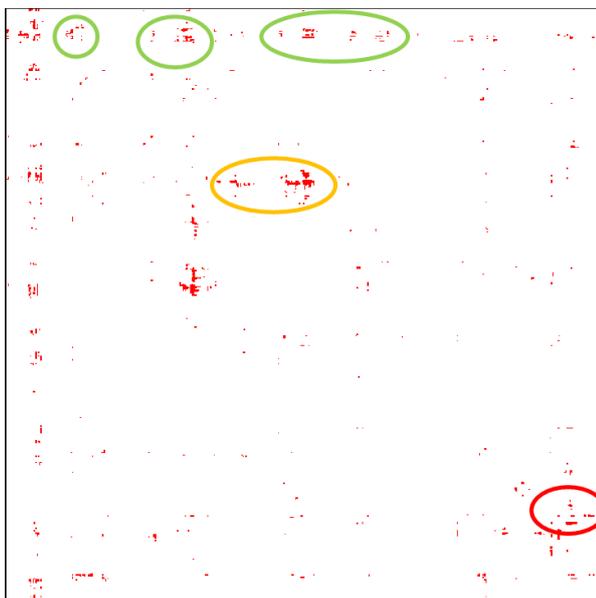


Figure 4.133. Difference between the cross correlation matrix of W318C mutant and that of WT.

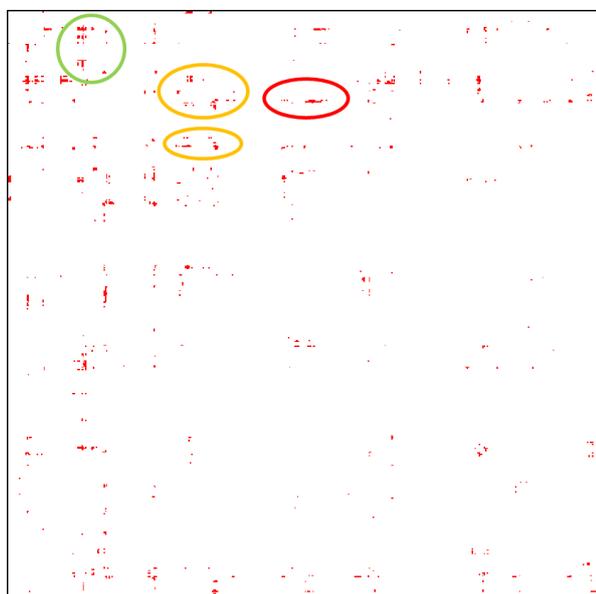


Figure 4.134. Difference between the cross correlation matrix of R\_5 bound WT and that of WT.

Difference matrices show that mutations cause higher differences in the communication of the residues which are located near the edges/surface of the protein like Cys161 and Leu261. The circled regions were mapped on the secondary structures of the monomers and shown as follows in Figure 4.135, Figure 4.136 and Figure 4.137.

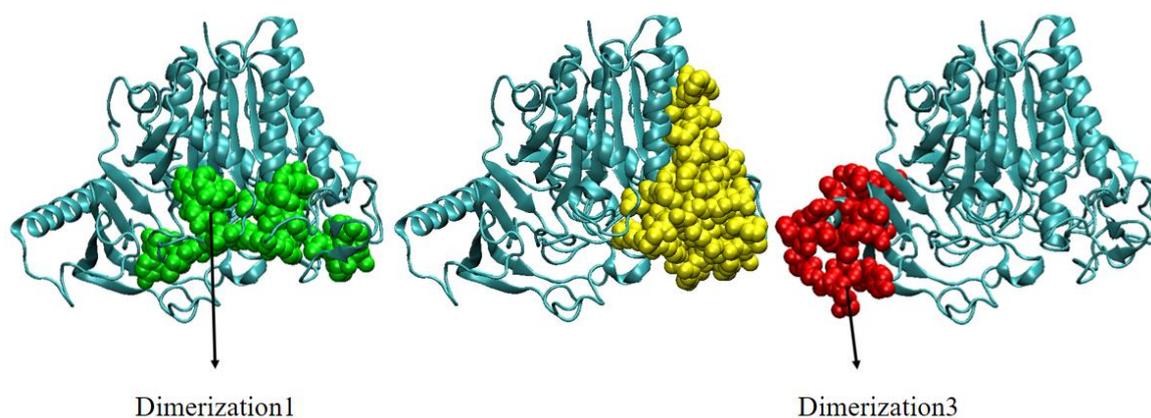


Figure 4.135. The regions that were circled in the difference matrix of W318C mutant monomer at neutral pH.

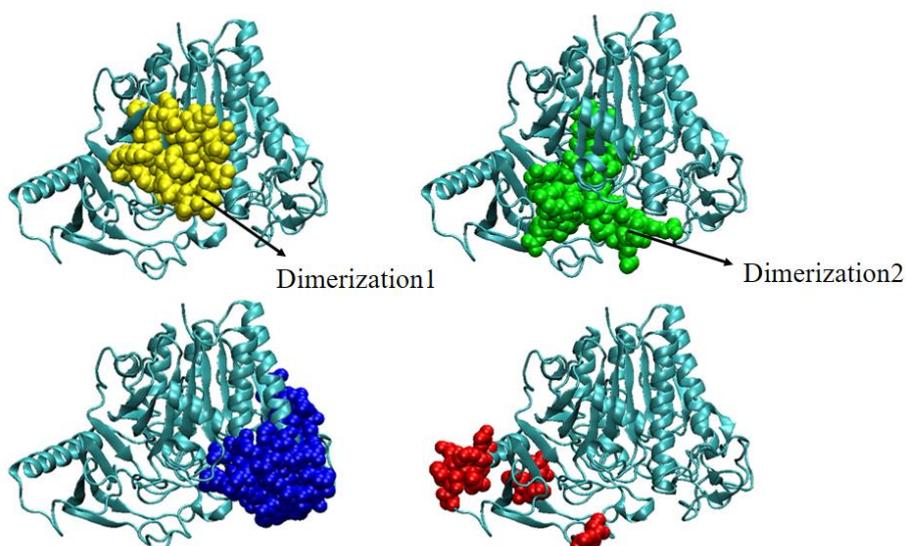


Figure 4.136. The regions that were circled in the difference matrix of E307K+T391S mutant monomer at neutral pH.

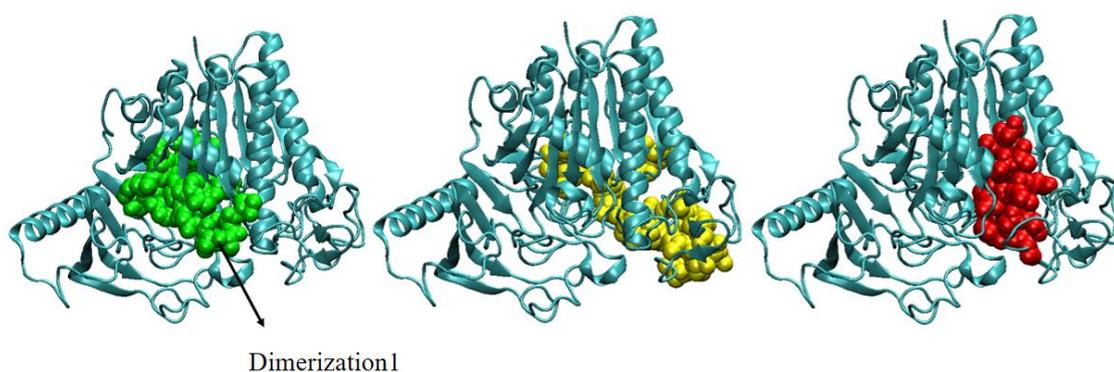


Figure 4.137. The regions that were circled in the difference matrix of R\_5 bound WT monomer at neutral pH.

The results of the cross correlation calculations on monomeric ARSA forms show that the correlation change between the residues of the monomers are seen mostly on the bottom part of the enzyme that is rich in loops. This region is probably affected by the changes on the structure and may be used as a hint to link the mutations effects on the enzyme dynamics and structure.

## 5. CONCLUSIONS AND RECOMMENDATIONS

### 5.1. Conclusions

The aim of this thesis was to find the effects of the mutations (W318C and E307K+T391S) on ARSA dynamics and structure and consequently to find a PC candidate for these ARSA mutants.

MD simulations on monomer ARSA forms at neutral and acidic milieu showed that mutations affected the dimerization region of the monomer structures, but the overall stability of both dimer and monomer forms is not affected by the mutations. Mutations do not affect the active site structure and do not trigger the escape of  $Mg^{2+}$  cation in the simulations. Also, they do not affect the stability of the Asp335-Arg370 region which protects the main glycosylation site from the stress of lysosomal environment. The RMSD profiles of the mutants were stable, suggesting structural stability, and the residue based RMSF profiles of the mutant forms were similar to that of WT, suggesting no major changes in dynamics upon mutation. The T274M and P377L mutants are known to fold incorrectly [4], but these mutants show no significant changes in structure or dynamics either. These results suggest that these mutant, if folded, can retain their WT conformation. The dimeric structure is also preserved at both pH values in the mutants despite instability in the dimerization regions. Dimerization3 region whose coordinates are missing in the crystal structure, is confirmed to be very flexible and in all dimer simulations, including those on the WT and mutants, but does not have a significant effect on dimerization of ARSA. When the dimerization regions are analyzed, and the trajectories are observed in detail, it can be concluded that the changes on the dynamics are important on these three regions in a decreasing order of importance on the structural stability as: Dimerization1, dimerization2, dimerization3. We suggest that the deviations that were observed in the dimerization1 region in monomeric forms is carried to the octamerization helix when ARSA is in its dimeric form, thus the effect of the mutation can be predicted by comparison.

The catalytic FGL69 residue can be hydrated or in diol form. When the MD simulations on monomer forms at acidic pH are compared, the hydration state of FGL does

not affect the stability of the WT and mutant monomers. Therefore, we continued with the nonhydrated form of ARSA for the MD simulations on dimer forms at acidic milieu. Moreover, for the hydration of ARSA, it should first form the stable dimeric or octameric form at acidic pH. Comparison of the average structure and the last frame coordinates of the dimer forms showed that there is a significant change in the association angle of E307K+T391S mutant. When the possible salt bridges of the simulations are examined, another remarkable difference showed up that the residue pair for the possible hydrogen bonds is different between the chains of C300F and E307K+T391S dimers at acidic pH. Additionally, there is an extra salt bridge in the simulation on E307K+T391S dimer which enhances the possibility of the association angle change between the chains of the dimer. C300F is known to be lacking octamerization due to the changed association angle between its monomers [61,62]. W318C mutant dynamics did not show any significant differences in terms of salt bridges or RMSF calculations, Trp318 is close to the first forming epitope (Arg19) and the signal peptide of ARSA. Our hypothesis is that the significant electrostatic change on this region may affect the folding of ARSA during its synthesis. Trp318 is also close to and T274M causes severe misfolding, suggesting that W318C mutation may cause severe misfolding.

The docking calculations show that two possible molecules can be proposed as PC candidates for ARSA: R\_5 and ZINC90709065. The stability of dimerization regions after binding of these two molecules changes and the dynamics of the dimerization regions become more rigid. These two molecules also preserve their places during the simulations.

## 5.2. Recommendations

ARSA monomer's crystal structure consists of 485 amino acids and it was deposited into PDB as a correctly folded mature form. On this form, it is hard to observe the effects of an only missense mutation. The structure and dynamics of the protein were not affected by the mutations and comparison with the other mutants whose mechanism of inactivation is known guided our conclusion. Performing folding studies on ARSA mutant instead of working on mature crystal structure can provide further information.

The binding of the natural ligand can be highly dependent on the water molecules in the active site since we know that they play crucial role in the reaction of ARSA [1]. Docking calculations can be repeated in the presence of active site water molecules. The results should be compared with the known ligand bound crystal structure.

Moreover, the water molecules around the protein and especially the ones near the oligomerization interfaces should be examined carefully. When we observed the trajectories, we saw that some water bridges are formed in the dimer interface. It would be better to focus more on the dynamics of some water molecules that are important at the specific regions. Adding these water molecules into the hydrogen bond occupancy calculations can provide information on their roles.

Longer simulations would provide better understanding about enzyme dynamics and the effects of the mutations on the protein structure and dynamics. Simulations can be extended from 50 ns to 1  $\mu$ s to observe the equilibrium dynamics of the proteins better.

## REFERENCES

1. G. Lukatela, N. Krauss, K. Theis, T. Selmer, V. Gieseimann, K. Von Figura, and W. Saenger, "Crystal structure of human arylsulfatase A: The aldehyde function and the metal ion at the active site suggest a novel mechanism for sulfate ester hydrolysis," *Biochemistry*, vol. 37, no. 97, pp. 3654–3664, 1998.
2. S. N. Shah, K. Minn, and F. Schoenfeld, "Arylsulfatase A and / 3-Galactosidase Activities in Leukocytes and Lymphocytes from Normal and Psychiatric Subjects," *Molecular and Chemical Neuropathology*, vol. 24, pp. 43–52, 1995.
3. M. Schenk, C. a K. Koppisetty, D. C. Santos, E. Carmona, S. Bhatia, P.-G. Nyholm, and N. Tanphaichitr, "Interaction of arylsulfatase-A (ASA) with its natural sulfoglycolipid substrates: a computational and site-directed mutagenesis study.," *Glycoconjugate Journal*, vol. 26, no. 8, pp. 1029–45, 2009.
4. P. Poeppel, M. M. Abouzied, C. Völker, and V. Gieselmann, "Misfolded endoplasmic reticulum retained subunits cause degradation of wild-type subunits of arylsulfatase A heteromers," *Federation of European Biochemical Societies Journal*, vol. 277, no. 16, pp. 3404–3414, 2010.
5. R. von Bülow, B. Schmidt, T. Dierks, N. Schwabauer, K. Schilling, E. Weber, I. Usón, and K. von Figura, "Defective oligomerization of arylsulfatase a as a cause of its instability in lysosomes and metachromatic leukodystrophy.," *The Journal of Biological Chemistry*, vol. 277, no. 11, pp. 9455–61, 2002.
6. M. T. Porter, a L. Fluharty, and H. Kihara, "Metachromatic leukodystrophy: arylsulfatase-A deficiency in skin fibroblast cultures.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 62, no. 3, pp. 887–891, 1969.
7. P. Poeppel, M. Habetha, A. Marcão, H. Büssow, L. Berna, and V. Gieselmann, "Missense mutations as a cause of metachromatic leukodystrophy: Degradation of

- arylsulfatase A in the endoplasmic reticulum,” *Federation of European Biochemical Societies Journal*, vol. 272, pp. 1179–1188, 2005.
8. A. Özkan and H. A. Özkara, “Metachromatic leukodystrophy: Biochemical characterization of two (p.307Glu→Lys, p.318Trp→Cys) arylsulfatase A mutations,” *Intractable & Rare Diseases Research*, vol. 5, no. 4, pp. 280–283, 2016.
  9. N. V Olkhovich, N. Takamura, N. a Pichkur, N. G. Gorovenko, K. Aoyagi, and S. Yamashita, “Novel mutations in arylsulfatase A gene in three Ukrainian families with metachromatic leukodystrophy.,” *Molecular Genetics and Metabolism*, vol. 80, no. 3, pp. 360–3, 2003.
  10. E. F. Neufeld, “Lysosomal Storage Diseases,” *Annual Review of Biochemistry*, vol. 60, 1991.
  11. V. Gieselmann, J. Zlotogora, a Harris, D. a Wenger, and C. P. Morris, “Molecular genetics of metachromatic leukodystrophy.,” *Human Mutation*, vol. 4, no. 4, pp. 233–42, 1994.
  12. Rauch, D. Hartmann, V. P. M. Saravanan, P. P. De Deyn, R. D’Hooge, A. M. Van Der Linden, and N. Schaeren-Wiemers, “Metachromatic leukodystrophy: consequences of sulphatide accumulation.,” *Acta Paediatrica Supplement*, vol. 92, p. 74–79; discussion 45, 2003.
  13. V. Gieselmann and I. Krägeloh-Mann, “Metachromatic leukodystrophy an update,” *Neuropediatrics*, vol. 41, no. 1, pp. 1–6, 2010.
  14. V. Gieselmann, a Polten, J. Kreysing, and K. von Figura, “Arylsulfatase A pseudodeficiency: loss of a polyadenylation signal and N-glycosylation site.,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 86, no. 23, pp. 9436–9440, 1989.
  15. V. Gieselmann, a L. Fluharty, T. Tønnesen, and K. Von Figura, “Mutations in the

- arylsulfatase A pseudodeficiency allele causing metachromatic leukodystrophy.," *American Journal of Human Genetics*, vol. 49, no. 2, pp. 407–13, 1991.
16. B. Eng, L. N. Nakamura, N. O'Reilly, N. Schokman, M. M. J. Nowaczyk, W. Krivit, and J. S. Wayne, "Identification of nine novel arylsulfatase a (ARSA) gene mutations in patients with metachromatic leukodystrophy (MLD)," *Human Mutation*, vol. 22, no. 5, pp. 418–419, 2003.
  17. S. Kafert, U. Heinisch, J. Zlotogora, and V. Gieselmann, "A missense mutation P136L in the arylsulfatase A gene causes instability and loss of activity of the mutant enzyme.," *Hum. Genet.*, vol. 95, no. 2, pp. 201–4, 1995.
  18. Ikeda, A. and Moore, T., *Metachromatic Leukodystrophy Treatment & Management: Medical Care, Consultations*. 2017, Emedicine.medscape.com. Available at: <http://emedicine.medscape.com/article/951840-treatment> [Accessed 10 Jul. 2017].
  19. U. Matzner and V. Gieselmann, "Gene therapy of metachromatic leukodystrophy.," *Expert Opinion on Biological Therapy*, vol. 5, no. 1, pp. 55–65, 2005.
  20. A. Biffi, G. Lucchini, A. Rovelli, and M. Sessa, "Metachromatic leukodystrophy: an overview of current and prospective treatments," *Bone Marrow Transplant.*, vol. 42, pp. S2–S6, 2008.
  21. O. N. Koç, J. Day, M. Nieder, S. L. Gerson, H. M. Lazarus, and W. Krivit, "Allogeneic mesenchymal stem cell infusion for treatment of metachromatic leukodystrophy (MLD) and Hurler syndrome (MPS-IH)," *Bone Marrow Transplant.*, vol. 30, no. 4, pp. 215–222, 2002.
  22. N. Miyake, K. Miyake, S. Karlsson, and T. Shimada, "Successful treatment of metachromatic leukodystrophy using bone marrow transplantation of HoxB4 overexpressing cells," *The American Society of Gene & Cell Therapy*, vol. 18, no. 7, pp. 1373–1378, 2010.

23. K. von Figura, F. Steckel, and a Hasilik, “Juvenile and adult metachromatic leukodystrophy: partial restoration of arylsulfatase A (cerebroside sulfatase) activity by inhibitors of thiol proteinases.,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 80, no. 19, pp. 6066–6070, 1983.
24. A. Mühlstein, S. Gelperina, E. Shipulo, O. Maksimenko, and J. Kreuter, “Arylsulfatase a bound to poly(butyl cyanoacrylate) nanoparticles for enzyme replacement therapy - Physicochemical evaluation,” *Pharmazie*, vol. 69, no. 7, pp. 518–524, 2014.
25. A. Bockenhoff, S. Cramer, P. Wolte, S. Knieling, C. Wohlenberg, V. Gieselmann, H.-J. Galla, and U. Matzner, “Comparison of Five Peptide Vectors for Improved Brain Delivery of the Lysosomal Enzyme Arylsulfatase A,” *Journal of Neuroscience*, vol. 34, no. 9, pp. 3122–3129, 2014.
26. H. Kitano, “Computational systems biology,” *Nature*, vol. 420, pp. 206–210, November, 2002.
27. D. E. Berman, D. Ringe, G. A. Petsko, and S. A. Small, “The Use of Pharmacological Retromer Chaperones in Alzheimer's Disease and other Endosomal-related Disorders,” *Neurotherapeutics*, vol. 12, no. 1, pp. 12–18, 2015.
28. R. L. Lieberman, J. A. D’Aquino, D. Ringe, and G. A. Petsko, “Effects of pH and iminosugar pharmacological chaperones on lysosomal glycosidase structure and stability,” *Biochemistry*, vol. 48, no. 22, pp. 4816–4827, 2009.
29. A. N. Gupta, K. Neupane, N. Rezajoei, L. M. Cortez, V. L. Sim, and M. T. Woodside, “Pharmacological chaperone reshapes the energy landscape for folding and aggregation of the prion protein.,” *Nature Communications*, vol. 7, no. 3, p. 12058, 2016.
30. V. Bernier, M. Lagacé, D. G. Bichet, and M. Bouvier, “Pharmacological chaperones: Potential treatment for conformational diseases,” *Trends in Endocrinology and Metabolism*, vol. 15, no. 5, pp. 222–228, 2004.

31. J. P. Morello, M. Bouvier, U. E. Petäjä-Repo, and D. G. Bichet, "Pharmacological chaperones: A new twist on receptor folding," *Trends in Pharmacological Sciences*, vol. 21, no. 12, pp. 466–469, 2000.
32. T. Arakawa, D. Ejima, Y. Kita, and K. Tsumoto, "Small molecule pharmacological chaperones: From thermodynamic stabilization to pharmaceutical drugs," *Biochimica et Biophysica Acta - Proteins and Proteomics*, vol. 1764, no. 11, pp. 1677–1687, 2006.
33. D. Ringe and G. a Petsko, "What are pharmacological chaperones and why are they interesting?," *Journal of Biology*, vol. 8, no. 9, p. 80, 2009.
34. A. Banning, C. Gülec, J. Rouvinen, S. J. Gray, and R. Tikkanen, "Identification of Small Molecule Compounds for Pharmacological Chaperone Therapy of Aspartylglucosaminuria," *Scientific Reports*, vol. 6, no. August, p. 37583, 2016.
35. G. Parenti, G. Andria, and A. Ballabio, "Lysosomal Storage Diseases: From Pathophysiology to Therapy," *Annual Review of Medicine*, vol. 66, no. 1, pp. 471–486, 2015.
36. E. Onder, I. Sinici, F. Müjgan Sönmez, M. Topçu, and H. A. Ozkara, "Identification of two novel arylsulfatase A mutations with a polymorphism as a cause of metachromatic leukodystrophy.," *Neurol. Res.*, vol. 31, no. 1, pp. 60–6, 2009.
37. C. R. Gault, L. M. Obeid, and Y. A. Hannun, "An overview of sphingolipid metabolism: From synthesis to breakdown," *Advances in Experimental Medicine and Biology*, vol. 688, pp. 1–23, 2010.
38. F. M. Platt, "Sphingolipid lysosomal storage disorders.," *Nature*, vol. 510, no. 7503, pp. 68–75, 2014.
39. V. Albinet, M.-L. Bats, C. Bedia, F. Sabourdy, V. Garcia, B. Ségui, N. Andrieu-Abadie, T. Hornemann, and T. Levade, "Sphingolipids: Basic Science and Drug Development," *Handbook of Experimental Pharmacology*, vol. 215, no. 215, pp. 127–52, 2013.

40. T. Kolter, "A view on sphingolipids and disease," *Chemistry and Physics of Lipids*, vol. 164, no. 6, pp. 590–606, 2011.
41. Y. Kacher and A. H. Futerman, "Genetic diseases of sphingolipid metabolism: Pathological mechanisms and therapeutic options," *Federation of European Biochemical Societies Letters*, vol. 580, no. 23, pp. 5510–5517, 2006.
42. T. Kolter and K. Sandhoff, "Sphingolipid metabolism diseases," *Biochimica et Biophysica Acta - Biomembranes*, vol. 1758, no. 12, pp. 2057–2079, 2006.
43. M. L. Barth, a Fensom, and a Harris, "Identification of seven novel mutations associated with metachromatic leukodystrophy.," *Human Mutations*, vol. 6, no. 2, pp. 170–6, 1995.
44. L. Stevens, F. T. Neuropsychiatric, and S. Hospital, "Purification and Properties A from Human Urine \* of Arylsulfatase," *The Journal of Biological Chemistry*, vol. 260, no. 7, pp. 2495-2501, 1975.
45. R. von Bülow, B. Schmidt, T. Dierks, K. von Figura, and I. Usón, "Crystal structure of an enzyme-substrate complex provides insight into the interaction between human arylsulfatase A and its substrates during catalysis.," *Journal of Molecular Biology*, vol. 305, no. 2, pp. 269–77, 2001.
46. A. Waldow, B. Schmidt, T. Dierks, R. Von Bülow, and K. Von Figura, "Amino Acid Residues Forming the Active Site of Arylsulfatase A," *Journal of Biological Chemistry*, vol. 274, no. 18, pp. 12284–12288, 1999.
47. J. O'Brien and Y. Kishimoto, "Saposin proteins: structure, function, and role in human lysosomal storage disorders.," *Federation of American Societies for Experimental Biology Journal*, vol. 5, no. 1, pp. 301–308, 1991.
48. A. Polten, A. Fluthary, C. Fluthary et al., "Molecular basis of different forms of Metachromatic Leukodystrophy," *The New England Journal of Medicine*, vol. 324, no.

- 1, pp. 18-22, 1991.
49. H. J. Sommerlade, T. Selmer, A. Ingendoh, V. Gieselmann, K. Von Figura, K. Neifer, and B. Schmidt, "Glycosylation and phosphorylation of arylsulfatase A," *Journal of Biological Chemistry*, vol. 269, no. 33, pp. 20977–20981, 1994.
50. M. Chruszcz, P. Laidler, M. Monkiewicz, E. Ortlund, L. Lebioda, and K. Lewinski, "Crystal structure of a covalent intermediate of endogenous human arylsulfatase A," *Journal of Inorganic Biochemistry*, vol. 96, no. 2–3, pp. 386–392, 2003.
51. P. Vagedes, W. Saenger, and E.-W. Knapp, "Driving forces of protein association: the dimer-octamer equilibrium in arylsulfatase A," *Biophysics Journal*, vol. 83, no. 12, pp. 3066–3078, 2002.
52. R. R. Abzalimov, C. E. Bobst, P. a Salinas, P. Savickas, J. J. Thomas, and I. a Kaltashov, "Studies of pH-dependent self-association of a recombinant form of arylsulfatase A with electrospray ionization mass spectrometry and size-exclusion chromatography.," *Analytical Chemistry*, vol. 85, no. 3, pp. 1591–6, 2013.
53. A Knaust, B. Schmidt, T. Dierks, R. von Bülow, and K. von Figura, "Residues critical for formylglycine formation and/or catalytic activity of arylsulfatase A.," *Biochemistry*, vol. 37, no. 40, pp. 13941–6, 1998.
54. T. Dierks, B. Schmidt, and K. von Figura, "Conversion of cysteine to formylglycine: a protein modification in the endoplasmic reticulum.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 94, no. 11, pp. 11963–11968, 1997.
55. M. Recksiek, T. Selmer, T. Dierks, B. Schmidt, and K. Von Figura, "Sulfatases, trapping of the sulfated enzyme intermediate by substituting the active site formylglycine," *Journal of Biological Chemistry*, vol. 273, no. 11, pp. 6096–6103, 1998.
56. M. J. Appel and C. R. Bertozzi, "Formylglycine, a post-translationally generated residue

with unique catalytic capabilities and biotechnology applications,” *American Chemical Society Chemical Biology*, vol. 10, no. 1, pp. 72–84, 2015.

57. D. Roeser, A. Preusser-Kunze, B. Schmidt, K. Gasow, J. G. Wittmann, T. Dierks, K. von Figura, and M. G. Rudolph, “A general binding mechanism for all human sulfatases by the formylglycine-generating enzyme.,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, no. 1, pp. 81–86, 2006.
58. R. B. Von Low, B. Schmidt, T. Dierks, N. Schwabauer, K. Schilling, E. Weber, I. Usón, and K. Von Figura, “Defective oligomerization of arylsulfatase A as a cause of its instability in lysosomes and metachromatic leukodystrophy,” *Journal of Biological Chemistry*, vol. 277, no. 11, pp. 9455–9461, 2002.
59. G. von Heijne, “The Signal peptide,” *Journal of Membrane Biology*, vol. 115, pp. 195–201, 1990.
60. J. S. Bonifacino and L. M. Traub, “Signals for sorting of transmembrane proteins to endosomes and lysosomes.,” *Annual Review of Biochemistry*, vol. 72, pp. 395–447, 2003.
61. A. Marcão, J. E. Azevedo, V. Gieselmann, and M. C. Sá Miranda, “Oligomerization capacity of two arylsulfatase A mutants: C300F and P425T,” *Biochemical and Biophysical Research Communications*, vol. 306, pp. 293–297, 2003.
62. A. Marcão, O. Amaral, E. Pinto, R. Pinto, and M. C. S. Miranda, “Metachromatic Leucodystrophy in Portugal — Finding of Four New Molecular Lesions : C300F ,” *Human Mutation*, vol. 232, no. 1, pp. 1-6, 1999.
63. P. Luzi, M. A. Rafi, H. Z. Rao, and D. A. Wenger, “Sixteen novel mutations in the arylsulfatase A gene causing metachromatic leukodystrophy,” *Gene*, vol. 530, no. 2, pp. 323–328, 2013.
64. R. L. Stevens, a L. Fluharty, H. Kihara, M. M. Kaback, L. J. Shapiro, B. Marsh, K.

- Sandhoff, and G. Fischer, "Cerebroside sulfatase activator deficiency induced metachromatic leukodystrophy.," *American Journal of Human Genetics*, vol. 33, no. 6, pp. 900–906, 1981.
65. R. Draghia, F. Letourneur, C. Drugan, J. Manicom, C. Blanchot, A. Kahn, L. Poenaru, and C. Caillaud, "Metachromatic Leukodystrophy : Identification of the First Deletion in Exon I and of Nine Novel Point Mutations in the Arylsulfatase A Gene," *Human Mutation*, vol. 242, no. July 1996, pp. 234–242, 1997.
66. M. Adenot and R. Lahana, "Blood-Brain Barrier Permeation Models: Discriminating between Potential CNS and Non-CNS Drugs Including P-Glycoprotein Substrates," *Journal of Chemical Information and Computer Sciences*, vol. 44, no. 1, pp. 239–248, 2004.
67. D. D. Jhala, S. S. Chettiar, and J. K. Singh, "Optimization and Validation of an In Vitro Blood Brain Barrier Permeability Assay Using Artificial Lipid Membrane," *Journal of Bioequivalence & Bioavailability*, no. 14, pp. 10–14, 2012.
68. U. Manual, *BioLuminate 1.9.*, Schrödinger, LLC, New York, NY, 2015.
69. J. Schymkowitz, J. Borg, F. Stricher, R. Nys, F. Rousseau, and L. Serrano, "The FoldX web server: An online force field," *Nucleic Acids Research*, vol. 33, no. 1. 2, pp. 382–388, 2005.
70. P. P. Guide, "Protein Preparation Guide," *Molecules*, vol. 2007, no. 8, pp. 428–432, 2011.
71. U. Manual, *LigPrep 2.3*, Schrödinger, LLC, New York, NY, 2015.
72. U. Manual, *Maestro 10.2*, Schrödinger, LLC, New York, NY, 2015.
73. L. Schrödinger, *Glide 5.5. User Manual*, Schrödinger, LLC, New York, 2009.

74. P. Fomby and A. J. Cherlin, "Assessing the performance of the MM/PBSA and MM/GBSA methods: II. The accuracy of ranking poses generated from docking," *National Institutes of Health Public Access*, vol. 72, no. 2, pp. 181–204, 2011.
75. Humphrey, W., Dalke, A. and Schulten, K., "VMD - Visual Molecular Dynamics" *Journal of Molecular Graphics*, vol. 14, no. 1, pp. 33-38, 1996
76. Schrödinger Suite 2009 *Virtual Screening Workflow; Glide version 5.5*, Schrödinger, LLC, New York, NY, 2006; *LigPrep version 2.3*, Schrödinger, LLC, New York, NY, 2006; *QikProp version 3.2*, Schrödinger, LLC, New York, NY, 2006.
77. J. W. Ponder and D. A. Case, "Force fields for protein simulations," *Advances in Protein Chemistry*, vol. 66, pp. 27–85, 2003.
78. A. J. Clark, P. Tiwary, K. Borrelli, S. Feng, E. B. Miller, R. Abel, R. A. Friesner, and B. J. Berne, "Prediction of Protein-Ligand Binding Poses via a Combination of Induced Fit Docking and Metadynamics Simulations," *Journal of Chemical Theory and Computation*, vol. 12, no. 6, pp. 2990–2998, 2016.
79. L. Schrödinger, *Canvas 1.5*, Schrödinger, LLC, New York, NY, 2015.
80. D. Bajusz, A. Rácz, and K. Héberger, "Why is Tanimoto index an appropriate choice for fingerprint-based similarity calculations?," *Journal of Cheminformatics*, vol. 7, no. 1, pp. 1–13, 2015.
81. C. G. Wermuth, C. R. Ganellin, P. Lindberg, and L. a. Mitscher, "Glossary of Terms Used in Medicinal Chemistry," *Pure and Applied Chemistry*, vol. 70, no. 5, pp. 1129–1143, 1998.
82. A. Gaulton, L. J. Bellis, A. P. Bento, J. Chambers, M. Davies, A. Hersey, Y. Light, S. McGlinchey, D. Michalovich, B. Al-Lazikani, and J. P. Overington, "ChEMBL: A large-scale bioactivity database for drug discovery," *Nucleic Acids Research*, vol. 40, no. D1, pp. 1100–1107, 2012.

83. A. Gaulton, A. Hersey, M. L. Nowotka, A. Patricia Bento, J. Chambers, D. Mendez, P. Mutowo, F. Atkinson, L. J. Bellis, E. Cibrian-Uhalte, M. Davies, N. Dedman, A. Karlsson, M. P. Magarinos, J. P. Overington, G. Papadatos, I. Smit, and A. R. Leach, "The ChEMBL database in 2017," *Nucleic Acids Research*, vol. 45, no. D1, pp. D945–D954, 2017.
84. J. Psychiatry, "Relationship of Neuroleptic Drug Effects at Brain Dopamine, Serotonin, alpha-Adrenergic, and Histamine Receptors to Clinical Potency," *The American Journal of Psychiatry*, vol. 137, no. 12, pp. 1518–1522, 1980.
85. J. M. Thomas, M. W. Grinstaff, and H. D. Lutz, "The Design of Leadlike Combinatorial Libraries," *Angewandte Chemie International Edition*, vol. 3773, no. 24, pp. 3743–3748, 2016.
86. D. Leung, C. Hardouin, D. L. Boger, and B. F. Cravatt, "Discovering potent and selective reversible inhibitors of enzymes in complex proteomes," *Nature Biotechnology*, vol. 21, no. 6, pp. 687–91, 2003.
87. P. P. Roy, J. T. Leonard, and K. Roy, "Exploring the impact of size of training sets for the development of predictive QSAR models," *Chemometrics and Intelligent Laboratory Systems*, vol. 90, no. 1, pp. 31–42, 2008.
88. S. L. Dixon, A. M. Smondjrev, E. H. Knoll, S. N. Rao, D. E. Shaw, and R. A. Friesner, "PHASE: A new engine for pharmacophore perception, 3D QSAR model development, and 3D database screening: 1. Methodology and preliminary results," *Journal of Computer-Aided Molecular Design*, vol. 20, no. 10–11, pp. 647–671, 2006.
89. C. Abad-Zapatero, "Ligand efficiency indices for effective drug discovery," *Expert Opinion on Drug Discovery*, vol. 2, no. 4, pp. 469–488, 2007.
90. C. Abad-Zapatero and J. T. Metz, "Ligand efficiency indices as guideposts for drug discovery," *Drug Discovery Today*, vol. 10, no. 7, pp. 464–469, 2005.

91. A. L. Hopkins, G. M. Keserú, P. D. Leeson, D. C. Rees, and C. H. Reynolds, “The Role of Ligand Efficiency Measures in Drug Discovery,” *Nature Reviews*, vol. 13, no. 2, pp. 105–121, 2014.
92. D. E. Shaw, *Desmond User’s Guide Desmond Version 3.0*, Version 0.5.3, New York, NY, 2011.
93. L. Schrödinger, *Prime*, Version 2.1, Schrödinger, LLC, New York, NY, 2009.
94. P. H. Hünenberger, A. E. Mark, and W. F. van Gunsteren, “Fluctuation and cross-correlation analysis of protein motions observed in nanosecond molecular dynamics simulations,” *Journal of Molecular Biology*, vol. 252, no. 4, pp. 492–503, 1995.
95. K. Kasahara, I. Fukuda, and H. Nakamura, “A novel approach of dynamic cross correlation analysis on molecular dynamics simulations and its application to Ets1 dimer-DNA complex,” *PLoS One*, vol. 9, no. 11, 2014.
96. N. W. Venables, M. D. Smith, and C. T. R., "An Introduction to R: Version 3.4.0," vol. 1, pp. 1–15, 2017.
97. G. Madhavi Sastry, M. Adzhigirey, T. Day, R. Annabhimoju, and W. Sherman, “Protein and ligand preparation: Parameters, protocols, and influence on virtual screening enrichments,” *Journal of Computer-Aided Molecular Design*, vol. 27, no. 3, pp. 221–234, 2013.
98. Humphrey, W., Dalke, A. and Schulten, K., "VMD - Visual Molecular Dynamics", *Journal of Molecular Graphics*, 1996, vol. 14, pp. 33-38..
99. S. Genheden and U. Ryde, “Comparison of end-point continuum-solvation methods for the calculation of protein-ligand binding free energies,” *Proteins: Structure, Function and Bioinformatics*, vol. 80, no. 5, pp. 1326–1342, 2012.
100. D. C. Rapaport, “Molecular dynamics simulation,” *Computing in Science*

*Engineering*, vol. 1, no. 1, pp. 537–542, 1999.

101. Repo, H., *Structural Studies on Lysosomal Proteins Research*, Ph.D., University of Helsinki, 2014.
102. M. Y. F. Virgens, L. Pol-Fachin, H. Verli, and M. L. Saraiva-Pereira, “Effects of glycosylation and pH conditions in the dynamics of human arylsulfatase A.,” *Journal of Biomolecular Structure & Dynamics*, no. 4, pp. 1–13, 2013.
103. P. Laidler and A. Litynska, “Arylsulfatase A from human placenta possesses only high mannose-type glycans.,” *The International Journal of Biochemistry & Cell Biology*, vol. 29, no. 3, pp. 475–483, 1997.
104. C. H. Hill, A. H. Viuff, S. J. Spratley, S. Salamone, S. H. Christensen, R. J. Read, N. W. Moriarty, H. H. Jensen, and J. E. Deane, “Azasugar inhibitors as pharmacological chaperones for Krabbe disease,” *Chemical Science*, vol. 6, no. 3, pp. 3075–3086, 2015.
105. J. M. Luco, “Prediction of the brain-blood distribution of a large set of drugs from structurally derived descriptors using partial least-squares (PLS) modeling.,” *Journal of Chemical Information and Computer Sciences*, vol. 39, no. 2, pp. 396–404, 1999.
106. S. Vilar, M. Chakrabarti, and S. Costanzi, “Prediction of passive blood-brain partitioning: Straightforward and effective classification models based on in silico derived physicochemical descriptors,” *Journal of Molecular Graphics and Modelling*, vol. 28, no. 8, pp. 899–903, 2010.
107. Y. H. Zhao, M. H. Abraham, A. Ibrahim, P. V. Fish, S. Cole, M. L. Lewis, M. J. De Groot, and D. P. Reynolds, “Predicting penetration across the blood-brain barrier from simple descriptors and fragmentation schemes,” *Journal of Chemical Information and Modeling*, vol. 47, no. 1, pp. 170–175, 2007.
108. D. A. Thaisrivongs, S. P. Miller, C. Molinaro, Q. Chen, Z. J. Song, L. Tan, L. Chen, W. Chen, A. Lekhal, S. K. Pulicare, and Y. Xu, “Synthesis of Verubecestat, a BACE1

- Inhibitor for the Treatment of Alzheimer's Disease," *Organic Letters*, vol. 18, no. 22, pp. 5780–5783, 2016.
109. A. Delgado, J. Casas, A. Llebaria, J. L. Abad, and G. Fabrias, "Inhibitors of sphingolipid metabolism enzymes," *Biochimica et Biophysica Acta - Biomembranes*, vol. 1758, no. 12, pp. 1957–1977, 2006.
  110. S. L. Planey and R. Kumar, "Lipophilicity Indices for Drug Development," *Journal of Applied Biopharmaceutics and Pharmacokinetics*, vol. 1, no. 1, pp. 31–36, 2013.
  111. A. L. Hopkins, G. M. Keserü, P. D. Leeson, D. C. Rees, and C. H. Reynolds, "The role of ligand efficiency metrics in drug discovery.," *Nature reviews. Drug discovery*, vol. 13, no. 2, pp. 105–21, 2014.
  112. C. Arenz, "Cellular Physiology and Biochemistr y Biochemistry Small Molecule Inhibitors of Acid Sphingomyeli- nase," *Cellular Physiology and Biochemistry Review*, vol. 26, pp. 1–8, 2010.
  113. A. G. Roth, D. Drescher, Y. Yang, S. Redmer, S. Uhlig, and C. Arenz, "Potent and selective inhibition of acid sphingomyelinase by bisphosphonates.," *Angewandte Chemie (International Edition in English)*., vol. 48, no. 41, pp. 7560–7563, 2009.
  114. B. E. Smid, J. M. F. G. Aerts, R. G. Boot, G. E. Linthorst, and C. E. M. Hollak, "Pharmacological small molecules for the treatment of lysosomal storage disorders," *Expert Opinion on Investigational Drugs*, vol. 19, no. 11, pp. 1367–1379, 2010.
  115. I. Bendikov-Bar, G. Maor, M. Filocamo, and M. Horowitz, "Ambroxol as a pharmacological chaperone for mutant glucocerebrosidase," *Blood Cells, Molecules, and Diseases*, vol. 50, no. 2, pp. 141–145, 2013.

## APPENDIX A: ANALYSIS OF MOLECULAR DYNAMICS SIMULATIONS

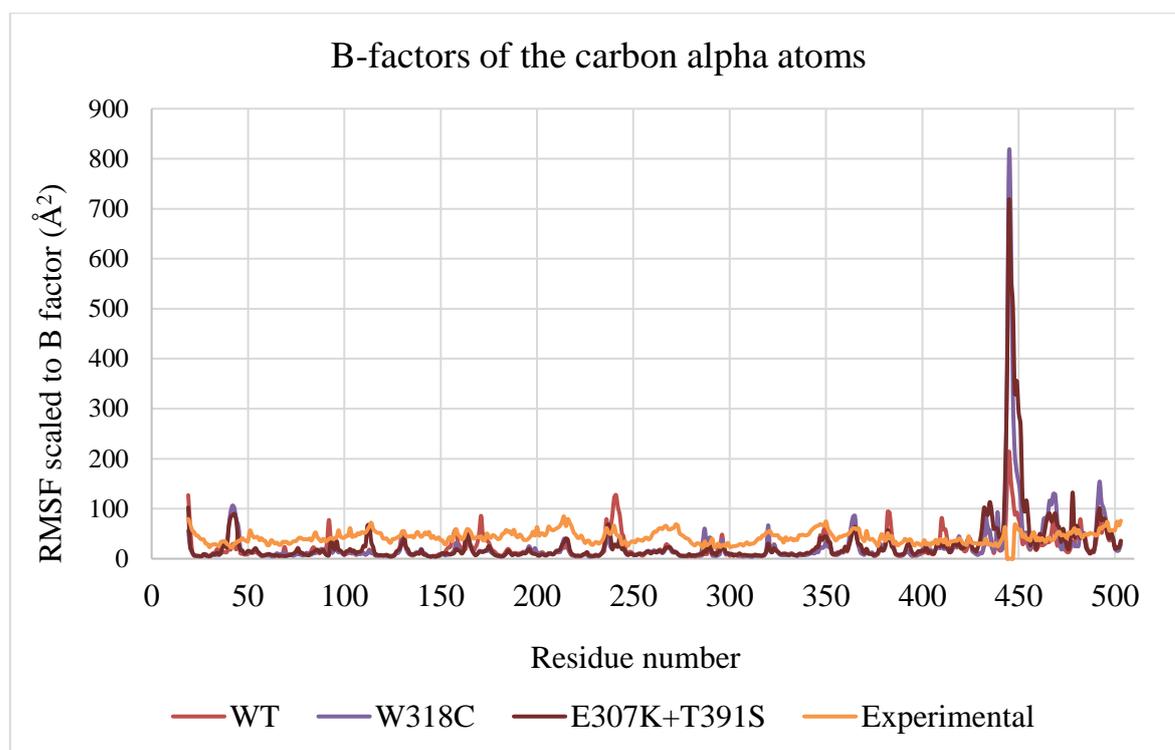


Figure A.1. RMSF results of the carbon alpha atom of the all residues of WT, W318C mutant, E307K+T391S mutant forms, and the experimental values from PBD are given for the whole trajectories. The calculated results were scaled to B factor, and the experimental B factor values are also shown.

Table A.1. Hydrogen bond occupancies of the selected residues are given in percentage for all monomer forms of ARSA at neutral pH.

WT			W318C		
residue pair		%	residue pair		%
ARG19	ASP216	116.46%	ARG19	ASP216	90.77%
ARG19	GLU273	77.80%	ARG19	GLU273	4.00%
ARG19	ASP216	12.31%	ARG19	GLU273	2.38%
ASP29	HIE227	65.44%	ASP29	SPC1601	25.89%
ASP29	SPC12087	26.71%	ASP29	HIE227	22.50%
ASP29	SPC120	11.19%	ASP29	SPC8759	12.57%
ARG73	ASP29	55.12%	ARG73	ASP29	73.62%
ARG73	LYN123	27.22%	ARG73	ASP281	68.89%
ARG73	FGL69	25.93%	ARG73	SPC8759	51.61%
ARG73	ASP281	7.97%	ARG73	UNK1	51.02%
ARG73	TRP124	6.13%	ARG73	FGL69	42.91%
ASP169	ARG291	25.70%	ASP169	TYR230	31.36%
ASP169	TYR230	6.23%	ASP169	SPC2459	22.55%
ASP169	TYR230	2.42%	ASP169	SPC2459	22.09%
GLY171	ASP152	0.65%	GLY171	CYS156	32.97%
ASP173	GLY154	0.98%	ASP173	ASP152	11.83%
GLU188	LEU181	62.27%	GLU188	LEU181	112.71%
GLU188	ARG200	16.36%	GLU188	ARG200	2.64%
GLU240	ASP249	47.37%	GLU240	ASP249	75.30%
ARG241	SER237	6.84%	ARG241	SER237	32.69%
TRP318	GLU272	34.79%	CYS318	SPC1979	67.50%
TRP318	ILE322	31.39%	CYS318	SPC12407	18.51%
TRP318	SPC1193	20.17%	CYS318	SPC7463	10.66%
GLU329	SER368	47.19%	GLU329	SER368	22.68%
GLU329	SER368	13.68%	GLU329	SPC10979	6.01%
LYS395	GLU307	162.01%	LYS395	GLU307	152.71%
LYS395	GLU437	95.11%	LYS395	GLU437	134.09%

Table A.1. Hydrogen bond occupancies of the selected residues are given in percentage for all monomer forms of ARSA at neutral pH (cont.).

WT			W318C		
residue pair		%	residue pair		%
LYS395	TYR429	66.77%	LYS395	TYR429	93.81%
LYS395	TYR429	35.23%	LYS395	THR305	23.60%
LYS395	THR305	32.63%	LYS395	TYR429	8.41%
PHE398	PHE387	92.59%	PHE398	PHE387	49.60%
PHE398	HIS423	47.44%	PHE398	HIS423	21.55%
GLU307	GLU437	74.38%	GLU307	GLU437	70.76%
E307K+T391S			WT-R_5		
residue pair		%	residue pair		%
ARG19	GLU273	98.81%	ARG19	GLU273	66.79%
ARG19	ASP216	58.98%	ARG19	GLU273	18.94%
ARG19	SPC1315	1.44%	ARG19	ASP216	15.67%
ASP29	HIE227	41.52%	ASP29	HIE227	105.91%
ASP29	SPC14608	27.12%	SPC29	GLU329	83.83%
ASP29	LYN123	16.46%	SPC29	THR327	67.34%
ARG73	LYN123	72.47%	ARG73	ASP281	37.78%
ARG73	ASP29	47.89%	ARG73	FGL69	29.09%
ARG73	SER225	31.57%	ARG73	LYS123	9.01%
ARG73	FGL69	17.46%	ARG73	ALA224	7.14%
ARG73	TRP124	11.05%	ARG73	TRP124	4.87%
ASP169	TYR230	34.25%	ASP169	ARG291	11.42%
ASP169	SPC6016	5.97%	ASP169	TYR230	1.89%
ASP169	SPC11197	4.80%	ASP169	SPC8043	0.67%
GLY171	CYS156	33.16%	GLY171	SPC771	36.38%
ASP173	ASP152	33.43%	ASP173	LEU176	3.05%
GLU188	LEU181	97.12%	GLU188	LEU181	61.18%
GLU188	ARG200	2.78%	GLU188	ARG200	18.91%
GLU240	ASP249	69.17%	GLU240	ASP249	60.65%

Table A.1. Hydrogen bond occupancies of the selected residues are given in percentage for all monomer forms of ARSA at neutral pH (cont.).

ARG241	SER237	22.55%	ARG241	SER237	8.48%
TRP318	GLU272	31.47%	TRP318	GLU272	33.52%
TRP318	SPC3924	13.47%	TRP318	ILE322	23.30%
TRP318	ILE322	7.35%	TRP318	SPC2369	15.91%
GLU329	SER368	19.84%	GLU329	SER368	71.13%
GLU329	SPC12724	7.95%	GLU329	SPC29	24.51%
LYS395	GLU437	107.50%	LYS395	GLU307	146.99%
LYS395	TYR429	67.34%	LYS395	GLU437	108.67%
LYS395	SPC11160	19.84%	LYS395	TYR429	81.46%
LYS395	THR305	14.90%	LYS395	TYR429	36.88%
LYS395	TYR429	7.28%	LYS395	THR305	36.72%
PHE398	PHE387	41.06%	PHE398	PHE387	62.04%
PHE398	HIS423	29.01%	PHE398	HIS423	49.15%
LYN307	GLU437	32.74%	GLU307	GLU437	41.44%
W318C-R_5			T274M		
residue pair		%	residue pair		%
ARG19	ASP216	83.59%	ARG19	GLU273	116.00%
ARG19	GLU273	22.72%	ARG19	ASP216	60.96%
ARG19	SPC12615	2.61%	ARG19	GLU272	1.60%
ASP29	SPC10730	50.20%	ASP29	HIE227	49.02%
SPC29	LYN123	45.16%	ASP29	SPC11133	22.42%
SPC29	HIE227	19.11%	ASP29	SPC1853	16.94%
ARG73	ASP281	118.02%	ARG73	ASP281	110.27%
ARG73	TRP124	76.55%	ARG73	SPC14690	65.55%
ARG73	SER150	33.81%	ARG73	FGL69	65.10%
ARG73	ASP29	25.28%	ARG73	ASP29	38.48%
ARG73	FGL69	12.41%	ARG73	TYR222	13.51%

Table A.1. Hydrogen bond occupancies of the selected residues are given in percentage for all monomer forms of ARSA at neutral pH (cont.).

W318C-R_5			T274M		
residue pair		%	residue pair		%
ASP169	SPC6818	18.02%	ASP169	TYR230	57.84%
ASP169	ARG291	13.30%	ASP169	SPC7031	6.49%
ASP169	SPC6818	10.82%	ASP169	SPC11088	5.71%
GLY171	CYS156	24.21%	GLY171	CYS156	41.21%
ASP173	ASP152	25.95%	ASP173	ASP152	2.53%
GLU188	LEU181	98.83%	GLU188	LEU181	117.71%
GLU188	ARG200	5.96%	GLU188	ARG200	0.88%
GLU240	ASP249	84.59%	GLU240	ASP249	73.84%
ARG241	SER237	21.21%	ARG241	SER237	37.88%
CYS318	SPC13416	20.22%	TRP318	GLU272	47.19%
CYS318	SPC12023	19.66%	TRP318	SPC10223	40.26%
CYS318	SPC3364	18.31%	TRP318	SPC2782	39.20%
GLU329	SER368	18.78%	GLU329	SER368	33.91%
GLU329	SPC12169	1.77%	GLU329	SPC13885	9.81%
LYS395	GLU307	139.75%	LYS395	GLU307	157.96%
LYS395	GLU437	112.80%	LYS395	GLU437	114.96%
LYS395	TYR429	78.49%	LYS395	TYR429	91.37%
LYS395	THR305	29.98%	LYS395	THR305	25.30%
LYS395	TYR429	20.82%	LYS395	TYR429	5.24%
PHE398	PHE387	45.88%	PHE398	PHE387	58.00%
PHE398	HIS423	22.82%	PHE398	HIS423	27.60%
GLU307	GLU437	124.46%	GLU307	GLU437	115.28%

Table A.1. Hydrogen bond occupancies of the selected residues are given in percentage for all monomer forms of ARSA at neutral pH (cont.).

W318C-ZINC			P377L		
residue pair		%	residue pair		%
GLU240	ASP249	52.65%	GLU240	ASP249	74.00%
ARG241	SER237	18.25%	ARG241	SER237	39.74%
CYS318	SPC7452	24.27%	TRP318	SPC2388	45.74%
CYS318	SPC663	13.97%	TRP318	SPC2204	27.23%
CYS318	SPC9269	13.86%	TRP318	GLU272	26.37%
GLU329	SER368	45.04%	SER368	GLU329	30.04%
GLU329	SPC146	1.71%	GLU329	SPC12977	7.08%
LYS395	GLU307	155.23%	LYS395	GLU307	160.88%
LYS395	GLU437	133.49%	LYS395	GLU437	107.34%
LYS395	TYR429	90.75%	LYS395	TYR429	85.34%
LYS395	THR305	33.26%	LYS395	THR305	19.15%
LYS395	TYR429	8.67%	LYS395	TYR429	11.82%
PHE398	PHE387	37.38%	PHE398	PHE387	47.34%
PHE398	HIS423	22.50%	PHE398	HIS423	27.17%
GLU307	GLU437	58.98%	GLU307	GLU437	85.44%

Table A.2. All possible salt bridges for ARSA monomers at neutral pH.

WT	W318C	E307K+T391S	T274M	P377L
ASP169-ARG291	ASP61-ARG390	ASP169-ARG291	ASP430-LYS433	ASP61-ARG311
ASP61-ARG311	GLU329-LYS367	GLU329-LYS367	ASP434-LYS433	ASP216-ARG19
GLU307-ARG311	ASP169-ARG291	ASP430-LYS433	GLU307-LYS395	GLU307-ARG311
ASP61-ARG390	ASP61-ARG311	ASP29-HIS229	ASP211-ARG143	ASP211-ARG143
ASP411-ARG299	GLU307-LYS395	ASP411-ARG299	ASP61-HIS328	ASP467-LYS463
GLU285-ARG288	GLU307-ARG311	GLU285-ARG288	ASP216-ARG217	ASP216-ARG217
ASP211-ARG143	ASP467-LYS463	ASP61-ARG390	ASP216-ARG19	GLU285-ARG288
ASP482-ARG97	ASP335-ARG370	ASP354-ARG84	ASP411-ARG299	ASP482-ARG97
ASP354-ARG84	ASP411-ARG299	ASP211-ARG143	ASP354-ARG84	ASP430-LYS393
ASP381-ARG384	GLU285-ARG288	ASP467-LYS463	GLU451-LYS393	GLU451-LYS433
GLU285-HIS405	ASP381-ARG384	GLU273-ARG19	ASP61-ARG390	GLU273-ARG19
ASP207-ARG143	ASP354-ARG84	ASP281-ARG73	ASP467-LYS463	ASP354-ARG84
ASP281-ARG73	ASP216-ARG19	ASP207-ARG143	GLU273-ARG19	ASP281-ARG73
GLU437-LYS395	ASP211-ARG143	ASP430-LYS393	GLU285-ARG288	ASP207-ARG143
ASP30-LYS302	ASP482-ARG97	ASP30-LYS302	ASP207-ARG143	ASP211-ARG214
ASP30-HIS229	GLU437-LYS395	GLU285-HIS405	ASP482-ARG97	ASP30-LYS302
ASP29-ARG73	ASP281-ARG73	ASP30-HIS229	GLU285-HIS405	GLU437-LYS395
ASP207-ARG214	GLU285-HIS405	ASP29-ARG73	ASP211-ARG214	ASP30-HIS229
ASP211-ARG214	ASP207-ARG143	ASP152-ARG288	ASP430-LYS393	ASP29-ARG73
ASP430-LYS433	ASP211-ARG214	ASP61-ARG311	ASP30-HIS229	GLU188-ARG200
ASP35-ARG244	ASP30-LYS302	GLU437-LYS395	ASP29-ARG73	ASP207-ARG214
GLU104-ARG80	ASP30-HIS229	GLU104-ARG80	GLU188-ARG200	ASP430-LYS433
GLU188-ARG200	ASP29-ARG73	GLU188-ARG200	ASP207-ARG214	ASP152-ARG288
GLU307-LYS395	GLU188-ARG200	ASP35-ARG244	GLU424-LYS463	GLU131-ARG97
ASP335-ARG370	ASP430-LYS433	GLU451-LYS393	ASP169-ARG291	ASP35-ARG244
GLU329-LYS367	GLU131-ARG97	ASP335-ARG370	GLU307-ARG311	GLU104-ARG80
GLU451-LYS393	ASP35-ARG244	ASP381-ARG384	ASP35-ARG244	ASP335-ARG370
ASP467-LYS463	GLU104-ARG80	ASP207-HIS206	GLU104-ARG80	GLU451-LYS393
ASP207-HIS206	GLU451-LYS393	ASP482-ARG97	GLU382-ARG384	ASP61-ARG390
GLU285-LYS302	ASP430-LYS393	GLU285-LYS302	ASP61-ARG311	ASP411-ARG299
ASP434-LYS433	GLU424-LYS463	ASP211-ARG214	GLU307-ARG390	GLU329-LYS367
ASP430-LYS393	GLU285-LYS302	ASP207-ARG214	GLU272-ARG19	GLU307-LYS395
GLU131-ARG496	ASP434-LYS433		ASP335-ARG370	ASP381-ARG384
	ASP207-ARG214		GLU329-LYS367	GLU424-LYS463
			ASP207-HIS206	GLU285-LYS302
			GLU285-LYS302	GLU285-HIS405
			ASP381-ARG384	ASP61-LYS393
			ASP281-ARG73	ASP169-ARG291

Table A.2. All possible salt bridges for ARSA monomers at neutral pH (cont.).

WT	W318C	E307K+T391S	T274M	P377L
			ASP30-LYS302	GLU437-ARG299
			GLU437-LYS395	
			GLU272-HIS321	

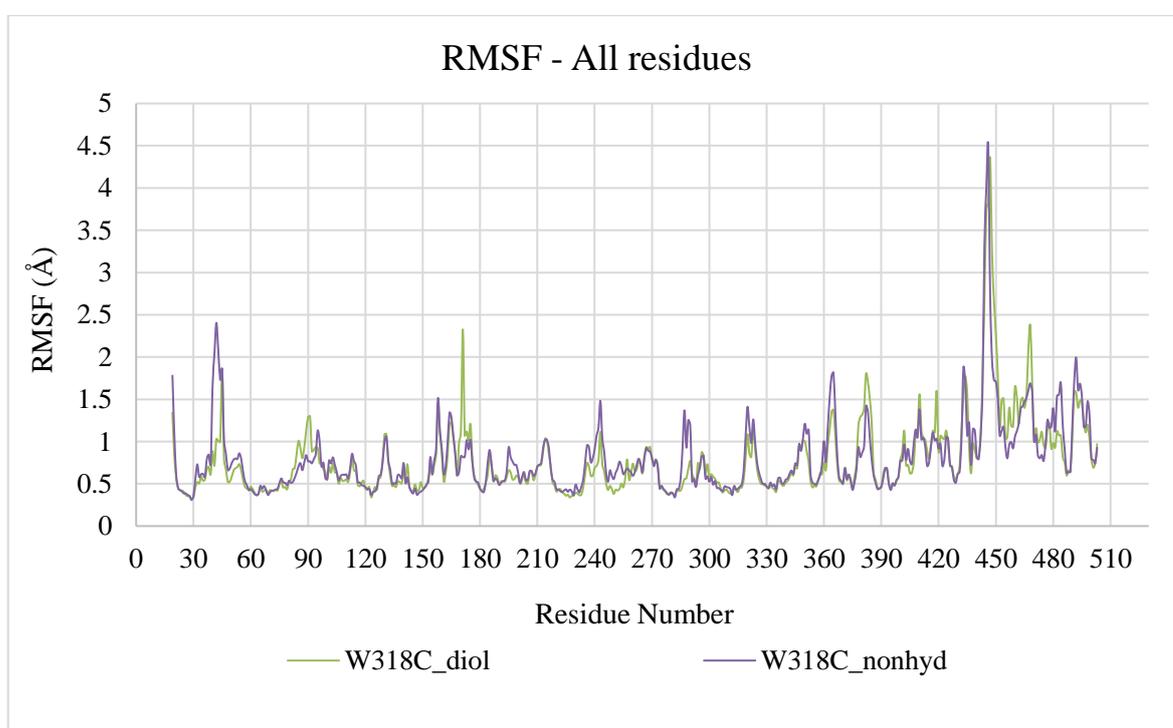


Figure A.2. Comparison of the RMSF results of both forms of W318C mutant ARSA at acidic pH.

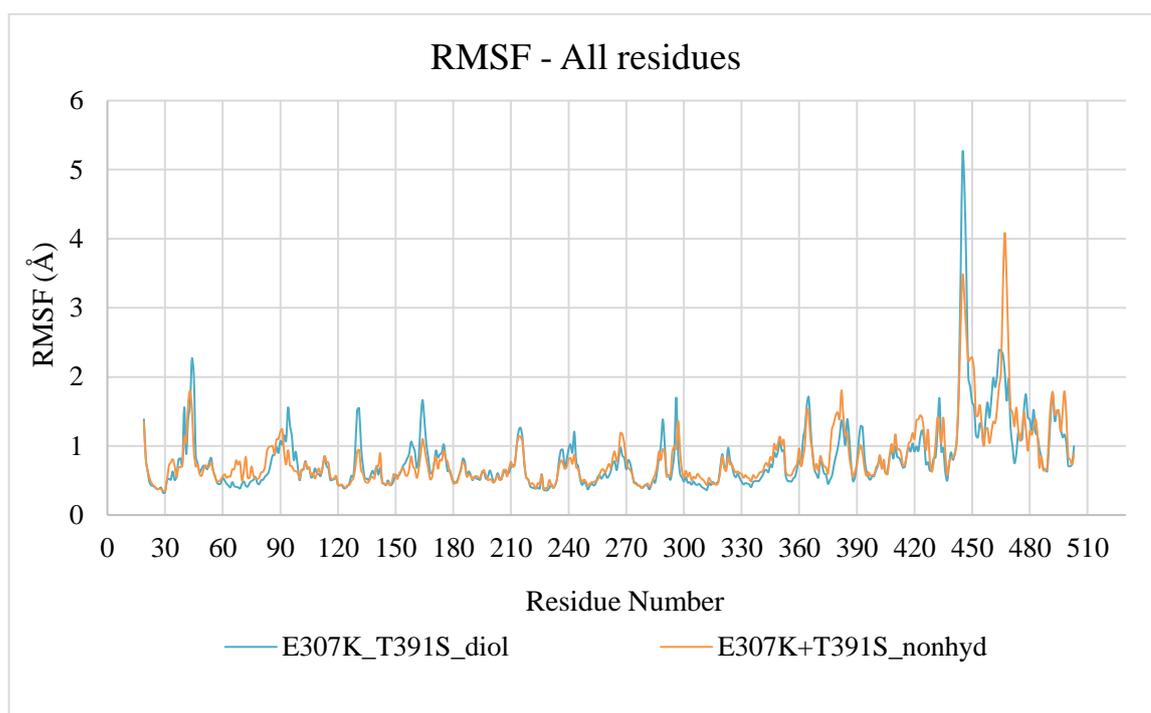


Figure A.3. Comparison of the RMSF results of both forms of E307K+T391S mutant ARSA at acidic pH.

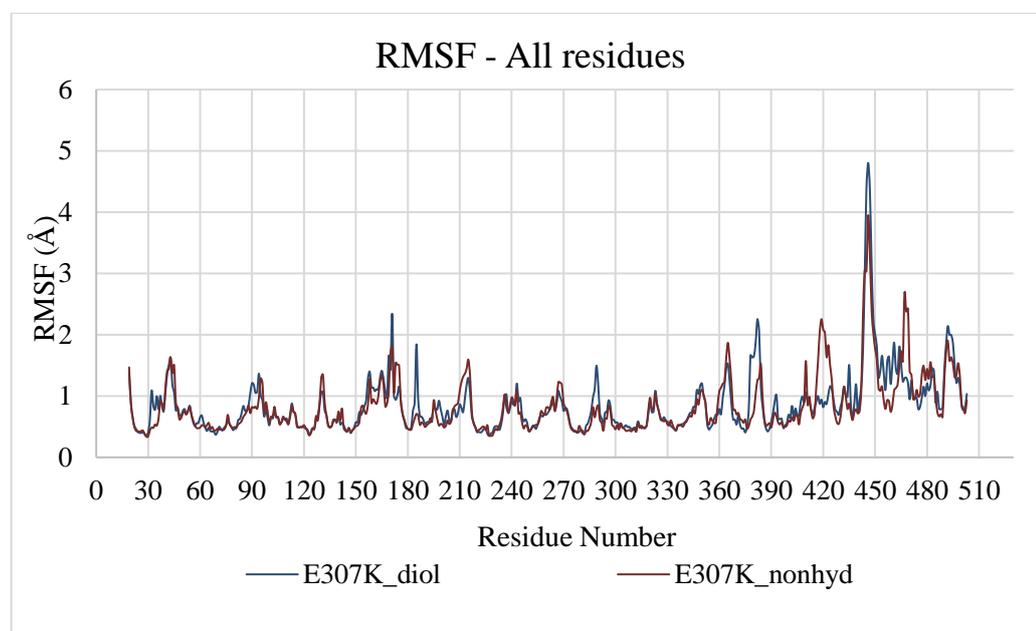


Figure A.4. Comparison of the RMSF results of both forms of E307K mutant ARSA at acidic pH

Table A.3. Hydrogen bond occupancies between the selected residues are given for the simulations at acidic pH.

WT						W318C					
diol		%	nonhyd		%	diol		%	nonhyd		%
ARG19	GLU273	90.3	ARG19	GLU273	113.1	ARG19	ASP216	65.2	ARG19	ASP216	50.3
ARG19	ASP216	31.4	ARG19	GLU272	2.1	ARG19	ASP216	12.3	ARG19	GLU273	48.8
ARG19	ASP216	1.5	ARG19	Water	1.3	ARG19	GLU273	6.6	ARG19	GLU273	14.8
ARG19	Water	1.2	ARG19	ASP216	1.1	ARG19	GLN213	6.4	ARG19	Water	4.0
ARG19	Water	1.0	ARG19	Water	1.1	ARG19	Water	3.8	ARG19	Water	3.1
ASP29	HIP227	70.7	ASP29	HIP227	87.9	ARG19	Water	2.0	ARG19	ASN22	1.9
LYS123	ASP29	60.1	LYS123	ASP29	59.2	ASP29	HIP227	72.6	ASP29	HIE227	64.0
Water	ASP29	10.2	Water	ASP29	1.3	LYS123	ASP29	71.5	LYS123	ASP29	58.4
ARG73	ASP29	5.1	ASP50	THR46	1.8	ARG73	ASP29	2.4	ARG73	ASP29	3.7
ASP50	THR46	1.6	ARG390	ASP61	74.2	ASP50	THR46	2.3	ASP50	THR46	20.2
ALA54	ASP50	1.0	THR60	ASP61	9.2	ALA54	ASP50	1.2	Water	ASP50	2.9
Water	ASP50	1.0	ALA331	ASP61	4.7	ARG390	ASP61	75.4	ALA54	ASP50	2.3
ARG390	ASP61	70.2	ARG311	ASP61	3.9	THR60	ASP61	9.2	THR46	ASP50	1.9
THR60	ASP61	11.6	Water	ASP61	2.6	ALA331	ASP61	5.7	ARG390	ASP61	73.7
ALA331	ASP61	5.3	Water	ASP61	2.3	ARG311	ASP61	4.1	THR60	ASP61	9.1
ARG311	ASP61	4.4	Water	ASP61	1.6	Water	ASP61	1.1	ALA331	ASP61	5.1
ARG73	ASP281	48.3	Water	ASP61	1.6	ARG73	ASP281	89.8	ARG311	ASP61	4.5
ARG73	Water	37.5	Water	ASP61	1.5	ARG73	Water	45.0	ARG73	ASP281	100.9
ARG73	ASP29	24.8	ARG73	ASP281	92.2	ARG73	SER225	27.5	ARG73	SER225	64.5
ARG73	LYS123	21.2	ARG73	SER225	36.0	ARG73	ASP29	18.9	ARG73	FGL69	20.2
Water	ARG73	14.9	ARG73	FGL69	16.8	Water	ARG73	15.2	ARG73	ASP29	15.1
ARG73	SER225	10.0	ARG73	ALA224	6.9	ARG73	Water	6.8	ARG73	LYS123	12.3
ARG73	FGL69	7.3	ARG73	ASP29	3.0	ARG73	ALA28	4.6	ARG73	ALA224	2.6
ARG73	ALA224	6.6	ARG73	LYS123	1.3	ARG73	ALA224	2.7	ARG97	ASP482	96.5
ARG73	ALA28	6.5	ARG73	TRP124	1.2	ARG73	LYS123	2.2	ARG97	GLY480	24.9
ARG97	ASP482	88.5	ARG97	ASP482	82.9	ARG73	FGL69	1.8	ARG97	ARG479	16.9
ARG97	GLY480	42.0	ARG97	GLY480	11.9	ARG97	ASP482	74.2	ARG97	Water	4.9
ARG97	Water	6.1	ARG97	GLY480	11.2	ARG97	GLY480	54.4	ARG97	Water	4.2
ARG97	ARG479	5.5	ARG97	ARG479	7.5	ARG97	ARG479	11.6	ARG97	Water	3.7

Table A.3. Hydrogen bond occupancies between the selected residues are given for the simulations at acidic pH (cont.).

WT					W318C						
diol		%	diol		%	diol		%	diol		%
ARG97	GLY480	4.2	ARG97	Water	3.8	ARG97	GLU131	8.0	ARG97	Water	2.8
ARG97	GLU131	4.1	ARG97	Water	3.2	ARG97	Water	5.5	ARG97	Water	2.7
ARG97	Water	4.0	ARG97	Water	3.2	ARG97	Water	2.4	ARG97	Water	2.6
ARG97	Water	3.8	ARG97	Water	2.1	ARG97	Water	2.4	ARG97	GLU131	2.6
ARG97	Water	3.4	ASP173	ASH152	2.0	ASH173	ASH152	19.4	ASH152	Water	7.6
ARG97	ARG479	3.0	Water	ASP169	11.1	ASH152	GLY171	4.1	ASH152	Water	7.2
ARG97	Water	2.5	ASP169	Water	7.7	ASH152	Water	2.2	ASH152	SER150	6.3
No Hbond for ASN152	No Hbond for ASN152	0.0	Water	ASP169	5.3	Water	ASH152	1.8	ASH152	Water	4.8
ASP169	Water	38.1	ASP169	Water	4.8	ASH152	Water	1.7	ASH152	Water	4.5
Water	ASP169	28.6	Water	ASP169	2.2	ASH152	Water	1.6	ASH152	Water	3.9
TYR230	ASP169	23.2	TYR230	ASP169	2.2	ASH152	Water	1.3	ASH152	Water	3.5
Water	ASP169	11.9	Water	ASP169	2.0	TYR230	ASP169	30.6	ASH152	Water	3.1
Water	ASP169	8.7	ARG143	ASP207	28.0	Water	ASP169	17.1	TYR230	ASP169	24.1
Water	ASP169	4.9	ASP207	ALA203	13.5	Water	ASP169	6.1	TYR230	ASP169	22.8
ARG288	ASP169	3.7	Water	ASP207	1.6	Water	ASP169	6.0	Water	ASP169	14.4
Water	GLU104	3.6	HIS229	TYR230	98.3	ASP169	Water	5.7	ASP169	Water	12.7
Water	GLN139	3.6	Water	HIS229	7.7	ASP169	Water	5.4	Water	ASP169	6.7
Water	ASP169	2.2	HIS229	HIS226	7.6	Water	ASP169	5.4	Water	ASP169	5.4
ARG143	ASP207	32.7	Water	HIS229	3.2	ASP169	Water	4.7	ASP169	Water	4.5
ASP207	ALA203	14.4	GLU272	GLU272	29.3	Water	ASP169	4.7	Water	ASP169	2.6
Water	ASP207	2.2	TRP318	GLU272	8.6	ARG291	ASP169	4.2	Water	ASP169	2.1
HIS229	TYR230	19.0	GLU307	GLU437	96.4	ARG143	ASP207	25.8	Water	ASP169	2.0
HIS229	ASP30	15.5	LYS395	GLU307	68.0	ASP207	ALA203	14.5	ASP207	ALA203	20.2
HIS229	Water	5.2	TYR63	GLU307	24.6	HIS229	TYR230	60.7	ARG143	ASP207	19.8
LYS123	HIS229	4.4	ARG311	GLU307	13.5	Water	HIS229	22.8	HIS229	TYR230	58.9
HIS229	HIS226	4.1	TRP318	Water	38.9	HIS229	HIS226	6.5	HIS229	HIS226	6.3
HIS229	Water	2.3	TRP318	GLU272	22.5	GLU307	GLU437	70.1	GLU307	GLU437	126.9
GLU272	GLU272	26.2	TRP318	Water	19.2	LYS395	GLU307	70.0	LYS395	GLU307	73.2
TRP318	GLU272	8.6	TRP318	Water	13.6	TYR63	GLU307	17.7	TYR63	GLU307	26.3
GLU307	GLU437	94.6	TRP318	Water	9.8	ARG311	GLU307	13.2	ARG311	GLU307	13.6
LYS395	GLU307	69.8	ILE322	TRP318	7.4	VAL310	GLU307	1.8	VAL310	GLU307	1.2

Table A.3. Hydrogen bond occupancies between the selected residues are given for the simulations at acidic pH (cont.).

WT						W318C					
diol		%	diol		%	diol		%	diol		%
TYR63	GLU307	18.4	TRP318	Water	4.5	CYS318	Water	12.5	CYS318	Water	29.6
ARG311	GLU307	12.1	TRP318	Water	4.3	CYS318	Water	12.1	CYS318	Water	15.1
TRP318	Water	25.3	TRP318	Water	3.9	CYS318	Water	12.1	CYS318	Water	12.0
TRP318	GLU272	24.6	ARG370	ASP335	77.2	CYS318	Water	11.4	CYS318	Water	11.4
TRP318	Water	23.6	PHE356	ASP335	24.9	CYS318	Water	10.7	CYS318	Water	8.6
TRP318	Water	9.5	SER332	ASP335	19.2	CYS318	Water	10.5	ILE322	CYS318	6.4
ILE322	TRP318	9.1	ASP335	SER332	5.8	ILE322	CYS318	9.7	Water	SER44	5.6
TRP318	Water	6.2	SER332	ASP335	5.1	CYS318	Water	6.1	CYS318	Water	4.3
TRP318	Water	6.1	ARG84	ASP354	93.0	CYS318	Water	5.5	CYS318	Water	3.9
TRP318	Water	5.7	LEU81	ASP354	40.9	CYS318	Water	3.4	CYS318	Water	3.9
TRP318	Water	5.3	ASP354	GLY79	25.3	CYS318	THR274	3.1	Water	ASP35	3.8
TRP318	Water	4.9	Water	ASP354	2.0	CYS318	Water	2.9	CYS318	Water	3.2
TRP318	Water	4.0	Water	ASP354	1.1	CYS318	Water	2.3	CYS318	THR274	2.9
ARG370	ASP335	68.8	LYS463	GLU424	8.3	CYS318	Water	2.3	ARG370	ASP335	74.1
SER332	ASP335	59.5	Water	GLU424	1.3	CYS318	Water	2.2	SER332	ASP335	52.6
PHE356	ASP335	22.6	Water	GLU424	1.1	ARG370	ASP335	73.3	PHE356	ASP335	18.0
SER332	ASP335	4.4	LYS433	ASP430	43.4	SER332	ASP335	58.5	SER332	ASP335	3.4
ARG84	ASP354	72.5	LYS433	Water	16.1	PHE356	ASP335	32.4	ASP335	SER332	2.8
Water	ASP354	22.2	LYS433	ASP430	9.1	SER332	ASP335	3.3	ARG84	ASP354	94.7
Water	ASP354	14.3	ARG217	ASP216	113.0	ARG84	ASP354	94.1	LEU81	ASP354	41.6
ASP354	Water	10.9	ARG217	GLY115	10.6	ASP354	GLY79	40.8	ASP354	GLY79	20.5
Water	ASP354	8.7	ARG217	GLN215	6.3	LEU81	ASP354	34.7	Water	ASP354	1.4
LEU81	ASP354	8.1	ARG217	ALA212	5.9	LYS463	GLU424	4.5	Water	ASP354	1.0
ASP354	Water	5.4	ARG217	Water	3.2	Water	GLU424	1.1	LYS463	GLU424	1.8
Water	ASP354	4.0	ARG217	ASP216	1.7	LYS433	ASP434	36.9	Water	GLU329	1.3
Water	ASP354	3.9	ARG217	ARG114	1.4	LYS433	Water	20.9	LYS433	ASP430	20.6
ASP354	GLY79	3.6	GLU240	ASP249	52.5	LYS433	Water	9.3	LYS433	ASP434	11.4
LYS463	GLU424	2.8	SER368	GLU329	28.4	LYS433	TYR439	4.8	LYS433	ASP430	4.9
LYS433	ASP434	30.3	LYS367	GLU329	1.4	LYS433	ASP430	3.8	LYS433	Water	4.5
LYS433	ASP430	10.2	Water	GLU329	1.4	LYS433	Water	2.0	LYS433	ASN438	2.8
LYS433	Water	10.0	ARG479	GLN476	17.5	LYS433	Water	1.9	LYS433	Water	2.1
LYS433	Water	4.5	ARG479	SER475	7.7	LYS433	ASP430	1.7	LYS433	Water	1.9
LYS433	Water	3.3	ARG97	ARG479	3.7	ARG217	ASP216	67.0	ARG217	ALA212	37.8
LYS433	Water	3.2	ARG479	Water	3.3	ARG217	ALA212	15.8	ARG217	GLY115	23.4
ARG217	GLY115	35.4	ARG479	ALA478	2.9	ARG217	GLY115	10.4	ARG217	GLN215	15.6
ARG217	ALA212	23.6	ARG479	Water	1.9	ARG217	GLN215	3.2	Water	Water	7.6
ARG217	GLN215	14.0	ASP482	SER95	41.1	ARG217	ASP216	2.1	ARG217	ASP216	3.6

Table A.3. Hydrogen bond occupancies between the selected residues are given for the simulations at acidic pH (cont.).

WT						W318C					
diol		%	diol		%	diol		%	diol		%
ARG217	ASP216	3.0	ARG97	ASP482	13.2	No 240	No 240	0.0	Water	Water	2.5
ARG217	Water	3.0	LEU485	ASP482	2.0	SER368	GLU329	33.2	Water	Water	2.2
GLU240	ASP249	70.3	ALA484	ASP482	1.5	Water	GLU329	1.7	ARG217	ASP216	1.5
SER368	GLU329	28.6				Water	GLU329	1.3	GLU240	ASP249	68.4
Water	GLU329	3.6				ARG479	GLN476	29.0	SER368	GLU329	29.2
Water	GLU329	3.2				ARG479	Water	7.5	Water	GLU329	2.4
Water	GLU329	3.0				ARG97	ARG479	5.7	Water	GLU329	2.4
Water	GLU329	2.5				ARG479	Water	5.6	Water	GLU329	2.0
ARG479	GLN476	10.2				ARG479	Water	3.6	Water	GLU329	1.9
ARG479	Water	7.9				ARG479	Water	3.2	ARG479	PRO474	18.2
ARG479	GLU481	7.2				ARG479	Water	2.5	ARG97	ARG479	8.4
ARG479	Water	7.1				ARG479	Water	2.4	ARG479	SER475	5.3
ARG479	Water	6.8				ASP482	SER95	38.5	ARG479	GLN476	4.6
ARG479	Water	4.1				ARG97	ASP482	17.2	ARG479	GLN476	2.4
ARG479	Water	3.9				Water	THR70	2.4	ARG479	MET85	2.3
ARG479	Water	3.8				ALA484	ASP482	1.9	ARG97	ASP482	19.4
ARG479	SER475	3.4				LEU485	ASP482	1.0	ASP482	SER95	18.8
ARG479	Water	3.2							LEU485	ASP482	1.6
ARG479	Water	3.2							ALA484	ASP482	1.3
ASP482	SER95	46.8									
ARG97	ASP482	19.4									
LEU485	ASP482	2.1									
ALA484	ASP482	1.4									
E307K+T391S						E307K					
diol		(%)	nonhyd		(%)	diol		(%)	nonhyd		(%)
ARG19	ASP216	90.2	ARG19	ASP216	100.2	ARG19	ASP216	89.8	ARG19	GLU273	78.7
ARG19	GLU273	11.4	ARG19	GLU273	6.0	ARG19	GLU273	54.9	ARG19	ASP216	29.1
ARG19	GLU273	2.3	ARG19	GLN213	4.2	ARG19	GLU273	5.3	ARG19	GLU273	3.5
ARG19	Water	1.4	ARG19	GLU273	2.1	ARG19	Water	2.5	ARG19	Water	2.1
ASP29	HIP227	54.3	ARG19	Water	2.0	LYS123	ASP29	64.7	ARG19	Water	1.9
LYS123	ASP29	48.0	ARG19	Water	1.8	ASP29	HIP227	46.7	LYS123	ASP29	62.0
ARG73	ASP29	41.4	ASP29	HIP227	78.1	ARG73	ASP29	13.1	ASP29	HIP227	60.6
Water	ASP29	16.1	LYS123	ASP29	61.8	ASP50	THR46	4.5	ARG73	ASP29	9.4
Water	GLU103	1.8	Water	ASP29	14.6	ARG311	ASP61	49.8	Water	ASP29	2.4
ASP50	THR46	2.6	Water	ASP29	6.6	ARG390	ASP61	28.1	ASP50	THR46	2.2
ALA54	ASP50	1.5	Water	ASP29	2.5	TYR63	ASP61	17.3	ARG311	ASP61	57.8
ARG311	ASP61	40.9	Water	ASP29	1.8	THR60	ASP61	10.9	ARG390	ASP61	39.6

Table A.3. Hydrogen bond occupancies between the selected residues are given for the simulations at acidic pH (cont.).

E307K+T391S						E307K					
diol		(%)	diol		(%)	diol		(%)	diol		(%)
ARG390	ASP61	22.4	Water	ASP29	1.4	ALA331	ASP61	4.1	THR60	ASP61	10.8
THR60	ASP61	11.4	ALA54	ASP50	1.2	Water	ASP61	3.9	ALA331	ASP61	2.3
Water	ASP61	7.7	ARG311	ASP61	50.8	ARG73	ASP281	73.0	ARG73	Water	90.6
Water	ASP61	6.6	TYR63	ASP61	32.8	ARG73	FGL69	52.0	ARG73	ASP29	46.8
ALA331	ASP61	6.6	ARG390	ASP61	19.4	ARG73	ASP29	40.2	ARG73	SER225	38.8
Water	ASP61	4.3	THR60	ASP61	10.9	ARG73	UNK1	36.0	ARG73	ASP281	16.9
ARG73	Water	124.6	ALA331	ASP61	6.2	ARG73	SER225	27.3	ARG73	LYS123	14.4
ARG73	ASP29	93.7	Water	ASP61	4.5	ARG73	LYS123	8.1	ARG73	FGL69	11.6
ARG73	FGL69	63.6	ARG73	ASP281	109.2	ARG97	ASP482	96.0	ARG73	ALA224	7.6
ARG73	UNK1	22.1	ARG73	Water	68.1	ARG97	ARG479	36.0	ARG97	ASP482	75.4
ARG73	Water	4.0	ARG73	FGL69	38.3	ARG97	GLY480	35.3	ARG97	ARG479	38.3
ARG73	ASP281	3.1	ARG73	UNK1	37.7	ARG97	Water	2.9	ARG97	GLY480	25.7
ARG97	ASP482	64.5	ARG73	Water	31.6	ARG97	Water	2.2	ARG97	GLU131	15.6
ARG97	GLU131	64.0	ARG73	ASP29	5.5	ASH173	ASH152	21.0	ARG97	Water	4.0
ARG97	GLY480	41.2	ARG73	SER225	4.9	ASH152	GLY171	4.1	ARG97	Water	3.9
ARG97	ARG479	28.5	ARG73	Water	4.6	ASH152	Water	3.9	ASH152	Water	1.7
ARG97	Water	17.3	ARG97	ASP482	69.2	ASH152	SER150	3.6	ASH152	Water	1.7
ARG97	Water	12.1	ARG97	GLY480	51.6	ASH152	Water	3.5	ASH152	Water	1.4
ARG97	Water	8.5	ARG97	GLU131	16.0	ASH152	Water	3.5	ASH152	Water	1.4
ARG97	Water	7.3	ARG97	ARG479	7.9	Water	ASP169	13.1	ASH152	Water	1.3
ARG97	Water	3.6	ARG97	Water	5.3	ASP169	Water	11.2	TYR230	ASP169	6.6
ARG97	Water	2.8	ARG97	Water	3.3	TYR230	ASP169	5.6	ASP169	Water	4.3
ASH173	ASH152	17.2	ASH152	HIS229	14.7	Water	ASP169	1.9	Water	ASP169	3.7
ASH152	SER150	7.1	ASH152	Water	2.6	ASP169	Water	1.0	ARG143	ASP207	42.6
ASH152	Water	3.3	ASH152	Water	2.1	ARG143	ASP207	19.8	ASP207	ALA203	9.7
ASH152	Water	3.1	ASH152	Water	1.8	ASP207	ALA203	17.7	HIS229	TYR230	73.0
Water	ASP169	25.1	ASP173	ASH152	1.6	Water	HIE151	9.7	Water	HIS229	20.4
TYR230	ASP169	25.0	ASH152	Water	1.6	Water	GLY171	8.7	HIS229	HIS226	3.4
ASP169	Water	16.2	TYR230	ASP169	43.2	Water	GLN153	7.8	Water	HIS229	1.2
TYR230	ASP169	9.4	Water	ASP169	8.6	Water	PRO155	2.7	LYS395	LYN307	37.0
Water	ASP169	5.3	ASP169	Water	8.0	HIS229	TYR230	82.1	LYN307	PRO435	23.0
Water	ASP169	3.5	ARG288	ASP169	4.8	HIS229	HIS226	7.9	ARG311	LYN307	15.2
ASP169	Water	3.3	Water	ASP169	4.7	LYN307	GLU437	56.0	LYN307	GLU437	7.6
Water	ASP169	3.0	Water	ASP169	3.0	LYN307	PRO435	14.2	LYN307	TYR63	3.2
ARG291	ASP169	2.7	ASP169	Water	2.9	ARG311	LYN307	13.6	TYR63	LYN307	1.4
ARG143	ASP207	35.3	ARG143	ASP207	30.7	TYR63	LYN307	10.5	TRP318	Water	26.7
ASP207	ALA203	13.2	ASP207	ALA203	14.6	Water	LYN307	6.0	TRP318	GLU272	21.3

Table A.3. Hydrogen bond occupancies between the selected residues are given for the simulations at acidic pH (cont.).

E307K+T391S						E307K					
diol		(%)	diol		(%)	diol		(%)	diol		(%)
HIS229	TYR230	60.2	HIS229	ASP30	69.9	LYN307	Water	4.9	TRP318	Water	19.8
HIS229	HIS226	4.7	HIS229	TYR230	23.3	TRP318	GLU272	26.0	TRP318	Water	15.9
LYN307	GLU437	65.7	ASH152	HIS229	6.4	TRP318	Water	22.2	TRP318	Water	13.7
LYN307	PRO435	31.1	HIS229	HIS226	2.8	TRP318	Water	21.9	TRP318	Water	12.8
ARG311	LYN307	18.7	LYN307	GLU437	54.9	TRP318	Water	18.2	ILE322	TRP318	10.1
ARG311	LYN307	2.4	LYN307	PRO435	22.6	TRP318	Water	11.9	TRP318	Water	4.6
VAL310	LYN307	1.5	ARG311	LYN307	10.4	TRP318	Water	9.0	TRP318	Water	3.9
TYR63	LYN307	1.3	Water	LYN307	5.5	ILE322	TRP318	7.1	ARG370	ASP335	67.2
TRP318	Water	69.2	TYR63	LYN307	5.1	TRP318	Water	5.7	SER332	ASP335	32.4
TRP318	GLU272	32.3	LYN307	Water	2.1	TRP318	Water	5.5	PHE356	ASP335	18.5
TRP318	Water	11.3	TRP318	Water	97.6	TRP318	Water	3.2	SER332	ASP335	14.5
TRP318	Water	10.7	TRP318	GLU272	21.2	ARG370	ASP335	73.2	ASP335	SER332	8.8
TRP318	Water	9.5	ILE322	TRP318	12.3	SER332	ASP335	54.2	ARG84	ASP354	96.2
ILE322	TRP318	7.0	TRP318	Water	7.1	PHE356	ASP335	25.9	LEU81	ASP354	45.5
Water	TYR63	1.5	TRP318	Water	5.3	ASP335	SER332	2.4	ASP354	GLY79	35.3
Water	GLU437	1.1	ARG370	ASP335	78.6	SER332	ASP335	2.3	No424	No424	0.0
ARG370	ASP335	73.7	SER332	ASP335	31.5	ARG84	ASP354	95.1	LYS433	ASP430	16.6
SER332	ASP335	58.1	PHE356	ASP335	17.9	LEU81	ASP354	41.8	LYS433	ASP430	3.0
PHE356	ASP335	23.0	ASP335	SER332	13.1	ASP354	GLY79	32.0	LYS433	ASP430	1.5
SER332	ASP335	3.0	ARG84	ASP354	92.1	Water	ASP354	1.6	ARG217	ASP211	69.7
ASP335	SER332	1.4	LEU81	ASP354	35.0	LYS463	GLU424	9.3	ARG217	Water	69.0
ARG84	ASP354	93.4	ASP354	GLY79	29.1	LYS433	ASP430	25.8	ARG217	GLY115	20.5
LEU81	ASP354	41.2	Water	ASP354	2.1	LYS433	ASP430	12.9	ARG217	ALA212	15.1
ASP354	GLY79	24.3	LYS463	GLU424	4.3	LYS433	Water	1.3	ARG217	GLN215	10.1
Water	ASP354	1.8	Water	GLU424	1.2	ARG217	GLY115	49.1	GLU240	ASP249	79.7
Water	ASP354	1.7	LYS433	ASP430	19.0	ARG217	ALA212	38.4	Water	GLU103	1.2
LYS463	GLU424	20.8	LYS433	ASP430	5.9	ARG217	GLN215	7.4	Water	GLU240	1.1
GLN460	GLU424	1.1	LYS433	ASP430	3.3	ARG217	Water	2.3	SER368	GLU329	24.5
LYS433	ASP430	22.1	ARG217	ALA212	35.7	ARG217	ASP216	1.6	Water	GLU329	19.6
LYS433	Water	12.5	ARG217	GLY115	26.9	ARG217	Water	1.4	Water	GLU329	9.5
LYS433	ASP430	11.6	ARG217	GLN215	17.1	GLU240	ASP249	59.6	GLN371	GLU329	7.0
ARG217	GLY115	45.8	ARG217	TYR116	3.4	Water	GLU240	3.1	Water	GLU329	4.5
ARG217	ALA212	35.2	ARG217	ASP216	3.3	Water	GLU240	1.7	Water	GLU329	2.8
ARG217	GLN215	11.2	ARG217	Water	1.4	SER368	GLU329	27.4	Water	GLU329	2.4
ARG217	ASP216	5.7	GLU240	ASP249	65.5	Water	GLU329	1.3	Water	GLU329	2.4
ARG217	TYR116	3.0	SER368	GLU329	27.9	Water	ASP35	1.3	Water	GLU329	2.2
ARG217	Water	2.0	Water	GLU329	9.0	ARG97	ARG479	17.8	Water	GLU329	2.1



Table A.4. The hydrogen bond occupancy percentages of the selected residues in the chains of dimer forms of ARSA at neutral pH.

WT					
Chain1		(%)	Chain2		(%)
ARG19	ASP216	99.10%	ARG19	ASP216	60.84%
ARG19	GLU273	54.77%	ARG19	GLU273	5.38%
ARG19	GLN213	8.25%	ARG19	GLN213	3.76%
ARG19	Water	1.79%	ARG19	ASP216	2.14%
ASP29	HIE227	75.42%	ASP29	HIE227	49.52%
ASP29	HIS229	46.61%	ASP29	ARG73	21.94%
ASP29	LYN123	2.65%	ASP29	Water	9.33%
ASP29	THR228	2.38%	ASP29	Water	8.42%
ASP61	THR60	54.63%	ASP61	ARG390	72.33%
ASP61	ARG390	26.29%	ASP61	THR60	38.36%
ASP61	Water	23.86%	ASP61	ALA331	19.57%
ASP61	ARG311	10.00%	ASP61	Water	6.56%
ASP61	Water	2.74%	ASP61	Water	4.52%
ASP61	Water	2.54%	ASP61	Water	3.16%
ARG73	ASP281	79.23%	ARG73	FGL69	83.69%
ARG73	SER225	53.34%	ARG73	ASP281	41.45%
ARG73	ALA224	12.37%	ARG73	Water	40.52%
ARG73	FGL69	10.36%	ARG73	ASP29	36.49%
ARG73	ASP29	8.14%	ARG73	TYR222	10.73%
ARG73	LYN123	6.74%	ARG73	LYN123	2.34%
ARG73	Water	4.63%	ARG73	Water	1.79%
ASP169	ARG291	40.81%	ASP169	ARG291	44.95%
ASP169	Water	6.10%	ASP169	TYR230	24.40%
ASP169	Water	5.42%	ASP169	Water	16.35%
ASP169	Water	3.33%	ASP169	TYR230	5.02%
ASP169	Water	2.95%	ASP169	Water	4.89%
ASP169	Water	2.80%	ASP169	Water	3.78%

Table A.4. The hydrogen bond occupancy percentages of the selected residues in the chains of dimer forms of ARSA at neutral pH (cont.).

WT					
Chain1		Chain1	Chain1		Chain1
ASP169	Water	2.49%	ASP169	Water	2.61%
GLY171	CYX156	11.07%	GLY171	CYX156	9.00%
GLY171	ASP169	0.89%	GLY171	ASP169	0.72%
GLY171	Water	0.48%	GLY171	GLN174	0.16%
ASP173	ASP152	2.90%	ASP173	ASP152	6.36%
ASP173	Water	1.44%	ASP173	GLN153	2.41%
ASP173	Water	1.33%	ASP173	Water	0.95%
GLU188	LEU181	99.75%	GLU188	LEU181	71.79%
GLU188	ARG200	40.45%	GLU188	ARG200	18.12%
GLU240	ASP249	102.58%	GLU240	ASP249	63.86%
GLU240	Water	0.20%	GLU240	Water	1.73%
ARG241	SER237	50.43%	ARG241	SER237	59.37%
ARG241	ASP249	8.85%	ARG241	ASP249	2.99%
ARG241	PHE238	1.61%	ARG241	PHE238	0.69%
ARG241	ALA239	1.16%	ARG241	Water	0.62%
GLU307	GLU437	82.78%	GLU307	GLU437	77.78%
GLU307	Water	15.78%	GLU307	TYR63	39.57%
GLU307	TYR63	12.22%	GLU307	ARG311	8.77%
GLU307	ARG311	9.48%	GLU307	ARG311	5.86%
TRP318	GLU272	28.49%	TRP318	ILE322	41.21%
TRP318	Water	12.84%	TRP318	GLU272	29.72%
TRP318	Water	5.74%	TRP318	Water	19.48%
TRP318	Water	2.53%	TRP318	Water	13.88%
TRP318	Water	2.51%	TRP318	Water	12.72%
GLU329	SER368	7.77%	GLU329	SER368	52.29%
GLU329	LYS367	1.49%	GLU329	SER368	18.57%
GLU329	Water	1.23%	GLU329	LYS367	12.25%

Table A.4. The hydrogen bond occupancy percentages of the selected residues in the chains of dimer forms of ARSA at neutral pH (cont.).

WT					
Chain1		Chain1	Chain1		Chain1
GLU329	Water	1.15%	GLU329	Water	2.03%
ASP335	ARG370	53.03%	ASP335	ARG370	89.86%
ASP335	SER332	47.57%	ASP335	PHE356	28.61%
ASP335	THR339	12.94%	ASP335	SER332	26.91%
ASP335	PHE356	8.82%	ASP335	THR339	21.63%
ASP335	THR339	0.65%	ASP335	Water	9.63%
LYS395	GLU307	145.78%	LYS395	GLU307	163.80%
LYS395	TYR429	61.66%	LYS395	TYR429	89.11%
LYS395	GLU437	48.82%	LYS395	GLU437	86.04%
LYS395	TYR429	22.40%	LYS395	THR305	24.70%
LYS395	Water	20.83%	LYS395	TYR429	0.09%
LYS395	THR305	15.75%			
PHE398	PHE387	54.02%	PHE398	PHE387	52.41%
PHE398	HIS423	35.93%	PHE398	HIS423	19.87%
ASP430	Water	27.50%	ASP430	SER432	51.00%
ASP430	LYS393	16.78%	ASP430	LYS393	38.64%
ASP430	LYS433	16.63%	ASP430	Water	30.32%
ASP430	SER432	16.63%	ASP430	Water	10.19%
ASP430	LYS433	8.64%	ASP430	Water	6.96%
ASP430	SER432	8.26%	ASP430	LYS433	5.09%
ASP430	Water	5.73%	ASP430	Water	3.87%
ASP430	LYS433	3.98%	ASP430	Water	3.17%
ASP430	ASP434	3.44%	ASP430	Water	3.07%
ARG479	PRO474	5.48%	ARG479	GLN476	14.57%
ARG479	GLN476	3.62%	ARG479	GLN476	7.08%
ARG479	ARG97	3.17%	ARG479	ARG97	3.23%
ARG479	Water	2.77%	ARG479	SER475	2.78%

Table A.4. The hydrogen bond occupancy percentages of the selected residues in the chains of dimer forms of ARSA at neutral pH (cont.).

WT					
Chain1		Chain1	Chain1		Chain1
ARG479	GLN476	2.37%	ARG479	Water	1.98%
ASP482	SER95	42.76%	ASP482	SER95	41.85%
ASP482	ALA484	21.40%	ASP482	ARG97	38.80%
ASP482	LEU485	7.99%	ASP482	Water	5.32%
ASP482	ARG97	2.73%	ASP482	Water	1.53%
GLN460	LEU456	20.61%	GLN460	LEU456	49.48%
GLN460	LEU456	6.33%	GLN460	GLU424	13.22%
GLN460	GLU424	1.99%	GLN460	LYS463	5.27%
GLN460	LYS457	0.82%	GLN460	ALA464	1.54%
W318C					
Chain1		(%)	Chain2		(%)
ARG19	ASP216	102.13%	ARG19	GLU273	78.09%
ARG19	GLU273	78.08%	ARG19	ASP216	69.92%
ARG19	GLN213	16.11%	ARG19	GLU273	1.63%
ARG19	Water	5.36%	ARG19	Water	0.80%
ASP29	HIE227	37.24%	ASP29	LYN123	32.16%
ASP29	ARG73	34.73%	ASP29	HIE227	24.47%
ASP29	Water	23.28%	ASP29	THR228	6.73%
ASP29	LYN123	10.11%	ASP29	HIS229	3.94%
ASP61	ARG390	66.60%	ASP61	THR60	72.76%
ASP61	THR60	24.27%	ASP61	ARG390	24.51%
ASP61	ALA331	18.21%	ASP61	Water	5.39%
ASP61	Water	5.60%	ASP61	ARG311	5.30%
ASP61	Water	3.41%	ASP61	Water	2.06%
ASP61	Water	1.47%	ASP61	Water	1.22%
ARG73	ASP29	77.95%	ARG73	ASP281	85.34%
ARG73	Water	56.99%	ARG73	FGL69	51.15%

Table A.4. The hydrogen bond occupancy percentages of the selected residues in the chains of dimer forms of ARSA at neutral pH (cont.).

W318C					
Chain1		Chain1	Chain1		Chain1
ARG73	Water	54.92%	ARG73	UNK1	50.36%
ARG73	FGL69	33.42%	ARG73	ASP29	34.95%
ARG73	ASP281	14.04%	ARG73	SER225	29.24%
ARG73	SER225	11.48%	ARG73	LYN123	11.85%
ARG73	LEU77	11.18%	ARG73	TYR222	0.39%
ASP169	TYR230	78.14%	ASP169	ARG291	16.21%
ASP169	Water	45.89%	ASP169	Water	3.40%
ASP169	Water	10.92%	ASP169	Water	1.93%
ASP169	Water	8.66%	ASP169	TYR230	1.73%
ASP169	Water	4.40%	ASP169	Water	1.28%
ASP169	Water	3.42%	ASP169	Water	0.91%
ASP169	Water	2.60%	ASP169	Water	0.54%
GLY171	CYX156	3.74%	GLY171	TYR230	24.32%
GLY171	ASP169	0.90%	GLY171	CYX156	0.65%
GLY171	GLN174	0.46%	GLY171	Water	0.40%
ASP173	ASP152	30.98%	ASP173	ASP152	35.81%
ASP173	GLN153	7.06%	ASP173	GLY154	7.83%
ASP173	GLY154	1.67%	ASP173	Water	0.71%
GLU188	LEU181	83.42%	GLU188	LEU181	84.84%
GLU188	ARG200	7.51%	GLU188	ARG200	23.68%
GLU240	ASP249	59.43%	GLU240	ASP249	85.40%
GLU240	Water	2.06%	GLU240	Water	0.26%
ARG241	SER237	54.96%	ARG241	SER237	40.02%
ARG241	ASP249	4.80%	ARG241	ASP249	9.56%
ARG241	PHE238	0.67%	ARG241	ALA239	3.11%
ARG241	Water	0.56%	ARG241	PHE238	2.67%
GLU307	GLU437	138.20%	GLU307	GLU437	45.32%

Table A.4. The hydrogen bond occupancy percentages of the selected residues in the chains of dimer forms of ARSA at neutral pH (cont.).

W318C					
Chain1		Chain1	Chain1		Chain1
GLU307	TYR63	62.87%	GLU307	TYR63	21.72%
GLU307	ARG311	2.61%	GLU307	ARG311	18.79%
GLU307	ARG311	0.84%	GLU307	ARG311	0.66%
CYS318	ILE322	38.35%	CYS318	Water	6.35%
CYS318	Water	10.72%	CYS318	Water	6.02%
CYS318	Water	9.19%	CYS318	Water	6.01%
CYS318	Water	9.04%	CYS318	Water	3.17%
CYS318	Water	8.28%	CYS318	THR274	2.70%
GLU329	SER368	36.90%	GLU329	SER368	12.54%
GLU329	LYS367	11.49%	GLU329	Water	8.68%
GLU329	Water	1.04%	GLU329	Water	4.62%
GLU329	Water	1.00%	GLU329	Water	4.31%
ASP335	ARG370	88.52%	ASP335	SER332	63.08%
ASP335	PHE356	41.61%	ASP335	ARG370	38.25%
ASP335	THR339	21.10%	ASP335	PHE356	8.15%
ASP335	SER332	6.57%	ASP335	THR339	6.39%
LYS395	GLU307	165.63%	LYS395	GLU307	166.97%
LYS395	GLU437	148.49%	LYS395	GLU437	67.38%
LYS395	TYR429	95.96%	LYS395	TYR429	59.27%
LYS395	THR305	23.31%	LYS395	THR305	46.06%
LYS395	TYR429	1.22%	LYS395	TYR429	20.54%
PHE398	PHE387	55.44%	PHE398	PHE387	54.46%
PHE398	HIS423	31.88%	PHE398	HIS423	35.07%
ASP430	SER432	37.37%	ASP430	LYS433	46.42%
ASP430	LYS393	34.32%	ASP430	Water	30.31%
ASP430	Water	20.35%	ASP430	Water	19.90%
ASP430	Water	6.40%	ASP430	Water	9.24%

Table A.4. The hydrogen bond occupancy percentages of the selected residues in the chains of dimer forms of ARSA at neutral pH (cont.).

W318C					
Chain1		Chain1	Chain1		Chain1
ASP430	Water	3.73%	ASP430	Water	7.04%
ASP430	LYS433	2.53%	ASP430	Water	6.02%
ASP430	Water	2.43%	ASP430	Water	4.75%
ASP430	Water	1.51%	ASP430	SER432	4.66%
ASP430	Water	0.91%	ASP430	Water	4.27%
ARG479	SER475	15.18%	ARG479	SER475	15.18%
ARG479	GLN476	10.62%	ARG479	GLN476	10.62%
ARG479	Water	6.66%	ARG479	Water	6.66%
ARG479	SER475	6.25%	ARG479	SER475	6.25%
ARG479	Water	2.88%	ARG479	Water	2.88%
ASP482	ARG97	67.90%	ASP482	ARG97	67.90%
ASP482	Water	1.75%	ASP482	Water	1.75%
ASP482	Water	1.56%	ASP482	Water	1.56%
ASP482	Water	1.30%	ASP482	Water	1.30%
GLN460	LEU456	11.90%	GLN460	LEU456	25.01%
GLN460	LEU456	1.69%	GLN460	GLU424	5.53%
GLN460	LYS457	1.32%	GLN460	ALA464	2.36%
GLN460	LYS457	0.74%	GLN460	LYS463	0.52%
E307K+T391S					
Chain1		(%)	Chain2		(%)
ARG19	GLU273	120.95%	ARG19	GLU273	155.82%
ARG19	ASP216	15.22%	ARG19	GLU273	4.24%
ARG19	GLU272	5.43%	ARG19	ASP216	1.58%
ARG19	Water	2.23%	ARG19	ASP216	0.83%
ASP29	HIE227	67.34%	ASP29	HIE227	41.79%
ASP29	LYN123	46.86%	ASP29	ARG73	38.25%
ASP29	Water	4.48%	ASP29	Water	8.12%

Table A.4. The hydrogen bond occupancy percentages of the selected residues in the chains of dimer forms of ARSA at neutral pH (cont.).

E307K+T391S					
Chain1		(%)	Chain2		(%)
ASP29	ARG73	2.58%	ASP29	Water	0.57%
ASP61	THR60	64.49%	ASP61	ARG390	49.75%
ASP61	ARG311	29.43%	ASP61	ALA331	31.31%
ASP61	TYR63	16.27%	ASP61	TYR63	7.77%
ASP61	ARG390	10.40%	ASP61	THR60	4.32%
ASP61	Water	0.82%	ASP61	Water	2.66%
ASP61	Water	0.71%	ASP61	Water	2.53%
ARG73	ASP29	59.31%	ARG73	FGL69	130.30%
ARG73	ASP281	43.33%	ARG73	ASP29	95.18%
ARG73	Water	40.75%	ARG73	ASP281	54.70%
ARG73	SER225	33.13%	ARG73	UNK1	45.70%
ARG73	FGL69	29.07%	ARG73	Water	44.88%
ARG73	Water	23.87%	ARG73	TRP124	3.89%
ARG73	Water	21.80%	ARG73	LYN123	0.59%
ASP169	Water	2.95%	ASP169	ARG291	70.69%
ASP169	Water	2.27%	ASP169	TYR230	3.15%
ASP169	ARG291	1.88%	ASP169	Water	2.84%
ASP169	Water	0.40%	ASP169	Water	1.90%
ASP169	Water	0.35%	ASP169	Water	1.50%
ASP169	Water	0.33%	ASP169	Water	1.44%
ASP169	Water	0.30%	ASP169	TYR230	1.38%
GLY171	ASP152	44.36%	GLY171	CYX156	0.42%
GLY171	CYX156	1.81%	GLY171	TYR230	0.35%
GLY171	Water	0.27%	GLY171	Water	0.29%
ASP173	GLY175	3.27%	ASP173	GLY154	7.82%
ASP173	Water	3.23%	ASP173	ASP152	7.04%
ASP173	Water	2.70%	ASP173	GLN153	3.06%

Table A.4. The hydrogen bond occupancy percentages of the selected residues in the chains of dimer forms of ARSA at neutral pH (cont.).

E307K+T391S					
Chain1		(%)	Chain2		(%)
GLU188	LEU181	121.72%	GLU188	LEU181	62.15%
GLU188	ARG200	7.86%	GLU188	ARG200	34.43%
GLU240	ASP249	93.84%	GLU240	ASP249	50.33%
GLU240	Water	0.23%	GLU240	Water	0.89%
ARG241	SER237	46.52%	ARG241	SER237	68.83%
ARG241	ASP249	7.39%	ARG241	ASP249	4.43%
ARG241	PHE238	1.24%	ARG241	PHE238	1.21%
ARG241	Water	0.65%	ARG241	ALA239	0.74%
LYN307	GLU437	126.35%	LYN307	PRO435	48.01%
LYN307	ARG311	13.56%	LYN307	GLU437	42.78%
LYN307	Water	10.13%	LYN307	ARG311	13.84%
LYN307	TYR63	9.19%	LYN307	ARG311	9.87%
LYN307	Water	9.12%	LYN307	Water	0.96%
TRP318	GLU272	23.86%	TRP318	ILE322	38.17%
TRP318	Water	6.76%	TRP318	GLU272	24.60%
TRP318	Water	6.09%	TRP318	Water	13.28%
TRP318	Water	5.20%	TRP318	Water	10.06%
TRP318	Water	4.34%	TRP318	Water	9.15%
GLU329	SER368	18.55%	GLU329	SER368	24.84%
GLU329	Water	1.22%	GLU329	LYS367	9.52%
GLU329	Water	0.62%	GLU329	Water	0.89%
GLU329	Water	0.48%	GLU329	Water	0.88%
ASP335	ARG370	67.91%	ASP335	ARG370	85.20%
ASP335	SER332	69.85%	ASP335	PHE356	30.83%
ASP335	PHE356	0.55%	ASP335	THR339	4.74%
ASP335	THR339	0.08%	ASP335	SER332	3.21%
LYS395	GLU437	125.66%	LYS395	THR305	84.86%

Table A.4. The hydrogen bond occupancy percentages of the selected residues in the chains of dimer forms of ARSA at neutral pH (cont.).

E307K+T391S					
Chain I		Chain I	Chain I		Chain I
LYS395	TYR429	82.71%	LYS395	TYR429	84.65%
LYS395	Water	35.66%	LYS395	GLU437	79.13%
LYS395	HIE397	23.27%	LYS395	Water	9.46%
LYS395	Water	21.90%	LYS395	TYR429	5.52%
LYS395	TYR429	4.30%	LYS395	Water	0.26%
PHE398	PHE387	38.66%	PHE398	PHE387	51.10%
PHE398	HIS423	10.10%	PHE398	HIS423	14.68%
ASP430	SER432	37.05%	ASP430	Water	52.80%
ASP430	LYS433	26.85%	ASP430	LYS393	31.03%
ASP430	SER432	19.28%	ASP430	SER432	17.93%
ASP430	Water	11.12%	ASP430	Water	10.07%
ASP430	Water	8.43%	ASP430	Water	8.61%
ASP430	Water	5.60%	ASP430	LYS433	8.09%
ASP430	Water	4.41%	ASP430	LYS433	5.11%
ASP430	Water	4.31%	ASP430	Water	5.08%
ASP430	Water	3.80%	ASP430	Water	2.52%
ARG479	SER96	5.97%	ARG479	GLN476	9.44%
ARG479	GLN476	5.93%	ARG479	Water	2.53%
ARG479	SER475	3.21%	ARG479	ARG97	2.27%
ARG479	Water	2.57%	ARG479	SER475	2.22%
ARG479	ALA478	2.19%	ARG479	ALA478	1.74%
ASP482	SER95	22.98%	ASP482	SER95	46.63%
ASP482	ALA484	18.42%	ASP482	ARG97	31.58%
ASP482	ARG97	12.42%	ASP482	ALA484	0.49%
ASP482	SER95	1.70%	ASP482	Water	0.40%
GLN460	LEU456	32.28%	GLN460	GLU424	26.75%
GLN460	LYS463	19.68%	GLN460	LEU456	23.58%

Table A.5. All possible salt bridges of the monomers forms of WT and W318C mutant monomers at neutral pH.

WT	W318C	WT-R_5	W318C-R_5	W318C-ZINC
ASP169-ARG291	ASP61-ARG390	ASP61-LYS393	ASP482-ARG97	ASP381-ARG384
ASP61-ARG311	GLU329-LYS367	ASP207-ARG143	ASP281-ARG73	ASP354-ARG84
GLU307-ARG311	ASP169-ARG291	GLU424-LYS463	ASP61-ARG390	ASP216-ARG19
ASP61-ARG390	ASP61-ARG311	ASP169-ARG288	GLU437-LYS395	ASP207-ARG143
ASP411-ARG299	GLU307-LYS395	GLU329-LYS367	ASP207-ARG143	ASP216-ARG217
GLU285-ARG288	GLU307-ARG311	ASP354-ARG84	ASP30-HIS229	GLU285-ARG288
ASP211-ARG143	ASP467-LYS463	ASP30-LYS302	ASP30-ARG288	GLU273-ARG19
ASP482-ARG97	ASP335-ARG370	ASP411-ARG299	ASP216-ARG217	ASP211-ARG143
ASP354-ARG84	ASP411-ARG299	GLU104-ARG80	GLU104-ARG80	ASP430-LYS433
ASP381-ARG384	GLU285-ARG288	GLU285-ARG288	GLU307-ARG311	ASP211-ARG214
GLU285-HIS405	ASP381-ARG384	ASP216-ARG19	ASP216-ARG19	ASP482-ARG97
ASP207-ARG143	ASP354-ARG84	ASP35-ARG244	ASP152-ARG288	ASP207-ARG214
ASP281-ARG73	ASP216-ARG19	GLU131-ARG97	ASP434-LYS433	ASP281-ARG73
GLU437-LYS395	ASP211-ARG143	ASP207-ARG214	ASP207-ARG214	ASP30-LYS302
ASP30-LYS302	ASP482-ARG97	ASP430-LYS393	ASP61-LYS393	GLU307-ARG311
ASP30-HIS229	GLU437-LYS395	ASP61-ARG311	ASP30-LYS302	ASP30-HIS229
ASP29-ARG73	ASP281-ARG73	GLU285-LYS302	GLU131-ARG97	ASP411-ARG299
ASP207-ARG214	GLU285-HIS405	GLU273-ARG19	ASP35-ARG244	ASP335-ARG370
ASP211-ARG214	ASP207-ARG143	GLU382-ARG384	GLU451-LYS393	ASP152-HIS229
ASP430-LYS433	ASP211-ARG214	ASP207-HIS206	ASP411-ARG299	ASP29-ARG73
ASP35-ARG244	ASP30-LYS302	GLU272-ARG19	GLU424-LYS463	GLU188-ARG200
GLU104-ARG80	ASP30-HIS229	ASP430-LYS433	ASP61-ARG311	GLU329-LYS367
GLU188-ARG200	ASP29-ARG73	GLU451-LYS393	ASP381-ARG384	ASP35-ARG244
GLU307-LYS395	GLU188-ARG200	GLU307-LYS395	GLU285-ARG288	ASP29-LYS123
ASP335-ARG370	ASP430-LYS433	ASP335-ARG370	ASP29-HIS229	ASP255-ARG241
GLU329-LYS367	GLU131-ARG97	ASP211-ARG214	ASP354-ARG84	GLU307-LYS395
GLU451-LYS393	ASP35-ARG244	GLU188-ARG200	ASP430-LYS393	GLU131-ARG97
ASP467-LYS463	GLU104-ARG80	ASP381-ARG384	GLU285-LYS302	ASP430-LYS393
ASP207-HIS206	GLU451-LYS393	ASP169-ARG291	GLU103-ARG479	GLU104-ARG80
GLU285-LYS302	ASP430-LYS393	GLU285-HIS405	GLU188-ARG200	GLU424-LYS463
ASP434-LYS433	GLU424-LYS463	ASP29-ARG73	ASP211-ARG214	GLU437-LYS395
ASP430-LYS393	GLU285-LYS302	ASP29-LYS123	ASP169-ARG291	ASP434-LYS433
GLU131-ARG496	ASP434-LYS433	GLU437-LYS395	ASP152-HIS229	GLU285-LYS302
	ASP207-ARG214	ASP61-ARG390	ASP430-LYS433	ASP152-ARG288
		ASP434-LYS433	ASP29-ARG73	GLU451-LYS393
		ASP216-ARG217	GLU329-LYS367	ASP169-ARG291
		ASP281-ARG73	GLU307-LYS395	ASP61-ARG311
		ASP482-ARG97	ASP335-ARG370	ASP61-ARG390

Table A.5. All possible salt bridges of the monomers forms of WT and W318C mutant monomers at neutral pH (cont.).

WT	W318C	WT-R_5	W318C-R_5	W318C-ZINC
		ASP152-LYS123	ASP211-ARG143	
		ASP211-ARG143	GLU273-ARG19	
		GLU307-ARG311		

Table A.6. All hydrogen bond occupancies of the dimer forms at acidic pH.

WT -nonhyd					
Chain1		(%)	Chain2		(%)
ARG19	ASP216	92.39%	ARG19	GLU273	132.76%
ARG19	GLU273	64.57%	ARG19	ASP216	57.79%
ARG19	SPC20469	0.67%	ARG19	SPC23839	1.76%
ARG19	SPC2669	0.61%	ARG19	ASP216	1.75%
ASP29	LYS123	76.46%	ASP29	HIP227	83.60%
ASP29	HIP227	52.07%	ASP29	ARG73	0.72%
ASP29	SPC23746	1.38%	ASP29	ASP30	0.44%
ASP29	THR228	0.70%	ASP29	LYS123	0.44%
ASP61	THR60	64.74%	ASP61	ARG390	80.30%
ASP61	ARG390	17.82%	ASP61	ALA331	13.11%
ASP61	ARG311	5.77%	ASP61	SPC26007	7.36%
ASP61	SPC30098	1.41%	ASP61	SPC18456	6.14%
ASP61	SPC26350	0.69%	ASP61	SPC24386	4.19%
ASP61	SPC5387	0.58%	ASP61	SPC32888	4.14%
ARG73	ASP281	42.60%	ARG73	ASP281	91.72%
ARG73	SPC29223	39.98%	ARG73	SPC31353	88.75%
ARG73	TRP124	28.26%	ARG73	FGL69	47.22%
ARG73	FGL69	28.14%	ARG73	SER225	25.51%
ARG73	SPC1095	22.85%	ARG73	PHE27	2.80%
ARG73	LYS123	18.01%	ARG73	ASP29	2.20%

Table A.6. All hydrogen bond occupancies of the dimer forms at acidic pH (cont.).

WT -nonhyd					
Chain1		(%)	Chain2		(%)
ARG73	SER225	14.37%	ARG73	LEU77	2.03%
ASP169	ARG291	1.52%	ASP169	ARG291	32.98%
ASP169	SPC1386	1.07%	ASP169	SPC482	26.17%
ASP169	SPC32816	1.05%	ASP169	TYR230	21.96%
ASP169	SPC32816	0.95%	ASP169	TYR230	19.52%
ASP169	SPC30924	0.89%	ASP169	SPC446	11.33%
ASP169	SPC25540	0.79%	ASP169	SPC511	8.98%
ASP169	SPC30910	0.63%	ASP169	SPC21526	3.78%
GLY171	TYR230	13.16%	GLY171	ASH152	6.99%
GLY171	CYX156	0.60%	GLY171	CYX156	6.64%
GLY171	SPC22043	0.48%	GLY171	TYR230	1.83%
ASP173	ASH152	32.83%	ASP173	ASH152	17.91%
ASP173	GLY154	24.54%	ASP173	GLY154	1.57%
ASP173	GLN153	7.84%	ASP173	SPC5139	1.32%
GLU188	LEU181	110.55%	GLU188	LEU181	57.19%
GLU188	ARG200	10.30%	GLU188	ARG200	48.24%
GLU240	ASP249	87.94%	GLU240	ASP249	42.20%
GLU240	SPC23075	0.32%	GLU240	SPC586	1.32%
ARG241	SER237	22.19%	ARG241	SER237	32.13%
ARG241	ALA239	14.33%	ARG241	SPC30437	9.10%
ARG241	ASP249	5.93%	ARG241	ALA239	5.76%
ARG241	GLN236	1.25%	ARG241	GLN236	2.98%
GLU307	GLU437	114.71%	GLU307	LYS395	65.44%
GLU307	ARG311	19.42%	GLU307	TYR63	57.29%
GLU307	TYR63	6.56%	GLU307	ARG311	26.34%
GLU307	ARG311	3.75%	GLU307	GLU437	18.90%
GLU307	LYS395	3.71%	GLU307	ARG311	2.43%

Table A.6. All hydrogen bond occupancies of the dimer forms at acidic pH (cont.).

WT -nonhyd					
Chain1		(%)	Chain2		(%)
TRP318	SPC15448	44.65%	TRP318	GLU272	49.24%
TRP318	GLU272	21.09%	TRP318	ILE322	46.38%
TRP318	SPC3974	18.49%	TRP318	SPC12966	30.01%
TRP318	SPC10188	10.49%	TRP318	SPC33050	11.27%
TRP318	SPC33069	8.67%	TRP318	SPC11776	10.73%
GLU329	SER368	25.85%	GLU329	SER368	27.93%
GLU329	SPC23929	0.75%	GLU329	LYS367	6.71%
GLU329	SPC383	0.53%	GLU329	SPC30904	3.34%
GLU329	SPC20208	0.51%	GLU329	SPC2290	1.42%
ASP335	ARG370	45.31%	ASP335	ARG370	86.63%
ASP335	SER332	33.32%	ASP335	PHE356	24.86%
ASP335	SER332	18.98%	ASP335	SER332	6.39%
ASP335	SER332	3.09%	ASP335	THR339	5.32%
ASP335	THR339	1.04%	ASP335	THR339	1.47%
ASP335	PHE356	0.52%	ASP335	SER332	1.23%
LYS395	GLU307	162.24%	LYS395	GLU437	167.95%
LYS395	GLU437	107.76%	LYS395	GLU307	162.45%
LYS395	TYR429	88.67%	LYS395	THR305	84.77%
LYS395	THR305	26.04%	LYS395	TYR429	60.60%
LYS395	TYR429	0.23%	LYS395	TYR429	24.24%
PHE398	PHE387	19.99%	PHE398	PHE387	32.53%
PHE398	HIP423	6.70%	PHE398	HIP423	8.03%
ASP430	SER432	39.94%	ASP430	LYS393	43.56%
ASP430	LYS433	23.96%	ASP430	SPC3644	33.42%
ASP430	SPC15521	22.66%	ASP430	SER432	28.90%
ASP430	SER432	10.50%	ASP430	SPC24810	27.03%
ASP430	ASP434	9.20%	ASP430	LYS433	25.41%

Table A.6. All hydrogen bond occupancies of the dimer forms at acidic pH (cont.).

WT -nonhyd					
Chain1		(%)	Chain2		(%)
ASP430	SPC449	5.77%	ASP430	SPC29759	24.35%
ASP430	SPC10327	5.75%	ASP430	SPC16352	7.36%
ASP430	SPC26492	4.34%	ASP430	LYS433	7.34%
ASP430	SPC463	2.64%	ASP430	SPC31617	3.23%
ARG479	GLN476	27.94%	ARG479	PRO474	11.96%
ARG479	ARG97	21.47%	ARG479	GLN476	7.82%
ARG479	SER475	10.78%	ARG479	SER475	4.18%
ARG479	GLN476	8.52%	ARG479	SPC25932	2.05%
ARG479	GLU103	8.03%	ARG479	SPC6217	2.01%
ASP482	SER95	98.62%	ASP482	ARG97	29.05%
ASP482	ALA484	12.57%	ASP482	SER95	15.35%
ASP482	LEU485	0.84%	ASP482	SPC28209	1.53%
ASP482	SPC13640	0.44%	ASP482	SPC7746	1.50%
GLN460	LEU456	41.94%	GLN460	GLU424	30.67%
GLN460	GLU424	24.33%	GLN460	LEU456	12.36%
GLN460	LYS463	6.01%	GLN460	LEU456	2.70%
GLN460	ALA464	2.80%	GLN460	SPC13153	2.15%
W318C - nonhyd					
Chain1		(%)	Chain2		(%)
ARG19	GLU273	40.46%	ARG19	ASP216	65.94%
ARG19	ASP216	33.80%	ARG19	GLU273	51.46%
ARG19	GLU273	10.85%	ARG19	GLN213	13.48%
ARG19	ALA212	4.29%	ARG19	SPC12203	4.65%
ASP29	LYS123	72.61%	ASP29	HIP227	57.86%
ASP29	HIP227	41.41%	ASP29	ARG73	15.13%
ASP29	THR228	1.44%	ASP29	LYS123	0.17%
ASP29	ASP30	0.26%	ASP29	ASP30	0.12%

Table A.6. All hydrogen bond occupancies of the dimer forms at acidic pH (cont.).

W318C - nonhyd					
Chain1		(%)	Chain2		(%)
ASP61	THR60	60.83%	ASP61	ARG390	71.13%
ASP61	ARG311	2.86%	ASP61	SPC4259	39.84%
ASP61	ARG390	2.51%	ASP61	ALA331	13.09%
ASP61	SPC7150	1.02%	ASP61	SPC34073	6.04%
ASP61	SPC4411	1.01%	ASP61	SPC8102	3.92%
ASP61	SPC5906	1.00%	ASP61	SPC6566	2.88%
ARG73	ASP281	133.11%	ARG73	FGL69	80.43%
ARG73	SPC606	49.36%	ARG73	ASP29	76.59%
ARG73	FGL69	30.10%	ARG73	ASP281	46.81%
ARG73	SER225	23.24%	ARG73	SER225	28.04%
ARG73	SPC27364	22.67%	ARG73	LYS123	14.10%
ARG73	SPC32071	3.35%	ARG73	SPC29139	7.49%
ARG73	ASP29	3.33%	ARG73	SPC15144	5.08%
ASP169	ARG291	27.87%	ASP169	SPC485	46.85%
ASP169	SPC7207	24.32%	ASP169	TYR230	33.05%
ASP169	SPC34444	7.64%	ASP169	TYR230	32.68%
ASP169	SPC7207	7.23%	ASP169	SPC448	15.39%
ASP169	SPC29769	4.82%	ASP169	SPC32371	15.10%
ASP169	SPC27040	3.60%	ASP169	SPC492	12.10%
ASP169	SPC34444	2.18%	ASP169	SPC11955	11.89%
GLY171	CYX156	37.51%	GLY171	CYX156	17.06%
GLY171	ASP169	1.35%	GLY171	ASP169	0.48%
GLY171	GLN174	0.76%	GLY171	GLN174	0.36%
ASP173	ASH152	10.59%	ASP173	ASH152	18.02%
ASP173	GLY154	1.14%	ASP173	SPC19857	1.13%
ASP173	ASH152	0.85%	ASP173	SPC3955	0.68%
GLU188	LEU181	108.06%	GLU188	LEU181	55.82%

Table A.6. All hydrogen bond occupancies of the dimer forms at acidic pH (cont.).

W318C - nonhyd					
Chain1		(%)	Chain2		(%)
GLU188	ARG200	7.54%	GLU188	ARG200	51.28%
GLU240	ASP249	96.04%	GLU240	ASP249	49.59%
GLU240	SPC26866	0.25%	GLU240	SPC26109	1.31%
ARG241	SER237	52.39%	ARG241	SER237	84.43%
ARG241	ASP249	8.26%	ARG241	ASP249	3.57%
ARG241	ALA239	4.62%	ARG241	SPC20257	0.68%
ARG241	GLN236	0.84%	ARG241	PHE238	0.63%
GLU307	GLU437	145.54%	GLU307	LYS395	74.96%
GLU307	TYR63	16.20%	GLU307	GLU437	68.49%
GLU307	ARG311	12.62%	GLU307	TYR63	47.65%
GLU307	ARG311	7.99%	GLU307	ARG311	32.68%
GLU307	LYS395	6.57%	GLU307	ARG311	1.28%
CYS318	SPC24860	34.22%	CYS318	ILE322	36.56%
CYS318	SPC1584	28.61%	CYS318	SPC20723	16.30%
CYS318	SPC10815	27.30%	CYS318	SPC4797	13.99%
CYS318	SPC23702	15.01%	CYS318	SPC10202	8.53%
CYS318	SPC33000	4.70%	CYS318	SPC3158	6.29%
GLU329	SER368	36.27%	GLU329	SER368	21.99%
GLU329	SPC32870	1.75%	GLU329	LYS367	8.42%
GLU329	SPC8996	1.44%	GLU329	SPC22488	6.09%
GLU329	SPC11958	1.00%	GLU329	SPC3580	3.89%
ASP335	ARG370	70.10%	ASP335	ARG370	84.38%
ASP335	SER332	25.70%	ASP335	PHE356	36.38%
ASP335	SER332	14.81%	ASP335	THR339	9.07%
ASP335	SER332	11.08%	ASP335	SER332	5.99%
ASP335	THR339	0.71%	ASP335	SER332	5.10%
ASP335	PHE356	0.26%	ASP335	THR339	1.07%

Table A.6. All hydrogen bond occupancies of the dimer forms at acidic pH (cont.).

W318C - nonhyd					
Chain1		(%)	Chain2		(%)
LYS395	GLU307	161.12%	LYS395	GLU307	160.61%
LYS395	GLU437	133.03%	LYS395	GLU437	120.66%
LYS395	TYR429	96.62%	LYS395	THR305	85.11%
LYS395	THR305	48.15%	LYS395	TYR429	71.14%
LYS395	TYR429	3.04%	LYS395	TYR429	27.09%
PHE398	PHE387	29.29%	PHE398	PHE387	43.82%
PHE398	HIP423	10.92%	PHE398	HIP423	3.70%
ASP430	SPC465	30.90%	ASP430	SPC29606	38.11%
ASP430	SER432	24.96%	ASP430	LYS393	35.11%
ASP430	LYS433	16.01%	ASP430	SER432	29.37%
ASP430	SER432	13.95%	ASP430	LYS433	16.52%
ASP430	SPC16575	10.38%	ASP430	SPC4871	9.41%
ASP430	SPC19401	6.81%	ASP430	SPC29008	8.16%
ASP430	ASP434	4.88%	ASP430	SPC4286	7.65%
ASP430	SPC426	1.90%	ASP430	LYS433	7.58%
ASP430	SPC486	1.47%	ASP430	SPC19332	5.67%
ARG479	SPC23506	9.16%	ARG479	SPC28084	4.88%
ARG479	SPC18355	8.36%	ARG479	ARG97	3.70%
ARG479	SPC13481	6.84%	ARG479	SPC14597	3.50%
ARG479	GLN476	5.53%	ARG479	PRO474	3.24%
ARG479	SPC7284	5.12%	ARG479	SPC15444	2.85%
ASP482	SER95	71.68%	ASP482	SER95	48.93%
ASP482	ALA484	17.14%	ASP482	ARG97	30.93%
ASP482	LEU485	2.54%	ASP482	SPC35124	1.58%
ASP482	SPC16841	0.56%	ASP482	SPC5888	0.95%
GLN460	LEU456	55.32%	GLN460	LEU456	33.39%
GLN460	ALA464	16.25%	GLN460	GLU424	25.25%

Table A.6. All hydrogen bond occupancies of the dimer forms at acidic pH (cont.).

W318C - nonhyd					
Chain1		(%)	Chain2		(%)
GLN460	GLU424	15.26%	GLN460	SPC19963	0.79%
GLN460	SPC32340	0.72%	GLN460	SPC33848	0.63%
WT #2 - nonhyd					
Chain1		(%)	Chain2		(%)
ARG19	GLU273	66.23%	ARG19	GLU273	155.98%
ARG19	ASP216	58.42%	ARG19	ASP216	8.93%
ARG19	ASP216	46.61%	ARG19	GLU273	8.34%
ARG19	SPC26113	17.06%	ARG19	SPC350	1.27%
ASP29	LYS123	72.66%	ASP29	HIP227	74.94%
ASP29	HIP227	64.89%	ASP29	SPC29805	14.31%
ASP29	SPC23074	2.94%	ASP29	SPC34972	5.43%
ASP29	THR228	0.60%	ASP29	ARG73	4.73%
ASP61	THR60	72.57%	ASP61	ARG390	69.73%
ASP61	ARG390	25.62%	ASP61	ALA331	13.32%
ASP61	ARG311	6.80%	ASP61	SPC16808	9.34%
ASP61	SPC30052	2.78%	ASP61	SPC12263	6.65%
ASP61	SPC3683	1.25%	ASP61	SPC5013	6.13%
ASP61	SPC19635	1.04%	ASP61	SPC12432	4.59%
ARG73	ASP281	104.33%	ARG73	FGL69	60.89%
ARG73	SPC1090	80.42%	ARG73	SPC29805	44.29%
ARG73	FGL69	39.27%	ARG73	TRP124	41.86%
ARG73	SER225	22.26%	ARG73	SPC23013	40.77%
ARG73	ASP29	8.52%	ARG73	SPC9604	37.57%
ARG73	SPC27695	4.44%	ARG73	ASP29	31.30%
ARG73	ALA224	0.85%	ARG73	SPC32388	24.97%
ASP169	SPC30912	31.14%	ASP169	ARG291	48.33%
ASP169	ARG291	26.20%	ASP169	SPC461	2.85%

Table A.6. All hydrogen bond occupancies of the dimer forms at acidic pH (cont.).

WT #2 - nonhyd					
Chain1		(%)	Chain2		(%)
ASP169	SPC30912	8.60%	ASP169	TYR230	2.80%
ASP169	SPC29696	2.37%	ASP169	SPC23019	2.00%
ASP169	SPC29696	1.13%	ASP169	SPC22450	1.81%
ASP169	TYR230	0.59%	ASP169	SPC27665	1.57%
ASP169	SPC11638	0.54%	ASP169	SPC31660	1.57%
GLY171	CYX156	35.05%	GLY171	TYR230	9.36%
GLY171	GLN174	0.99%	GLY171	CYX156	0.47%
GLY171	ASP169	0.41%	GLY171	SPC12883	0.44%
ASP173	ASH152	16.77%	ASP173	ASH152	34.38%
ASP173	SPC23486	0.99%	ASP173	GLY154	5.84%
ASP173	SPC16163	0.88%	ASP173	GLN153	0.95%
GLU188	LEU181	105.82%	GLU188	LEU181	55.53%
GLU188	ARG200	10.37%	GLU188	ARG200	52.30%
GLU240	ASP249	91.68%	GLU240	ASP249	37.84%
GLU240	SPC6727	0.34%	GLU240	SPC22685	2.08%
ARG241	SER237	52.51%	ARG241	SER237	28.57%
ARG241	ASP249	8.41%	ARG241	ALA239	4.68%
ARG241	SPC34891	0.97%	ARG241	SPC3015	4.65%
ARG241	SPC9367	0.52%	ARG241	SPC2543	1.69%
GLU307	GLU437	121.66%	GLU307	LYS395	47.69%
GLU307	ARG311	21.80%	GLU307	TYR63	40.75%
GLU307	TYR63	13.40%	GLU307	ARG311	33.70%
GLU307	ARG311	7.34%	GLU307	GLU437	24.65%
GLU307	VAL310	2.35%	GLU307	VAL310	1.39%
TRP318	SPC2780	28.83%	TRP318	SPC11526	47.73%
TRP318	SPC2078	24.05%	TRP318	ILE322	44.32%
TRP318	GLU272	17.97%	TRP318	GLU272	21.18%

Table A.6. All hydrogen bond occupancies of the dimer forms at acidic pH (cont.).

WT #2 - nonhyd					
Chain1		(%)	Chain2		(%)
TRP318	SPC32682	15.91%	TRP318	SPC1996	12.87%
TRP318	SPC30995	9.72%	TRP318	SPC806	6.81%
GLU329	SER368	25.98%	GLU329	SER368	21.32%
GLU329	SPC3492	0.51%	GLU329	LYS367	11.15%
GLU329	SPC11138	0.39%	GLU329	SPC6806	2.01%
GLU329	SPC20014	0.38%	GLU329	SPC12174	1.43%
ASP335	SER332	51.40%	ASP335	ARG370	90.41%
ASP335	ARG370	38.79%	ASP335	PHE356	31.07%
ASP335	SER332	11.01%	ASP335	SER332	7.35%
ASP335	THR339	7.24%	ASP335	SER332	2.80%
ASP335	PHE356	5.62%	ASP335	THR339	1.79%
ASP335	SER332	1.88%	ASP335	THR339	0.86%
LYS395	GLU307	109.35%	LYS395	GLU307	146.67%
LYS395	TYR429	84.53%	LYS395	GLU437	146.30%
LYS395	GLU437	71.61%	LYS395	TYR429	65.45%
LYS395	THR305	21.92%	LYS395	THR305	32.96%
LYS395	TYR429	3.96%	LYS395	TYR429	11.78%
PHE398	PHE387	18.89%	PHE398	PHE387	35.56%
PHE398	HIP423	4.48%	PHE398	HIP423	5.43%
ASP430	SPC20343	76.02%	ASP430	SPC3755	35.45%
ASP430	SER432	34.12%	ASP430	SPC7417	31.88%
ASP430	LYS433	20.18%	ASP430	SPC4254	18.84%
ASP430	SER432	9.59%	ASP430	LYS393	16.44%
ASP430	LYS393	7.91%	ASP430	SER432	12.87%
ASP430	SPC474	2.60%	ASP430	SPC18501	8.17%
ASP430	SPC455	2.49%	ASP430	SPC30264	7.86%
ASP430	ASP434	1.86%	ASP430	SPC23974	3.96%

Table A.6. All hydrogen bond occupancies of the dimer forms at acidic pH (cont.).

WT #2 - nonhyd					
Chain1		(%)	Chain2		(%)
ASP430	LYS433	1.85%	ASP430	SPC26674	3.88%
ARG479	GLN476	21.73%	ARG479	GLN476	10.36%
ARG479	ARG97	6.70%	ARG479	SPC26924	3.63%
ARG479	SPC33662	5.28%	ARG479	SPC30422	3.41%
ARG479	SPC58	2.38%	ARG479	GLN476	3.35%
ARG479	SPC35048	2.17%	ARG479	SPC2128	2.99%
ASP482	ALA484	17.56%	ASP482	ARG97	40.21%
ASP482	SER95	15.26%	ASP482	SER95	20.94%
ASP482	ARG97	8.58%	ASP482	SPC23100	1.14%
ASP482	LEU485	2.51%	ASP482	SPC6174	1.08%
GLN460	LEU456	40.95%	GLN460	LEU456	28.99%
GLN460	GLU424	32.02%	GLN460	GLU424	20.50%
GLN460	LYS463	11.73%	GLN460	LEU456	2.41%
GLN460	LEU456	3.52%	GLN460	SPC18540	0.72%
C300F - nonhyd					
Chain1		(%)	Chain2		(%)
ARG19	ASP216	101.49%	ARG19	ASP216	109.31%
ARG19	SPC7159	7.35%	ARG19	GLU273	71.63%
ARG19	SPC13760	6.18%	ARG19	GLU273	5.14%
ARG19	SPC23943	5.22%	ARG19	SPC9665	0.98%
ASP29	LYS123	71.71%	ASP29	HIP227	54.23%
ASP29	HIP227	55.38%	ASP29	ARG73	10.45%
ASP29	SPC26607	19.44%	ASP29	SPC30056	7.40%
ASP29	THR228	1.11%	ASP29	SPC6008	3.56%
ASP61	THR60	68.77%	ASP61	ARG390	72.67%
ASP61	ARG390	12.83%	ASP61	SPC4404	27.59%
ASP61	ARG311	6.48%	ASP61	SPC17417	24.81%

Table A.6. All hydrogen bond occupancies of the dimer forms at acidic pH (cont.).

C300F - nonhyd					
Chain1		(%)	Chain2		(%)
ASP61	SPC9434	1.45%	ASP61	ALA331	17.00%
ASP61	SPC9402	1.40%	ASP61	SPC20965	15.92%
ASP61	ALA331	1.27%	ASP61	SPC35529	12.76%
ARG73	SPC19681	91.51%	ARG73	ASP29	57.35%
ARG73	FGL69	50.87%	ARG73	SPC30190	51.29%
ARG73	LYS123	43.05%	ARG73	FGL69	49.05%
ARG73	TRP124	32.98%	ARG73	SPC30056	20.88%
ARG73	SPC19355	30.43%	ARG73	SPC6008	19.32%
ARG73	SPC19344	26.26%	ARG73	LYS123	19.17%
ARG73	ASP281	21.52%	ARG73	TRP124	5.62%
ASP169	SPC30876	7.71%	ASP169	TYR230	70.30%
ASP169	SPC26666	7.11%	ASP169	SPC515	53.84%
ASP169	SPC21186	6.35%	ASP169	SPC514	37.79%
ASP169	SPC12952	6.00%	ASP169	ARG288	34.87%
ASP169	ARG291	5.29%	ASP169	ARG288	19.06%
ASP169	SPC12952	5.09%	ASP169	SPC514	1.81%
ASP169	SPC6103	4.77%	ASP169	TYR230	1.69%
GLY171	CYX156	37.82%	GLY171	SPC30521	2.60%
GLY171	ASP169	1.07%	GLY171	SPC6809	1.87%
GLY171	GLN174	0.56%	GLY171	SPC32345	1.62%
ASP173	SPC7521	0.98%	ASP173	GLY154	6.06%
ASP173	SPC5962	0.89%	ASP173	SPC24231	1.26%
ASP173	SPC16514	0.79%	ASP173	SPC20738	0.99%
GLU188	LEU181	112.90%	GLU188	LEU181	59.72%
GLU188	ARG200	10.10%	GLU188	ARG200	37.96%
GLU240	ASP249	101.01%	GLU240	ASP249	51.98%
GLU240	SPC2858	0.75%	GLU240	SPC19572	6.80%

Table A.6. All hydrogen bond occupancies of the dimer forms at acidic pH (cont.).

C300F - nonhyd					
Chain1		(%)	Chain2		(%)
ARG241	SER237	37.83%	ARG241	SER237	39.70%
ARG241	ALA239	5.67%	ARG241	ASP249	3.84%
ARG241	ASP249	5.17%	ARG241	PHE238	1.16%
ARG241	SPC23680	0.63%	ARG241	ALA239	1.13%
GLU307	GLU437	136.11%	GLU307	LYS395	67.16%
GLU307	ARG311	12.81%	GLU307	GLU437	65.15%
GLU307	TYR63	8.60%	GLU307	TYR63	48.01%
GLU307	ARG311	8.28%	GLU307	ARG311	28.37%
GLU307	LYS395	2.85%	GLU307	SPC26670	8.94%
TRP318	SPC10427	44.56%	TRP318	ILE322	42.36%
TRP318	SPC29096	37.25%	TRP318	SPC10528	38.95%
TRP318	SPC17740	17.60%	TRP318	GLU272	31.14%
TRP318	GLU272	14.29%	TRP318	SPC35052	10.34%
TRP318	SPC2675	0.12%	TRP318	SPC21213	4.64%
GLU329	SER368	20.89%	GLU329	SER368	24.65%
GLU329	SER368	2.06%	GLU329	SPC36028	2.78%
GLU329	SPC23650	0.98%	GLU329	LYS367	2.20%
GLU329	SPC27154	0.93%	GLU329	SPC14442	2.08%
ASP335	ARG370	58.92%	ASP335	ARG370	86.55%
ASP335	SER332	33.59%	ASP335	PHE356	30.37%
ASP335	SER332	8.00%	ASP335	SER332	6.77%
ASP335	THR339	6.90%	ASP335	SER332	5.46%
ASP335	SER332	3.41%	ASP335	THR339	2.52%
ASP335	SER332	0.15%	ASP335	THR339	2.27%
LYS395	GLU307	163.74%	LYS395	GLU307	166.40%
LYS395	GLU437	133.77%	LYS395	GLU437	98.92%
LYS395	TYR429	111.95%	LYS395	TYR429	76.78%

Table A.6. All hydrogen bond occupancies of the dimer forms at acidic pH (cont.).

C300F - nonhyd					
Chain1		(%)	Chain2		(%)
LYS395	THR305	34.16%	LYS395	THR305	44.92%
LYS395	TYR429	1.27%	LYS395	TYR429	12.99%
PHE398	PHE387	33.88%	PHE398	PHE387	46.46%
PHE398	HIP423	3.10%	PHE398	HIP423	6.30%
ASP430	SPC19752	92.50%	ASP430	LYS393	46.69%
ASP430	SER432	60.22%	ASP430	SPC30810	42.85%
ASP430	SER432	13.25%	ASP430	SER432	39.38%
ASP430	ASP434	6.92%	ASP430	SPC2859	13.38%
ASP430	LYS433	4.52%	ASP430	SPC17729	9.41%
ASP430	SPC473	2.73%	ASP430	SPC5297	6.81%
ASP430	SPC478	0.50%	ASP430	LYS433	6.06%
ASP430	SPC473	0.35%	ASP430	SPC30874	5.44%
ASP430	LYS393	0.18%	ASP430	SPC25712	5.23%
ARG479	SPC979	17.10%	ARG479	GLN476	1.68%
ARG479	GLN476	6.23%	ARG479	ALA478	1.41%
ARG479	ARG97	5.74%	ARG479	SPC7425	1.36%
ARG479	ALA478	5.73%	ARG479	ASP482	1.36%
ARG479	SPC24849	5.50%	ARG479	GLN476	1.27%
ASP482	SER95	73.70%	ASP482	ARG97	80.07%
ASP482	ALA484	14.79%	ASP482	SER95	30.02%
ASP482	ARG97	4.97%	ASP482	GLN486	8.47%
ASP482	LEU485	1.82%	ASP482	SPC17076	3.22%
GLN460	GLU424	49.48%	GLN460	GLU424	36.36%
GLN460	LEU456	41.93%	GLN460	LEU456	19.66%
GLN460	ALA464	8.64%	GLN460	LEU456	4.57%
GLN460	LEU456	2.71%	GLN460	LYS463	0.95%

Table A.6. All hydrogen bond occupancies of the dimer forms at acidic pH (cont.).

E307K+T391S - nonhyd					
Chain1		(%)	Chain2		(%)
ARG19	GLU273	124.29%	ARG19	ASP216	136.71%
ARG19	ASP216	20.50%	ARG19	GLU273	28.16%
ARG19	GLU273	9.75%	ARG19	ASP216	6.33%
ARG19	SPC27366	1.30%	ARG19	GLN213	4.23%
ASP29	LYS123	68.81%	ASP29	HIP227	78.45%
ASP29	HIP227	36.14%	ASP29	SPC29074	1.82%
ASP29	THR228	0.66%	ASP29	LYS123	0.98%
ASP29	SPC20940	0.37%	ASP29	ASP30	0.94%
ASP61	THR60	62.46%	ASP61	ARG390	41.50%
ASP61	ARG311	36.27%	ASP61	ARG311	23.66%
ASP61	SPC5915	7.77%	ASP61	ALA331	19.45%
ASP61	SPC28768	4.45%	ASP61	SPC4223	16.41%
ASP61	ARG390	4.15%	ASP61	TYR63	8.24%
ASP61	SPC29370	2.68%	ASP61	SPC17191	7.60%
ARG73	SPC30229	50.14%	ARG73	ASP281	90.57%
ARG73	FGL69	42.86%	ARG73	FGL69	87.33%
ARG73	SPC4578	32.42%	ARG73	SER225	31.23%
ARG73	TRP124	17.37%	ARG73	SPC30049	29.33%
ARG73	SPC1184	14.44%	ARG73	SPC30992	21.88%
ARG73	ASP281	10.90%	ARG73	SPC29083	7.16%
ARG73	ASP29	10.78%	ARG73	SPC29420	5.62%
ASP169	SPC29667	14.25%	ASP169	ARG291	43.53%
ASP169	SPC29667	13.96%	ASP169	TYR230	38.06%
ASP169	ARG291	7.94%	ASP169	SPC473	17.06%
ASP169	SPC32516	4.86%	ASP169	SPC21956	14.83%
ASP169	SPC14797	4.43%	ASP169	SPC6255	9.79%
ASP169	SPC32516	2.46%	ASP169	SPC12640	4.14%

Table A.6. All hydrogen bond occupancies of the dimer forms at acidic pH (cont.).

E307K+T391S - nonhyd					
Chain1		(%)	Chain2		(%)
ASP169	SPC14797	1.99%	ASP169	SPC4831	3.86%
GLY171	GLN174	16.69%	GLY171	CYX156	3.33%
GLY171	CYX156	2.99%	GLY171	ASP169	0.44%
GLY171	SPC8129	1.82%	GLY171	SPC18961	0.36%
ASP173	ASH152	52.79%	ASP173	ASH152	37.92%
ASP173	GLY154	15.15%	ASP173	GLY154	2.72%
ASP173	SPC881	2.04%	ASP173	SPC3043	1.32%
GLU188	LEU181	112.91%	GLU188	LEU181	58.46%
GLU188	ARG200	9.89%	GLU188	ARG200	44.85%
GLU240	ASP249	90.51%	GLU240	ASP249	51.62%
GLU240	SPC32825	0.36%	GLU240	LYS433	10.72%
ARG241	SER237	28.70%	ARG241	SER237	20.90%
ARG241	ALA239	12.48%	ARG241	SPC10941	9.54%
ARG241	SPC18623	7.77%	ARG241	ALA239	7.75%
ARG241	ASP249	3.47%	ARG241	GLN236	1.97%
LYN307	GLU437	42.65%	LYN307	GLU437	45.93%
LYN307	PRO435	38.09%	LYN307	TYR63	16.68%
LYN307	ARG311	15.91%	LYN307	ARG311	13.80%
LYN307	SPC464	8.01%	LYN307	LYS395	3.55%
LYN307	VAL310	2.91%	LYN307	GLU437	3.42%
TRP318	GLU272	49.83%	TRP318	ILE322	48.40%
TRP318	SPC31674	8.39%	TRP318	SPC7293	30.11%
TRP318	SPC11596	8.15%	TRP318	SPC29835	23.32%
TRP318	SPC16420	7.72%	TRP318	GLU272	18.98%
TRP318	SPC21989	6.76%	TRP318	SPC4085	2.92%
GLU329	SER368	6.44%	GLU329	SER368	29.60%
GLU329	SPC14534	0.66%	GLU329	SPC15613	6.08%

Table A.6. All hydrogen bond occupancies of the dimer forms at acidic pH (cont.).

E307K+T391S - nonhyd					
Chain1		(%)	Chain2		(%)
GLU329	SPC24526	0.55%	GLU329	SPC6557	5.64%
GLU329	SPC25708	0.43%	GLU329	LYS367	4.25%
ASP335	ARG370	65.99%	ASP335	ARG370	91.14%
ASP335	SER332	28.98%	ASP335	PHE356	20.82%
ASP335	SER332	21.61%	ASP335	SER332	16.54%
ASP335	SER332	2.61%	ASP335	THR339	0.71%
ASP335	THR339	0.68%	ASP335	THR339	7.00%
ASP335	PHE356	0.29%	ASP335	SER332	1.07%
LYS395	GLU437	108.49%	LYS395	GLU437	105.48%
LYS395	TYR429	96.20%	LYS395	THR305	77.32%
LYS395	SPC5977	49.80%	LYS395	TYR429	44.39%
LYS395	THR305	29.47%	LYS395	TYR429	24.70%
LYS395	SPC6289	5.04%	LYS395	SPC8829	23.77%
PHE398	PHE387	35.08%	PHE398	PHE387	43.11%
PHE398	SPC5045	0.22%	PHE398	HIP423	4.39%
ASP430	SPC1608	97.84%	ASP430	TYR439	87.79%
ASP430	LYS433	44.52%	ASP430	LYS433	46.01%
ASP430	SPC1608	37.57%	ASP430	SER432	33.24%
ASP430	ASH434	24.83%	ASP430	LYS393	11.33%
ASP430	SPC464	20.92%	ASP430	SPC32023	3.25%
ASP430	SER432	18.71%	ASP430	SPC1983	2.28%
ASP430	SPC29820	4.19%	ASP430	SPC5613	1.80%
ASP430	ASH434	2.13%	ASP430	SPC26543	1.72%
ASP430	TYR439	1.18%	ASP430	SPC32600	1.45%
ARG479	SER475	21.49%	ARG479	SER96	14.06%
ARG479	GLN476	6.48%	ARG479	GLN476	7.47%
ARG479	SPC22611	4.50%	ARG479	SER475	6.85%

Table A.6. All hydrogen bond occupancies of the dimer forms at acidic pH (cont.).

E307K+T391S - nonhyd					
Chain1		(%)	Chain2		(%)
ARG479	ARG97	2.12%	ARG479	PRO474	6.60%
ARG479	SPC3453	1.50%	ARG479	SPC26934	2.92%
ASP482	ALA484	25.43%	ASP482	ARG97	71.74%
ASP482	ARG97	4.68%	ASP482	SER95	22.07%
ASP482	LEU485	2.35%	ASP482	SPC1663	1.42%
ASP482	SER95	2.19%	ASP482	SPC3726	1.28%
GLN460	LEU456	47.43%	GLN460	LEU456	23.99%
GLN460	GLU424	9.82%	GLN460	GLU424	22.79%
GLN460	LYS463	5.81%	GLN460	SPC28962	0.54%
GLN460	ALA464	2.92%	GLN460	ALA464	0.39%

Table A.7. Possible salt bridge forming residues are listed for the dimers at acidic pH.

WT	
C1	C2
ASP61-ARG390	ASP30-HIS229
GLU307-LYS395	ASP430-LYS393
ASP354-ARG84	GLU329-LYS367
ASP381-ARG384	GLU103-ARG479
GLU285-LYS302	ASP169-ARG291
GLU329-LYS367	ASP207-ARG143
ASP35-ARG244	ASP29-ARG73
ASP169-ARG291	ASP281-ARG73
ASP430-LYS433	ASP430-LYS433
ASP207-ARG143	GLU131-ARG97
ASP430-LYS393	ASP61-ARG311
ASP50-ARG58_chainB	GLU285-LYS302

Table A.7. Possible salt bridge forming residues are listed for the dimers at acidic pH  
(cont.).

WT	
C1	C2
GLU451-LYS393	ASP381-ARG384
ASP216-ARG19	GLU437-LYS395
ASP411-ARG299	GLU307-ARG311
ASP29-HIS229	GLU329-LYS393
ASP482-ARG97	ASP30-LYS302
ASP211-ARG214	ASP211-ARG214
ASP30-LYS302	GLU451-LYS433
ASP29-LYS123	ASP29-LYS123
ASP207-ARG214	ASP354-ARG84
GLU437-LYS395	ASP211-ARG143
GLU131-ARG97	GLU104-ARG80
ASP30-HIS229	GLU451-LYS393
GLU188-ARG200	ASP482-ARG97
ASP335-ARG370	ASP335-ARG370
ASP211-ARG143	ASP216-ARG217
GLU103-ARG479	GLU285-ARG288
ASP29-ARG73	GLU273-ARG19
GLU307-ARG311	ASP216-ARG19
GLU104-ARG80	ASP467-LYS463
ASP61-ARG311	ASP207-ARG214
GLU273-ARG19	GLU188-ARG200
GLU424-LYS463	GLU424-LYS463
ASP467-LYS463	ASP411-ARG299
ASP281-ARG73	GLU307-LYS395
GLU451-LYS433	ASP61-ARG390
GLU285-ARG288	ASP35-ARG244
	ASP50-ARG58_chainA

Table A.7. Possible salt bridge forming residues are listed for the dimers at acidic pH  
(cont.).

W318C	
C1	C2
GLU307-ARG311	ASP207-ARG214
GLU188-ARG200	ASP430-LYS393
ASP30-LYS302	ASP482-ARG97
ASP354-ARG84	GLU285-LYS302
ASP335-ARG370	ASP169-ARG291
GLU329-LYS367	ASP35-ARG244
ASP430-LYS433	ASP61-ARG390
ASP430-LYS393	ASP207-ARG143
ASP207-ARG214	ASP381-ARG384
GLU240-ARG241	ASP216-ARG19
ASP411-ARG299	ASP354-ARG84
ASP216-ARG217	ASP211-ARG214
ASP61-ARG390	ASP30-LYS302
GLU285-ARG288	GLU437-LYS395
GLU307-LYS395	GLU451-LYS433
ASP381-ARG384	ASP29-LYS123
ASP61-ARG311	GLU131-ARG97
ASP482-ARG97	ASP335-ARG370
ASP207-ARG143	GLU329-LYS367
GLU437-LYS395	GLU451-LYS393
ASP35-ARG244	GLU188-ARG200
GLU424-LYS463	ASP211-ARG143
GLU451-LYS433	ASP30-HIS229
ASP30-HIS229	GLU424-LYS463
ASP29-ARG73	ASP29-ARG73
ASP29-LYS123	GLU307-ARG311
ASP211-ARG143	GLU104-ARG80

Table A.7. Possible salt bridge forming residues are listed for the dimers at acidic pH  
(cont.).

W318C	
C1	C2
GLU451-LYS393	ASP430-LYS433
ASP169-ARG291	ASP61-ARG311
ASP281-ARG73	ASP411-ARG299
ASP216-ARG19	ASP216-ARG217
GLU104-ARG80	GLU273-ARG19
ASP211-ARG214	GLU382-ARG384
GLU273-ARG19	GLU437-ARG299
GLU285-LYS302	ASP281-ARG73
ASP467-LYS463	GLU307-LYS395
ASP61-LYS393	GLU285-ARG288
GLU131-ARG97	ASP467-LYS463
WT-2	
C1	C2
GLU285-LYS302	ASP207-ARG214
ASP381-ARG384	ASP61-ARG390
GLU437-LYS395	GLU307-LYS395
GLU307-ARG311	ASP354-ARG84
ASP29-ARG73	ASP381-ARG384
ASP30-LYS302	GLU285-LYS302
ASP281-ARG73	ASP169-ARG291
ASP29-LYS123	ASP35-ARG244
ASP211-ARG143	GLU329-LYS367
ASP482-ARG97	ASP430-LYS433
GLU451-LYS433	ASP430-LYS393
GLU451-LYS393	ASP207-ARG143
ASP207-ARG143	ASP216-ARG19
ASP216-ARG217	GLU451-LYS393

Table A.7. Possible salt bridge forming residues are listed for the dimers at acidic pH  
(cont.).

WT-2	
C1	C2
ASP335-ARG370	ASP411-ARG299
GLU285-ARG288	ASP482-ARG97
ASP207-ARG214	ASP211-ARG214
GLU188-ARG200	ASP30-LYS302
GLU104-ARG80	ASP29-LYS123
ASP61-ARG311	GLU437-LYS395
GLU424-LYS463	GLU131-ARG97
ASP216-ARG19	GLU131-ARG496
GLU307-LYS395	GLU188-ARG200
ASP61-ARG390	ASP335-ARG370
GLU273-ARG19	ASP211-ARG143
ASP35-ARG244	ASP30-HIS229
ASP430-LYS433	ASP29-ARG73
ASP467-LYS463	GLU307-ARG311
ASP430-LYS393	GLU104-ARG80
ASP30-HIS229	ASP61-ARG311
GLU329-LYS367	ASP216-ARG217
ASP354-ARG84	GLU273-ARG19
ASP169-ARG291	GLU424-LYS463
GLU131-ARG97	ASP467-LYS463
GLU103-ARG479	ASP281-ARG73
ASP411-ARG299	GLU451-LYS433
ASP211-ARG214	GLU285-ARG288
	GLU240-ARG241

Table A.7. Possible salt bridge forming residues are listed for the dimers at acidic pH  
(cont.).

C300F	
C1	C2
GLU451-LYS393	ASP207-ARG143
ASP169-ARG288	ASP30-HIS229
GLU131-ARG97	ASP207-ARG214
ASP482-ARG97	ASP29-ARG73
ASP35-ARG244	ASP482-ARG479
ASP467-LYS463	GLU131-ARG97
ASP430-LYS433	ASP467-LYS463
ASP216-ARG19	ASP411-ARG299
ASP335-ARG370	ASP29-HIS229
ASP30-LYS302	ASP61-ARG311
ASP281-ARG73	ASP211-ARG143
ASP211-ARG214	GLU285-ARG288
ASP29-LYS123	GLU307-LYS395
ASP411-ARG299	ASP381-ARG384
ASP207-ARG214	GLU273-ARG19
GLU285-LYS302	ASP169-ARG291
GLU307-LYS395	ASP430-LYS433
ASP381-ARG384	GLU285-LYS302
GLU437-LYS395	ASP29-LYS123
GLU329-LYS367	ASP281-ARG73
ASP211-ARG143	ASP211-ARG214
ASP207-ARG143	ASP430-LYS393
ASP30-HIS229	ASP30-LYS302
ASP61-ARG311	GLU104-ARG80
ASP29-ARG73	ASP50-ARG58
GLU104-ARG80	GLU437-LYS395
GLU188-ARG200	ASP354-ARG84

Table A.7. Possible salt bridge forming residues are listed for the dimers at acidic pH  
(cont.).

C300F	
C1	C2
GLU424-LYS463	ASP169-ARG288
ASP169-ARG291	ASP61-ARG390
GLU285-ARG288	ASP335-ARG370
ASP61-ARG390	GLU451-LYS433
ASP430-LYS393	GLU307-ARG311
ASP354-ARG84	ASP482-ARG97
GLU307-ARG311	GLU451-LYS393
	GLU424-LYS463
	GLU481-ARG479
	GLU188-ARG200
	GLU240-ARG241
	GLU329-LYS367
	ASP35-ARG244
	ASP216-ARG19
E307K+T391S	
C1	C2
ASP30-LYS302	GLU329-LYS367
ASP169-ARG288	ASP354-ARG84
GLU188-ARG200	ASP430-LYS393
ASP411-ARG299	GLU131-ARG97
ASP381-ARG384	ASP35-ARG244
ASP50-ARG58_chainB	ASP381-ARG384
GLU285-LYS302	ASP169-ARG288
GLU424-LYS463	ASP216-ARG19
ASP29-ARG73	ASP61-ARG390
GLU131-ARG97	ASP29-HIS229
ASP61-ARG311	ASP467-LYS463

Table A.7. Possible salt bridge forming residues are listed for the dimers at acidic pH  
(cont.).

E307K+T391S	
C1	C2
ASP211-ARG214	GLU273-ARG19
ASP207-ARG214	ASP281-ARG73
ASP335-ARG370	ASP211-ARG214
ASP29-LYS123	ASP30-LYS302
ASP430-LYS433	ASP29-LYS123
ASP61-ARG390	ASP430-LYS433
GLU104-ARG80	ASP207-ARG214
GLU240_chainB-LYS433	GLU285-LYS302
ASP281-ARG73	ASP482-ARG97
ASP50_chainB-ARG58	ASP30-HIS229
ASP35-ARG244	ASP211-ARG143
ASP216-ARG217	ASP61-ARG311
ASP354-ARG84	ASP207-ARG143
GLU273-ARG19	GLU424-LYS463
GLU329-LYS367	ASP29-ARG73
ASP216-ARG19	GLU104-ARG80
GLU451-LYS393	ASP335-ARG370
GLU437-LYS395	ASP411-ARG299
GLU272-ARG19	GLU437-LYS395
ASP30-HIS229	GLU188-ARG200
ASP169-ARG291	GLU451-LYS393
ASP207-ARG143	ASP169-ARG291
GLU285-ARG288	GLU285-ARG288
ASP211-ARG143	GLU481-ARG97
ASP467-LYS463	
ASP482-ARG97	

## APPENDIX B: DOCKING CALCULATIONS AND PHYSICOCHEMICAL PROPERTIES OF THE MOLECULES

Table B.1. Converted IC<sub>50</sub> and pIC<sub>50</sub> values of the known bioactive molecules of ARSA are listed in the decreasing order of their PIC<sub>50</sub> values.

CHEMBL ID	STD VALUE	STD UNITS	IC <sub>50</sub> (M)	pIC <sub>50</sub>
CHEMBL219916	5.4	nM	5.4E-09	8.268
CHEMBL1729444	5.4	nM	5.4E-09	8.268
CHEMBL1407943	5.4	nM	5.4E-09	8.268
CHEMBL86304	5.4	nM	5.4E-09	8.268
CHEMBL501701	5.4	nM	5.4E-09	8.268
CHEMBL391997	5.4	nM	5.4E-09	8.268
CHEMBL273648	5.4	nM	5.4E-09	8.268
CHEMBL266195	5.4	nM	5.4E-09	8.268
CHEMBL1256759	5.4	nM	5.4E-09	8.268
CHEMBL37744	5.4	nM	5.4E-09	8.268
CHEMBL93247	5.4	nM	5.4E-09	8.268
CHEMBL13378	5.4	nM	5.4E-09	8.268
CHEMBL1256175	5.4	nM	5.4E-09	8.268
CHEMBL1609891	6	nM	6.0E-09	8.222
CHEMBL1256667	7.6	nM	7.6E-09	8.119
CHEMBL3199475	10.7	nM	1.07E-08	7.971
CHEMBL814	10.7	nM	1.07E-08	7.971
CHEMBL239243	10.7	nM	1.07E-08	7.971
CHEMBL576127	12	nM	1.2E-08	7.921
CHEMBL505	12	nM	1.2E-08	7.921
CHEMBL35482	13.5	nM	1.35E-08	7.870
CHEMBL679	13.5	nM	1.35E-08	7.870
CHEMBL76768	13.5	nM	1.35E-08	7.870
CHEMBL844	15.1	nM	1.51E-08	7.821

Table B.1. Converted IC<sub>50</sub> and pIC<sub>50</sub> values of the known bioactive molecules of ARSA are listed in the decreasing order of their PIC<sub>50</sub> values (cont.).

CHEMBL ID	STD VALUE	STD UNITS	IC <sub>50</sub> (M)	pIC <sub>50</sub>
CHEMBL274619	16.9	nM	1.69E-08	7.772
CHEMBL455983	16.9	nM	1.69E-08	7.772
CHEMBL454761	16.9	nM	1.69E-08	7.772
CHEMBL282199	19	nM	1.9E-08	7.721
CHEMBL3392362	19	nM	1.9E-08	7.721
CHEMBL85139	21.3	nM	2.13E-08	7.672
CHEMBL315157	21.3	nM	2.13E-08	7.672
CHEMBL614	21.3	nM	2.13E-08	7.672
CHEMBL63	23.9	nM	2.39E-08	7.622
CHEMBL1256283	26.9	nM	2.69E-08	7.570
CHEMBL302840	26.9	nM	2.69E-08	7.570
CHEMBL53929	26.9	nM	2.69E-08	7.570
CHEMBL723	26.9	nM	2.69E-08	7.570
CHEMBL1255941	30.1	nM	3.01E-08	7.521
CHEMBL262083	30.1	nM	3.01E-08	7.521
CHEMBL419815	30.1	nM	3.01E-08	7.521
CHEMBL468	30.1	nM	3.01E-08	7.521
CHEMBL1512503	30.1	nM	3.01E-08	7.521
CHEMBL1456087	30.1	nM	3.01E-08	7.521
CHEMBL182	30.1	nM	3.01E-08	7.521
CHEMBL407874	33.8	nM	3.38E-08	7.471
CHEMBL61593	33.8	nM	3.38E-08	7.471
CHEMBL274440	33.8	nM	3.38E-08	7.471
CHEMBL285755	33.8	nM	3.38E-08	7.471
CHEMBL275006	33.8	nM	3.38E-08	7.471
CHEMBL1369461	33.8	nM	3.38E-08	7.471
CHEMBL45891	37.9	nM	3.79E-08	7.421
CHEMBL458273	37.9	nM	3.79E-08	7.421

Table B.1. Converted IC<sub>50</sub> and pIC<sub>50</sub> values of the known bioactive molecules of ARSA are listed in the decreasing order of their PIC<sub>50</sub> values (cont.).

CHEMBL ID	STD VALUE	STD UNITS	IC <sub>50</sub> (M)	pIC <sub>50</sub>
CHEMBL1474884	42.6	nM	4.26E-08	7.371
CHEMBL134	42.6	nM	4.26E-08	7.371
CHEMBL15976	47.8	nM	4.78E-08	7.321
CHEMBL122270	47.8	nM	4.78E-08	7.321
CHEMBL1062	47.8	nM	4.78E-08	7.321
CHEMBL1527442	47.8	nM	4.78E-08	7.321
CHEMBL473104	53.6	nM	5.36E-08	7.271
CHEMBL347862	60.1	nM	6.01E-08	7.221
CHEMBL287045	60.1	nM	6.01E-08	7.221
CHEMBL3085495	60.1	nM	6.01E-08	7.221
CHEMBL1337170	60.1	nM	6.01E-08	7.221
CHEMBL185073	67.5	nM	6.75E-08	7.171
CHEMBL265301	67.5	nM	6.75E-08	7.171
CHEMBL132722	67.5	nM	6.75E-08	7.171
CHEMBL34095	75.7	nM	7.57E-08	7.121
CHEMBL1542	75.7	nM	7.57E-08	7.121
CHEMBL60518	75.7	nM	7.57E-08	7.121
CHEMBL431504	75.7	nM	7.57E-08	7.121
CHEMBL18686	75.7	nM	7.57E-08	7.121
CHEMBL1619272	75.7	nM	7.57E-08	7.121
CHEMBL103	75.7	nM	7.57E-08	7.121
CHEMBL1334976	75.7	nM	7.57E-08	7.121
CHEMBL1318812	84.9	nM	8.49E-08	7.071
CHEMBL609245	84.9	nM	8.49E-08	7.071
CHEMBL50267	84.9	nM	8.49E-08	7.071
CHEMBL16410	84.9	nM	8.49E-08	7.071
CHEMBL1329438	84.9	nM	8.49E-08	7.071
CHEMBL1079460	84.9	nM	8.49E-08	7.071

Table B.1. Converted IC<sub>50</sub> and pIC<sub>50</sub> values of the known bioactive molecules of ARSA are listed in the decreasing order of their PIC<sub>50</sub> values (cont.).

CHEMBL ID	STD VALUE	STD UNITS	IC <sub>50</sub> (M)	pIC <sub>50</sub>
CHEMBL1114	84.9	nM	8.49E-08	7.071
CHEMBL136737	95.3	nM	9.53E-08	7.021
CHEMBL7890	95.3	nM	9.53E-08	7.021
CHEMBL13852	95.3	nM	9.53E-08	7.021
CHEMBL1439172	95.3	nM	9.53E-08	7.021
CHEMBL407	95.3	nM	9.53E-08	7.021
CHEMBL585	106.9	nM	1.069E-07	6.971
CHEMBL1518938	106.9	nM	1.069E-07	6.971
CHEMBL1256841	106.9	nM	1.069E-07	6.971
CHEMBL250270	120	nM	1.2E-07	6.921
CHEMBL240390	120	nM	1.2E-07	6.921
CHEMBL97760	120	nM	1.2E-07	6.921
CHEMBL153057	120	nM	1.2E-07	6.921
CHEMBL435	151	nM	1.52E-07	6.821
CHEMBL1451377	151	nM	1.52E-07	6.821
CHEMBL698	151	nM	1.52E-07	6.821
CHEMBL279898	151	nM	1.52E-07	6.821
CHEMBL365809	151	nM	1.52E-07	6.821
CHEMBL175	151	nM	1.52E-07	6.821
CHEMBL8320	151	nM	1.52E-07	6.821
CHEMBL672	151	nM	1.52E-07	6.821
CHEMBL78150	169.4	nM	1.694E-07	6.771
CHEMBL7393	169.4	nM	1.694E-07	6.771
CHEMBL344127	169.4	nM	1.694E-07	6.771
CHEMBL15564	169.4	nM	1.694E-07	6.771
CHEMBL513	169.4	nM	1.694E-07	6.771
CHEMBL285629	190.1	nM	1.901E-07	6.721
CHEMBL87045	190.1	nM	1.901E-07	6.721

Table B.1. Converted IC<sub>50</sub> and pIC<sub>50</sub> values of the known bioactive molecules of ARSA are listed in the decreasing order of their PIC<sub>50</sub> values (cont.).

CHEMBL ID	STD VALUE	STD UNITS	IC <sub>50</sub> (M)	pIC <sub>50</sub>
CHEMBL517	190.1	nM	1.901E-07	6.721
CHEMBL67378	190.1	nM	1.901E-07	6.721
CHEMBL23330	190.1	nM	1.901E-07	6.721
CHEMBL16	213.3	nM	2.133E-07	6.671
CHEMBL415	213.3	nM	2.133E-07	6.671
CHEMBL457530	213.3	nM	2.133E-07	6.671
CHEMBL559612	213.3	nM	2.133E-07	6.671
CHEMBL998	213.3	nM	2.133E-07	6.671
CHEMBL337309	239.3	nM	2.393E-07	6.621
CHEMBL524	268.5	nM	2.685E-07	6.571
CHEMBL1088844	268.5	nM	2.685E-07	6.571
CHEMBL499	268.5	nM	2.685E-07	6.571
CHEMBL285843	268.5	nM	2.685E-07	6.571
CHEMBL1201754	301.3	nM	3.013E-07	6.521
CHEMBL1204876	301.3	nM	3.013E-07	6.521
CHEMBL278020	301.3	nM	3.013E-07	6.521
CHEMBL1194	301.3	nM	3.013E-07	6.521
CHEMBL568150	301.3	nM	3.013E-07	6.521
CHEMBL1201308	301.3	nM	3.013E-07	6.521
CHEMBL1160508	338.1	nM	3.381E-07	6.471
CHEMBL1046	338.1	nM	3.381E-07	6.471
CHEMBL1593558	379.3	nM	3.793E-07	6.421
CHEMBL645	379.3	nM	3.793E-07	6.421
CHEMBL41140	379.3	nM	3.793E-07	6.421
CHEMBL1485629	379.3	nM	3.793E-07	6.421
CHEMBL418333	379.3	nM	3.793E-07	6.421
CHEMBL165	379.3	nM	3.793E-07	6.421
CHEMBL1109	379.3	nM	3.793E-07	6.421

Table B.1. Converted IC<sub>50</sub> and pIC<sub>50</sub> values of the known bioactive molecules of ARSA are listed in the decreasing order of their PIC<sub>50</sub> values (cont.).

CHEMBL ID	STD VALUE	STD UNITS	IC <sub>50</sub> (M)	pIC <sub>50</sub>
CHEMBL3392493	379.3	nM	3.793E-07	6.421
CHEMBL386389	425.6	nM	4.256E-07	6.371
CHEMBL20799	425.6	nM	4.256E-07	6.371
CHEMBL87992	425.6	nM	4.256E-07	6.371
CHEMBL601364	477.5	nM	4.775E-07	6.321
CHEMBL69710	477.5	nM	4.775E-07	6.321
CHEMBL258805	477.5	nM	4.775E-07	6.321
CHEMBL1428	477.5	nM	4.775E-07	6.321
CHEMBL1562	535.8	nM	5.358E-07	6.271
CHEMBL356703	535.8	nM	5.358E-07	6.271
CHEMBL6640	601.2	nM	6.012E-07	6.221
CHEMBL753	601.2	nM	6.012E-07	6.221
CHEMBL1725279	601.2	nM	6.012E-07	6.221
CHEMBL1553629	674.6	nM	6.746E-07	6.171
CHEMBL93655	674.6	nM	6.746E-07	6.171
CHEMBL1519	674.6	nM	6.746E-07	6.171
CHEMBL39879	756.9	nM	7.569E-07	6.121
CHEMBL1618272	756.9	nM	7.569E-07	6.121
CHEMBL515679	756.9	nM	7.569E-07	6.121
CHEMBL570383	756.9	nM	7.569E-07	6.121
CHEMBL1266	756.9	nM	7.569E-07	6.121
CHEMBL8747	756.9	nM	7.569E-07	6.121
CHEMBL29292	849.2	nM	8.492E-07	6.071
CHEMBL33086	849.2	nM	8.492E-07	6.071
CHEMBL1256814	849.2	nM	8.492E-07	6.071
CHEMBL611	849.2	nM	8.492E-07	6.071
CHEMBL1079905	849.2	nM	8.492E-07	6.071
CHEMBL89598	849.2	nM	8.492E-07	6.071

Table B.1. Converted IC<sub>50</sub> and pIC<sub>50</sub> values of the known bioactive molecules of ARSA are listed in the decreasing order of their PIC<sub>50</sub> values (cont.).

CHEMBL ID	STD VALUE	STD UNITS	IC <sub>50</sub> (M)	pIC <sub>50</sub>
CHEMBL8066	849.2	nM	8.492E-07	6.071
CHEMBL8327	849.2	nM	8.492E-07	6.071
CHEMBL286204	849.2	nM	8.492E-07	6.071
CHEMBL2130690	849.2	nM	8.492E-07	6.071
CHEMBL212579	952.8	nM	9.528E-07	6.021
CHEMBL1322392	952.8	nM	9.528E-07	6.021
CHEMBL726	952.8	nM	9.528E-07	6.021
CHEMBL1188627	952.8	nM	9.528E-07	6.021
CHEMBL134752	952.8	nM	9.528E-07	6.021
CHEMBL20936	952.8	nM	9.528E-07	6.021
CHEMBL841	952.8	nM	9.528E-07	6.021
CHEMBL250711	952.8	nM	9.528E-07	6.021
CHEMBL148	952.8	nM	9.528E-07	6.021
CHEMBL1405	952.8	nM	9.528E-07	6.021
CHEMBL15023	952.8	nM	9.528E-07	6.021
CHEMBL1591046	1069.1	nM	1.0691E-06	5.971
CHEMBL90621	1069.1	nM	1.0691E-06	5.971
CHEMBL1257030	1069.1	nM	1.0691E-06	5.971
CHEMBL26736	1069.1	nM	1.0691E-06	5.971
CHEMBL1451	1069.1	nM	1.0691E-06	5.971
CHEMBL1255578	1069.1	nM	1.0691E-06	5.971
CHEMBL139000	1069.1	nM	1.0691E-06	5.971
CHEMBL360055	1069.1	nM	1.0691E-06	5.971
CHEMBL191044	1069.1	nM	1.0691E-06	5.971
CHEMBL280065	1069.1	nM	1.0691E-06	5.971
CHEMBL1231533	1199.5	nM	1.1995E-06	5.921
CHEMBL327708	1199.5	nM	1.1995E-06	5.921
CHEMBL1236602	1199.5	nM	1.1995E-06	5.921

Table B.1. Converted IC<sub>50</sub> and pIC<sub>50</sub> values of the known bioactive molecules of ARSA are listed in the decreasing order of their PIC<sub>50</sub> values (cont.).

CHEMBL ID	STD VALUE	STD UNITS	IC <sub>50</sub> (M)	pIC <sub>50</sub>
CHEMBL780	1199.5	nM	1.1995E-06	5.921
CHEMBL279556	1345.9	nM	1.3459E-06	5.871
CHEMBL1315261	1345.9	nM	1.3459E-06	5.871
CHEMBL1256746	1345.9	nM	1.3459E-06	5.871
CHEMBL8809	1345.9	nM	1.3459E-06	5.871
CHEMBL100622	1510.1	nM	1.5101E-06	5.821
CHEMBL1274	1510.1	nM	1.5101E-06	5.821
CHEMBL268868	1510.1	nM	1.5101E-06	5.821
CHEMBL299052	1694.4	nM	1.6944E-06	5.771
CHEMBL425686	1694.4	nM	1.6944E-06	5.771
CHEMBL1256391	1694.4	nM	1.6944E-06	5.771
CHEMBL1607704	1694.4	nM	1.6944E-06	5.771
CHEMBL637	1694.4	nM	1.6944E-06	5.771
CHEMBL267373	1694.4	nM	1.6944E-06	5.771
CHEMBL25147	1694.4	nM	1.6944E-06	5.771
CHEMBL1556654	1694.4	nM	1.6944E-06	5.771
CHEMBL489156	1694.4	nM	1.6944E-06	5.771
CHEMBL351464	1901.2	nM	1.9012E-06	5.721
CHEMBL1201287	1901.2	nM	1.9012E-06	5.721
CHEMBL1473475	1901.2	nM	1.9012E-06	5.721
CHEMBL2448607	1901.2	nM	1.9012E-06	5.721
CHEMBL18132	1901.2	nM	1.9012E-06	5.721
CHEMBL189382	1901.2	nM	1.9012E-06	5.721
CHEMBL1336469	1901.2	nM	1.9012E-06	5.721
CHEMBL1320902	1901.2	nM	1.9012E-06	5.721
CHEMBL1159717	1901.2	nM	1.9012E-06	5.721
CHEMBL1256693	2133.1	nM	2.1331E-06	5.671
CHEMBL570	2133.1	nM	2.1331E-06	5.671

Table B.1. Converted IC<sub>50</sub> and pIC<sub>50</sub> values of the known bioactive molecules of ARSA are listed in the decreasing order of their PIC<sub>50</sub> values (cont.).

CHEMBL ID	STD VALUE	STD UNITS	IC <sub>50</sub> (M)	pIC <sub>50</sub>
CHEMBL26998	2133.1	nM	2.1331E-06	5.671
CHEMBL56543	2133.1	nM	2.1331E-06	5.671
CHEMBL289277	2133.1	nM	2.1331E-06	5.671
CHEMBL1710306	2133.1	nM	2.1331E-06	5.671
CHEMBL27	2133.1	nM	2.1331E-06	5.671
CHEMBL1181010	2133.1	nM	2.1331E-06	5.671
CHEMBL135	2393.4	nM	2.3934E-06	5.621
CHEMBL15245	2393.4	nM	2.3934E-06	5.621
CHEMBL16687	2393.4	nM	2.3934E-06	5.621
CHEMBL557555	2393.4	nM	2.3934E-06	5.621
CHEMBL45281	2393.4	nM	2.3934E-06	5.621
CHEMBL611494	2393.4	nM	2.3934E-06	5.621
CHEMBL1256390	2393.4	nM	2.3934E-06	5.621
CHEMBL182461	2393.4	nM	2.3934E-06	5.621
CHEMBL295038	2393.4	nM	2.3934E-06	5.621
CHEMBL330129	2393.4	nM	2.3934E-06	5.621
CHEMBL28	2393.4	nM	2.3934E-06	5.621
CHEMBL273481	2393.4	nM	2.3934E-06	5.621
CHEMBL3	2393.4	nM	2.3934E-06	5.621
CHEMBL1256760	2393.4	nM	2.3934E-06	5.621
CHEMBL31088	2393.4	nM	2.3934E-06	5.621
CHEMBL286722	2393.4	nM	2.3934E-06	5.621
CHEMBL2449003	2393.4	nM	2.3934E-06	5.621
CHEMBL45244	2393.4	nM	2.3934E-06	5.621
CHEMBL1169388	2685.5	nM	2.6855E-06	5.571
CHEMBL297362	2685.5	nM	2.6855E-06	5.571
CHEMBL914	2685.5	nM	2.6855E-06	5.571
CHEMBL18840	2685.5	nM	2.6855E-06	5.571

Table B.1. Converted IC<sub>50</sub> and pIC<sub>50</sub> values of the known bioactive molecules of ARSA are listed in the decreasing order of their PIC<sub>50</sub> values (cont.).

CHEMBL ID	STD VALUE	STD UNITS	IC <sub>50</sub> (M)	pIC <sub>50</sub>
CHEMBL817	2685.5	nM	2.6855E-06	5.571
CHEMBL334378	2685.5	nM	2.6855E-06	5.571
CHEMBL1190	2685.5	nM	2.6855E-06	5.571
CHEMBL1199253	2685.5	nM	2.6855E-06	5.571
CHEMBL13791	2685.5	nM	2.6855E-06	5.571
CHEMBL23507	2685.5	nM	2.6855E-06	5.571
CHEMBL1256836	2685.5	nM	2.6855E-06	5.571
CHEMBL564	2685.5	nM	2.6855E-06	5.571
CHEMBL64239	2685.5	nM	2.6855E-06	5.571
CHEMBL45068	2685.5	nM	2.6855E-06	5.571
CHEMBL33720	2685.5	nM	2.6855E-06	5.571
CHEMBL2107804	2685.5	nM	2.6855E-06	5.571
CHEMBL1568698	3013.1	nM	3.0131E-06	5.521
CHEMBL90472	3013.1	nM	3.0131E-06	5.521
CHEMBL965	3013.1	nM	3.0131E-06	5.521
CHEMBL629	3013.1	nM	3.0131E-06	5.521
CHEMBL1097	3013.1	nM	3.0131E-06	5.521
CHEMBL376505	3013.1	nM	3.0131E-06	5.521
CHEMBL1256623	3013.1	nM	3.0131E-06	5.521
CHEMBL856	3013.1	nM	3.0131E-06	5.521
CHEMBL466465	3013.1	nM	3.0131E-06	5.521
CHEMBL300555	3013.1	nM	3.0131E-06	5.521
CHEMBL49247	3013.1	nM	3.0131E-06	5.521
CHEMBL1320667	3013.1	nM	3.0131E-06	5.521
CHEMBL1096979	3380.8	nM	3.3808E-06	5.471
CHEMBL1515756	3380.8	nM	3.3808E-06	5.471
CHEMBL1516111	3380.8	nM	3.3808E-06	5.471
CHEMBL25236	3380.8	nM	3.3808E-06	5.471

Table B.1. Converted IC<sub>50</sub> and pIC<sub>50</sub> values of the known bioactive molecules of ARSA are listed in the decreasing order of their PIC<sub>50</sub> values (cont.).

CHEMBL ID	STD VALUE	STD UNITS	IC <sub>50</sub> (M)	pIC <sub>50</sub>
CHEMBL15192	3380.8	nM	3.3808E-06	5.471
CHEMBL95	3380.8	nM	3.3808E-06	5.471
CHEMBL111583	3380.8	nM	3.3808E-06	5.471
CHEMBL267548	3380.8	nM	3.3808E-06	5.471
CHEMBL305881	3380.8	nM	3.3808E-06	5.471
CHEMBL166551	3380.8	nM	3.3808E-06	5.471
CHEMBL559934	3380.8	nM	3.3808E-06	5.471
CHEMBL1433	3793.3	nM	3.7933E-06	5.421
CHEMBL104	3793.3	nM	3.7933E-06	5.421
CHEMBL8618	3793.3	nM	3.7933E-06	5.421
CHEMBL258405	3793.3	nM	3.7933E-06	5.421
CHEMBL1489	3793.3	nM	3.7933E-06	5.421
CHEMBL279998	3793.3	nM	3.7933E-06	5.421
CHEMBL1256898	3793.3	nM	3.7933E-06	5.421
CHEMBL54	4256.2	nM	4.2562E-06	5.371
CHEMBL1398677	4256.2	nM	4.2562E-06	5.371
CHEMBL356388	4256.2	nM	4.2562E-06	5.371
CHEMBL51149	4256.2	nM	4.2562E-06	5.371
CHEMBL45816	4256.2	nM	4.2562E-06	5.371
CHEMBL265301	4256.2	nM	4.2562E-06	5.371
CHEMBL338790	4256.2	nM	4.2562E-06	5.371
CHEMBL1330596	4256.2	nM	4.2562E-06	5.371
CHEMBL39	4256.2	nM	4.2562E-06	5.371
CHEMBL4	4256.2	nM	4.2562E-06	5.371
CHEMBL8514	4256.2	nM	4.2562E-06	5.371
CHEMBL1396221	4256.2	nM	4.2562E-06	5.371
CHEMBL445102	4256.2	nM	4.2562E-06	5.371
CHEMBL39493	4256.2	nM	4.2562E-06	5.371

Table B.1. Converted IC<sub>50</sub> and pIC<sub>50</sub> values of the known bioactive molecules of ARSA are listed in the decreasing order of their PIC<sub>50</sub> values (cont.).

CHEMBL ID	STD VALUE	STD UNITS	IC <sub>50</sub> (M)	pIC <sub>50</sub>
CHEMBL485832	4256.2	nM	4.2562E-06	5.371
CHEMBL267930	4256.2	nM	4.2562E-06	5.371
CHEMBL1255647	4256.2	nM	4.2562E-06	5.371
CHEMBL26	4775.5	nM	4.7755E-06	5.321
CHEMBL284895	4775.5	nM	4.7755E-06	5.321
CHEMBL1591918	4775.5	nM	4.7755E-06	5.321
CHEMBL409	4775.5	nM	4.7755E-06	5.321
CHEMBL1358	4775.5	nM	4.7755E-06	5.321
CHEMBL1371	4775.5	nM	4.7755E-06	5.321
CHEMBL112816	4775.5	nM	4.7755E-06	5.321
CHEMBL1201213	4775.5	nM	4.7755E-06	5.321
CHEMBL477	4775.5	nM	4.7755E-06	5.321
CHEMBL36633	4775.5	nM	4.7755E-06	5.321
CHEMBL903	4775.5	nM	4.7755E-06	5.321
CHEMBL1473269	4775.5	nM	4.7755E-06	5.321
CHEMBL225230	4775.5	nM	4.7755E-06	5.321
CHEMBL1256480	4775.5	nM	4.7755E-06	5.321
CHEMBL1413406	5358.2	nM	5.3582E-06	5.271
CHEMBL18238	5358.2	nM	5.3582E-06	5.271
CHEMBL189	5358.2	nM	5.3582E-06	5.271
CHEMBL284609	5358.2	nM	5.3582E-06	5.271
CHEMBL289480	5358.2	nM	5.3582E-06	5.271
CHEMBL930	5358.2	nM	5.3582E-06	5.271
CHEMBL260563	5358.2	nM	5.3582E-06	5.271
CHEMBL452076	5358.2	nM	5.3582E-06	5.271
CHEMBL284116	5358.2	nM	5.3582E-06	5.271
CHEMBL1371156	5358.2	nM	5.3582E-06	5.271
CHEMBL83	5358.2	nM	5.3582E-06	5.271

Table B.1. Converted IC<sub>50</sub> and pIC<sub>50</sub> values of the known bioactive molecules of ARSA are listed in the decreasing order of their PIC<sub>50</sub> values (cont.).

CHEMBL ID	STD VALUE	STD UNITS	IC <sub>50</sub> (M)	pIC <sub>50</sub>
CHEMBL17127	5358.2	nM	5.3582E-06	5.271
CHEMBL1256351	5358.2	nM	5.3582E-06	5.271
CHEMBL280074	5358.2	nM	5.3582E-06	5.271
CHEMBL1412861	5358.2	nM	5.3582E-06	5.271
CHEMBL333826	5358.2	nM	5.3582E-06	5.271
CHEMBL471	5358.2	nM	5.3582E-06	5.271
CHEMBL1358565	5358.2	nM	5.3582E-06	5.271
CHEMBL1160723	6012	nM	6.012E-06	5.221
CHEMBL74926	6012	nM	6.012E-06	5.221
CHEMBL2110371	6012	nM	6.012E-06	5.221
CHEMBL51989	6012	nM	6.012E-06	5.221
CHEMBL38686	6012	nM	6.012E-06	5.221
CHEMBL1533156	6012	nM	6.012E-06	5.221
CHEMBL102740	6012	nM	6.012E-06	5.221
CHEMBL295212	6012	nM	6.012E-06	5.221
CHEMBL6437	6012	nM	6.012E-06	5.221
CHEMBL273094	6012	nM	6.012E-06	5.221
CHEMBL1162201	6012	nM	6.012E-06	5.221
CHEMBL1229908	6012	nM	6.012E-06	5.221
CHEMBL420746	6012	nM	6.012E-06	5.221
CHEMBL1591	6012	nM	6.012E-06	5.221
CHEMBL30	6309.6	nM	6.3096E-06	5.200
CHEMBL88553	6745.6	nM	6.7456E-06	5.171
CHEMBL578741	6745.6	nM	6.7456E-06	5.171
CHEMBL32503	6745.6	nM	6.7456E-06	5.171
CHEMBL58033	6745.6	nM	6.7456E-06	5.171
CHEMBL1187724	6745.6	nM	6.7456E-06	5.171
CHEMBL415324	6745.6	nM	6.7456E-06	5.171

Table B.1. Converted IC<sub>50</sub> and pIC<sub>50</sub> values of the known bioactive molecules of ARSA are listed in the decreasing order of their PIC<sub>50</sub> values (cont.).

CHEMBL ID	STD VALUE	STD UNITS	IC <sub>50</sub> (M)	pIC <sub>50</sub>
CHEMBL8145	6745.6	nM	6.7456E-06	5.171
CHEMBL589	6745.6	nM	6.7456E-06	5.171
CHEMBL367625	7079.5	nM	7.0795E-06	5.150
CHEMBL574181	7568.6	nM	7.5686E-06	5.121
CHEMBL2448494	7568.6	nM	7.5686E-06	5.121
CHEMBL46516	7568.6	nM	7.5686E-06	5.121
CHEMBL673	7568.6	nM	7.5686E-06	5.121
CHEMBL1452520	7568.6	nM	7.5686E-06	5.121
CHEMBL270299	7568.6	nM	7.5686E-06	5.121
CHEMBL596380	7568.6	nM	7.5686E-06	5.121
CHEMBL464859	7568.6	nM	7.5686E-06	5.121
CHEMBL288096	7568.6	nM	7.5686E-06	5.121
CHEMBL1200714	8492.1	nM	8.4921E-06	5.071
CHEMBL47940	8492.1	nM	8.4921E-06	5.071
CHEMBL282567	8492.1	nM	8.4921E-06	5.071
CHEMBL508329	8492.1	nM	8.4921E-06	5.071
CHEMBL283206	8492.1	nM	8.4921E-06	5.071
CHEMBL1423	8492.1	nM	8.4921E-06	5.071
CHEMBL1607809	8492.1	nM	8.4921E-06	5.071
CHEMBL543	8492.1	nM	8.4921E-06	5.071
CHEMBL23832	8492.1	nM	8.4921E-06	5.071
CHEMBL285913	8492.1	nM	8.4921E-06	5.071
CHEMBL22373	8492.1	nM	8.4921E-06	5.071
CHEMBL1416659	8492.1	nM	8.4921E-06	5.071
CHEMBL64545	8492.1	nM	8.4921E-06	5.071
CHEMBL274525	8492.1	nM	8.4921E-06	5.071
CHEMBL267936	8492.1	nM	8.4921E-06	5.071
CHEMBL419296	8492.1	nM	8.4921E-06	5.071

Table B.1. Converted IC<sub>50</sub> and pIC<sub>50</sub> values of the known bioactive molecules of ARSA are listed in the decreasing order of their PIC<sub>50</sub> values (cont.).

CHEMBL ID	STD VALUE	STD UNITS	IC <sub>50</sub> (M)	pIC <sub>50</sub>
CHEMBL22778	8492.1	nM	8.4921E-06	5.071
CHEMBL9666	8492.1	nM	8.4921E-06	5.071
CHEMBL13790	8492.1	nM	8.4921E-06	5.071
CHEMBL776	8492.1	nM	8.4921E-06	5.071
CHEMBL49	8492.1	nM	8.4921E-06	5.071
CHEMBL13045	8492.1	nM	8.4921E-06	5.071
CHEMBL101168	8492.1	nM	8.4921E-06	5.071
CHEMBL1556172	8492.1	nM	8.4921E-06	5.071
CHEMBL225284	8492.1	nM	8.4921E-06	5.071
CHEMBL47529	8492.1	nM	8.4921E-06	5.071
CHEMBL534	8492.1	nM	8.4921E-06	5.071
CHEMBL373937	8492.1	nM	8.4921E-06	5.071
CHEMBL488803	8492.1	nM	8.4921E-06	5.071
CHEMBL491960	8492.1	nM	8.4921E-06	5.071
CHEMBL1334874	8492.1	nM	8.4921E-06	5.071
CHEMBL63154	8492.1	nM	8.4921E-06	5.071
CHEMBL278332	8492.1	nM	8.4921E-06	5.071
CHEMBL52030	8492.1	nM	8.4921E-06	5.071
CHEMBL293749	8492.1	nM	8.4921E-06	5.071
CHEMBL1332270	8492.1	nM	8.4921E-06	5.071
CHEMBL245019	8492.1	nM	8.4921E-06	5.071
CHEMBL6567	8492.1	nM	8.4921E-06	5.071
CHEMBL54976	8492.1	nM	8.4921E-06	5.071
CHEMBL714	8492.1	nM	8.4921E-06	5.071
CHEMBL1006	8492.1	nM	8.4921E-06	5.071
CHEMBL19299	8492.1	nM	8.4921E-06	5.071
CHEMBL280563	8492.1	nM	8.4921E-06	5.071
CHEMBL1094	8492.1	nM	8.4921E-06	5.071

Table B.1. Converted IC<sub>50</sub> and pIC<sub>50</sub> values of the known bioactive molecules of ARSA are listed in the decreasing order of their PIC<sub>50</sub> values (cont.).

CHEMBL ID	STD VALUE	STD UNITS	IC <sub>50</sub> (M)	pIC <sub>50</sub>
CHEMBL36715	8492.1	nM	8.4921E-06	5.071
CHEMBL945	8492.1	nM	8.4921E-06	5.071
CHEMBL521	8492.1	nM	8.4921E-06	5.071
CHEMBL24938	8492.1	nM	8.4921E-06	5.071
CHEMBL239800	8492.1	nM	8.4921E-06	5.071
CHEMBL1623871	8492.1	nM	8.4921E-06	5.071
CHEMBL8706	8492.1	nM	8.4921E-06	5.071
CHEMBL1082832	8492.1	nM	8.4921E-06	5.071
CHEMBL145	8492.1	nM	8.4921E-06	5.071
CHEMBL253144	8492.1	nM	8.4921E-06	5.071
CHEMBL154	8492.1	nM	8.4921E-06	5.071
CHEMBL276727	8492.1	nM	8.4921E-06	5.071
CHEMBL628	8492.1	nM	8.4921E-06	5.071
CHEMBL93403	8492.1	nM	8.4921E-06	5.071
CHEMBL106265	8492.1	nM	8.4921E-06	5.071
CHEMBL1256686	8492.1	nM	8.4921E-06	5.071
CHEMBL17331	8492.1	nM	8.4921E-06	5.071
CHEMBL315795	8492.1	nM	8.4921E-06	5.071
CHEMBL64249	8492.1	nM	8.4921E-06	5.071
CHEMBL96	8492.1	nM	8.4921E-06	5.071
CHEMBL321691	8492.1	nM	8.4921E-06	5.071
CHEMBL155265	8492.1	nM	8.4921E-06	5.071
CHEMBL696	8912.5	nM	8.9125E-06	5.050
CHEMBL1096339	9528.3	nM	9.5283E-06	5.021
CHEMBL274323	9528.3	nM	9.5283E-06	5.021
CHEMBL425294	9528.3	nM	9.5283E-06	5.021
CHEMBL360052	9528.3	nM	9.5283E-06	5.021
CHEMBL277498	9528.3	nM	9.5283E-06	5.021

Table B.1. Converted IC<sub>50</sub> and pIC<sub>50</sub> values of the known bioactive molecules of ARSA are listed in the decreasing order of their PIC<sub>50</sub> values (cont.).

CHEMBL ID	STD VALUE	STD UNITS	IC <sub>50</sub> (M)	pIC <sub>50</sub>
CHEMBL275854	9528.3	nM	9.5283E-06	5.021
CHEMBL95758	9528.3	nM	9.5283E-06	5.021
CHEMBL433461	9528.3	nM	9.5283E-06	5.021
CHEMBL383144	9528.3	nM	9.5283E-06	5.021
CHEMBL534232	10691	nM	1.0691E-05	4.971
CHEMBL7568	10691	nM	1.0691E-05	4.971
CHEMBL1180496	10691	nM	1.0691E-05	4.971
CHEMBL333985	10691	nM	1.0691E-05	4.971
CHEMBL1622807	10691	nM	1.0691E-05	4.971
CHEMBL1402894	10691	nM	1.0691E-05	4.971
CHEMBL222519	10691	nM	1.0691E-05	4.971
CHEMBL597	10691	nM	1.0691E-05	4.971
CHEMBL30432	10691	nM	1.0691E-05	4.971
CHEMBL295467	10691	nM	1.0691E-05	4.971
CHEMBL489553	10691	nM	1.0691E-05	4.971
CHEMBL126077	10691	nM	1.0691E-05	4.971
CHEMBL1441661	10691	nM	1.0691E-05	4.971
CHEMBL84158	11995.5	nM	1.19955E-05	4.921
CHEMBL174271	11995.5	nM	1.19955E-05	4.921
CHEMBL88612	11995.5	nM	1.19955E-05	4.921
CHEMBL1117	11995.5	nM	1.19955E-05	4.921
CHEMBL774	11995.5	nM	1.19955E-05	4.921
CHEMBL1206	11995.5	nM	1.19955E-05	4.921
CHEMBL249089	11995.5	nM	1.19955E-05	4.921
CHEMBL379156	13459.1	nM	1.34591E-05	4.871
CHEMBL96926	13459.1	nM	1.34591E-05	4.871
CHEMBL277362	13459.1	nM	1.34591E-05	4.871
CHEMBL415879	13459.1	nM	1.34591E-05	4.871

Table B.1. Converted IC50 and pIC50 values of the known bioactive molecules of ARSA are listed in the decreasing order of their PIC50 values (cont.).

CHEMBL ID	STD VALUE	STD UNITS	IC50 (M)	pIC50
CHEMBL1568415	15101.4	nM	1.51014E-05	4.821
CHEMBL128365	15101.4	nM	1.51014E-05	4.821
CHEMBL304818	15101.4	nM	1.51014E-05	4.821
CHEMBL1417392	15101.4	nM	1.51014E-05	4.821
CHEMBL8867	15101.4	nM	1.51014E-05	4.821
CHEMBL1256514	16944.1	nM	1.69441E-05	4.771
CHEMBL1559663	16944.1	nM	1.69441E-05	4.771
CHEMBL1231700	16944.1	nM	1.69441E-05	4.771
CHEMBL26138	16944.1	nM	1.69441E-05	4.771
CHEMBL1330721	16944.1	nM	1.69441E-05	4.771
CHEMBL12587	16944.1	nM	1.69441E-05	4.771
CHEMBL3348825	16944.1	nM	1.69441E-05	4.771
CHEMBL1476294	16944.1	nM	1.69441E-05	4.771
CHEMBL280099	16944.1	nM	1.69441E-05	4.771
CHEMBL109037	16944.1	nM	1.69441E-05	4.771
CHEMBL517986	16944.1	nM	1.69441E-05	4.771
CHEMBL1643	16944.1	nM	1.69441E-05	4.771
CHEMBL402468	16944.1	nM	1.69441E-05	4.771
CHEMBL291143	19011.5	nM	1.90115E-05	4.721
CHEMBL168276	19011.5	nM	1.90115E-05	4.721
CHEMBL51483	19011.5	nM	1.90115E-05	4.721
CHEMBL1256291	19011.5	nM	1.90115E-05	4.721
CHEMBL18331	19011.5	nM	1.90115E-05	4.721
CHEMBL371811	19011.5	nM	1.90115E-05	4.721
CHEMBL976	19011.5	nM	1.90115E-05	4.721
CHEMBL260374	19011.5	nM	1.90115E-05	4.721
CHEMBL1435579	19011.5	nM	1.90115E-05	4.721
CHEMBL297784	19011.5	nM	1.90115E-05	4.721

Table B.1. Converted IC<sub>50</sub> and pIC<sub>50</sub> values of the known bioactive molecules of ARSA are listed in the decreasing order of their PIC<sub>50</sub> values (cont.).

CHEMBL ID	STD VALUE	STD UNITS	IC <sub>50</sub> (M)	pIC <sub>50</sub>
CHEMBL417799	21331.3	nM	2.13313E-05	4.671
CHEMBL459016	21331.3	nM	2.13313E-05	4.671
CHEMBL588	21331.3	nM	2.13313E-05	4.671
CHEMBL20	21331.3	nM	2.13313E-05	4.671
CHEMBL1224512	21331.3	nM	2.13313E-05	4.671
CHEMBL68500	21331.3	nM	2.13313E-05	4.671
CHEMBL375270	21331.3	nM	2.13313E-05	4.671
CHEMBL460574	21331.3	nM	2.13313E-05	4.671
CHEMBL1329455	21331.3	nM	2.13313E-05	4.671
CHEMBL8197	21331.3	nM	2.13313E-05	4.671
CHEMBL1257003	21331.3	nM	2.13313E-05	4.671
CHEMBL258764	21331.3	nM	2.13313E-05	4.671
CHEMBL52	21331.3	nM	2.13313E-05	4.671
CHEMBL355496	21331.3	nM	2.13313E-05	4.671
CHEMBL1193571	21331.3	nM	2.13313E-05	4.671
CHEMBL72	21331.3	nM	2.13313E-05	4.671
CHEMBL460	21331.3	nM	2.13313E-05	4.671
CHEMBL416755	21331.3	nM	2.13313E-05	4.671
CHEMBL76903	21331.3	nM	2.13313E-05	4.671
CHEMBL269733	21331.3	nM	2.13313E-05	4.671
CHEMBL106516	21331.3	nM	2.13313E-05	4.671
CHEMBL10247	21331.3	nM	2.13313E-05	4.671
CHEMBL123	22387.2	nM	2.23872E-05	4.650
CHEMBL429095	23934.1	nM	2.39341E-05	4.621
CHEMBL69863	23934.1	nM	2.39341E-05	4.621
CHEMBL1622540	23934.1	nM	2.39341E-05	4.621
CHEMBL164	23934.1	nM	2.39341E-05	4.621
CHEMBL27673	23934.1	nM	2.39341E-05	4.621

Table B.1. Converted IC<sub>50</sub> and pIC<sub>50</sub> values of the known bioactive molecules of ARSA are listed in the decreasing order of their PIC<sub>50</sub> values (cont.).

CHEMBL ID	STD VALUE	STD UNITS	IC <sub>50</sub> (M)	pIC <sub>50</sub>
CHEMBL22207	23934.1	nM	2.39341E-05	4.621
CHEMBL488025	23934.1	nM	2.39341E-05	4.621
CHEMBL58	23934.1	nM	2.39341E-05	4.621
CHEMBL311498	23934.1	nM	2.39341E-05	4.621
CHEMBL9514	23934.1	nM	2.39341E-05	4.621
CHEMBL184	23934.1	nM	2.39341E-05	4.621
CHEMBL224864	23934.1	nM	2.39341E-05	4.621
CHEMBL279956	23934.1	nM	2.39341E-05	4.621
CHEMBL281594	23934.1	nM	2.39341E-05	4.621
CHEMBL1201356	23934.1	nM	2.39341E-05	4.621
CHEMBL1330831	23934.1	nM	2.39341E-05	4.621
CHEMBL266084	23934.1	nM	2.39341E-05	4.621
CHEMBL303958	23934.1	nM	2.39341E-05	4.621
CHEMBL276832	26854.5	nM	2.68545E-05	4.571
CHEMBL341898	26854.5	nM	2.68545E-05	4.571
CHEMBL86931	26854.5	nM	2.68545E-05	4.571
CHEMBL1411717	26854.5	nM	2.68545E-05	4.571
CHEMBL1160025	26854.5	nM	2.68545E-05	4.571
CHEMBL605003	26854.5	nM	2.68545E-05	4.571
CHEMBL498	26854.5	nM	2.68545E-05	4.571
CHEMBL255044	26854.5	nM	2.68545E-05	4.571
CHEMBL326958	26854.5	nM	2.68545E-05	4.571
CHEMBL744	26854.5	nM	2.68545E-05	4.571
CHEMBL185	26854.5	nM	2.68545E-05	4.571
CHEMBL15770	26854.5	nM	2.68545E-05	4.571
CHEMBL279229	26854.5	nM	2.68545E-05	4.571
CHEMBL59	26854.5	nM	2.68545E-05	4.571
CHEMBL1628215	30131.3	nM	3.01313E-05	4.521

Table B.1. Converted IC<sub>50</sub> and pIC<sub>50</sub> values of the known bioactive molecules of ARSA are listed in the decreasing order of their PIC<sub>50</sub> values (cont.).

CHEMBL ID	STD VALUE	STD UNITS	IC <sub>50</sub> (M)	pIC <sub>50</sub>
CHEMBL537	30131.3	nM	3.01313E-05	4.521
CHEMBL1065	30131.3	nM	3.01313E-05	4.521
CHEMBL1437	30131.3	nM	3.01313E-05	4.521
CHEMBL1740	30131.3	nM	3.01313E-05	4.521
CHEMBL222280	30131.3	nM	3.01313E-05	4.521
CHEMBL1004	30131.3	nM	3.01313E-05	4.521
CHEMBL1374843	30131.3	nM	3.01313E-05	4.521
CHEMBL3351078	30131.3	nM	3.01313E-05	4.521
CHEMBL1551767	33807.8	nM	3.38078E-05	4.471
CHEMBL445	33807.8	nM	3.38078E-05	4.471
CHEMBL70565	33807.8	nM	3.38078E-05	4.471
CHEMBL1255835	33807.8	nM	3.38078E-05	4.471
CHEMBL76589	33807.8	nM	3.38078E-05	4.471
CHEMBL1255834	33807.8	nM	3.38078E-05	4.471
CHEMBL1201288	33807.8	nM	3.38078E-05	4.471
CHEMBL1455	33807.8	nM	3.38078E-05	4.471
CHEMBL56393	33807.8	nM	3.38078E-05	4.471
CHEMBL161343	33807.8	nM	3.38078E-05	4.471
CHEMBL1328898	33807.8	nM	3.38078E-05	4.471
CHEMBL102714	33807.8	nM	3.38078E-05	4.471
CHEMBL268229	33807.8	nM	3.38078E-05	4.471
CHEMBL50	33807.8	nM	3.38078E-05	4.471
CHEMBL319244	33807.8	nM	3.38078E-05	4.471
CHEMBL40998	37933	nM	3.7933E-05	4.421
CHEMBL1213033	37933	nM	3.7933E-05	4.421
CHEMBL240773	37933	nM	3.7933E-05	4.421
CHEMBL459	37933	nM	3.7933E-05	4.421
CHEMBL1256028	37933	nM	3.7933E-05	4.421

Table B.1. Converted IC<sub>50</sub> and pIC<sub>50</sub> values of the known bioactive molecules of ARSA are listed in the decreasing order of their PIC<sub>50</sub> values (cont.).

CHEMBL ID	STD VALUE	STD UNITS	IC <sub>50</sub> (M)	pIC <sub>50</sub>
CHEMBL401570	37933	nM	3.7933E-05	4.421
CHEMBL452861	37933	nM	3.7933E-05	4.421
CHEMBL109684	37933	nM	3.7933E-05	4.421
CHEMBL94	37933	nM	3.7933E-05	4.421
CHEMBL258767	37933	nM	3.7933E-05	4.421
CHEMBL1590464	37933	nM	3.7933E-05	4.421
CHEMBL1008	37933	nM	3.7933E-05	4.421
CHEMBL1373981	37933	nM	3.7933E-05	4.421
CHEMBL1079	37933	nM	3.7933E-05	4.421
CHEMBL73711	37933	nM	3.7933E-05	4.421
CHEMBL1332680	37933	nM	3.7933E-05	4.421
CHEMBL71752	37933	nM	3.7933E-05	4.421
CHEMBL539947	37933	nM	3.7933E-05	4.421
CHEMBL66089	37933	nM	3.7933E-05	4.421
CHEMBL299683	37933	nM	3.7933E-05	4.421
CHEMBL1241268	37933	nM	3.7933E-05	4.421
CHEMBL600	37933	nM	3.7933E-05	4.421
CHEMBL1191361	37933	nM	3.7933E-05	4.421
CHEMBL320820	37933	nM	3.7933E-05	4.421
CHEMBL617	37933	nM	3.7933E-05	4.421
CHEMBL1448	37933	nM	3.7933E-05	4.421
CHEMBL72365	37933	nM	3.7933E-05	4.421
CHEMBL1238	37933	nM	3.7933E-05	4.421
CHEMBL811	37933	nM	3.7933E-05	4.421
CHEMBL17157	37933	nM	3.7933E-05	4.421
CHEMBL598952	37933	nM	3.7933E-05	4.421
CHEMBL900	37933	nM	3.7933E-05	4.421
CHEMBL1179189	37933	nM	3.7933E-05	4.421

Table B.1. Converted IC<sub>50</sub> and pIC<sub>50</sub> values of the known bioactive molecules of ARSA are listed in the decreasing order of their PIC<sub>50</sub> values (cont.).

CHEMBL ID	STD VALUE	STD UNITS	IC <sub>50</sub> (M)	pIC <sub>50</sub>
CHEMBL1398096	37933	nM	3.7933E-05	4.421
CHEMBL261557	37933	nM	3.7933E-05	4.421
CHEMBL492008	37933	nM	3.7933E-05	4.421
CHEMBL22870	37933	nM	3.7933E-05	4.421
CHEMBL1316321	37933	nM	3.7933E-05	4.421
CHEMBL35354	37933	nM	3.7933E-05	4.421
CHEMBL1359148	37933	nM	3.7933E-05	4.421
CHEMBL479	37933	nM	3.7933E-05	4.421
CHEMBL226115	37933	nM	3.7933E-05	4.421
CHEMBL1200604	37933	nM	3.7933E-05	4.421
CHEMBL784	37933	nM	3.7933E-05	4.421
CHEMBL1760	37933	nM	3.7933E-05	4.421
CHEMBL1370772	37933	nM	3.7933E-05	4.421
CHEMBL1257015	37933	nM	3.7933E-05	4.421
CHEMBL1256364	37933	nM	3.7933E-05	4.421
CHEMBL187734	37933	nM	3.7933E-05	4.421
CHEMBL1621	37933	nM	3.7933E-05	4.421
CHEMBL1009	37933	nM	3.7933E-05	4.421
CHEMBL429298	37933	nM	3.7933E-05	4.421
CHEMBL54661	37933	nM	3.7933E-05	4.421
CHEMBL1404	37933	nM	3.7933E-05	4.421
CHEMBL1201353	37933	nM	3.7933E-05	4.421
CHEMBL318111	37933	nM	3.7933E-05	4.421
CHEMBL1358525	37933	nM	3.7933E-05	4.421
CHEMBL423	37933	nM	3.7933E-05	4.421
CHEMBL708	37933	nM	3.7933E-05	4.421
CHEMBL14563	37933	nM	3.7933E-05	4.421
CHEMBL19130	37933	nM	3.7933E-05	4.421

Table B.1. Converted IC<sub>50</sub> and pIC<sub>50</sub> values of the known bioactive molecules of ARSA are listed in the decreasing order of their PIC<sub>50</sub> values (cont.).

CHEMBL ID	STD VALUE	STD UNITS	IC <sub>50</sub> (M)	pIC <sub>50</sub>
CHEMBL88621	37933	nM	3.7933E-05	4.421
CHEMBL127421	37933	nM	3.7933E-05	4.421
CHEMBL1256871	37933	nM	3.7933E-05	4.421
CHEMBL960	37933	nM	3.7933E-05	4.421
CHEMBL190	37933	nM	3.7933E-05	4.421
CHEMBL23472	37933	nM	3.7933E-05	4.421
CHEMBL267777	37933	nM	3.7933E-05	4.421
CHEMBL259388	37933	nM	3.7933E-05	4.421
CHEMBL281232	37933	nM	3.7933E-05	4.421
CHEMBL129801	37933	nM	3.7933E-05	4.421
CHEMBL56731	37933	nM	3.7933E-05	4.421
CHEMBL441618	42561.5	nM	4.25615E-05	4.371

## APPENDIX C: R CODES FOR CROSS CORRELATION CALCULATIONS

```
setwd("/home/path/to/ARSA_cross_corR_files/")
trj<-readMat('trj_matrix_3d.mat')
trj<-as.numeric(unlist(trj), ncol = 3, byrow = TRUE)
dim(trj)<-c(2084,3,485)
ort<-readMat('mean.mat')
ort<-as.numeric(unlist(ort), ncol = 3, byrow = TRUE)
dim(ort)<-c(1,3,485)
delta<-array(0, c(2084,3,485))
for (i in 1:2084) {
  delta[i,]<-trj[i,]-ort[1,]
}
matrix<-array(0, c(2084,485,485))
for (i in 1:2084) {
  for (j in 1:485) {
    for (k in 1:485) {
      matrix[i,j,k]<-sum(delta[i,j]*delta[i,k])
    }
  }
  print(i/2084*100)
}
num = array(0,c(485,485))
for (j in 1:485){
  num[,j] = apply(matrix[,j],2,mean)
  print(j/485*100)
```

```
}  
  
#####  
  
sqr<-delta^2  
rdeno=array(0,c(2084,485))  
for (i in 1:485) {  
  rdeno[,i]<-apply(sqr[,i], 1, sum)  
}  
orti<-colMeans(rdeno)  
befsq<-array(0, c(485,485))  
for (i in 1:485){  
  for (j in 1:485) {  
    befsqr[i,j]<-orti[i]*orti[j]  
  }  
}  
denominator<-sqrt(befsq)  
C<-num/denominator  
write.csv(C, file = "cross_corr.csv")
```

## APPENDIX D: CHEMICAL STRUCTURES OF DE NOVO SYNTHESIZED MOLECULES

Table D.1. The SMILES codes of the de novo synthesized molecules that were used in docking calculations

SMILES codes	Molecule name
<chem>C1C[NH+](C)CCN1CN2C(=O)C(\c(c23)cc([N+])([O-])=O)cc3)=N/NC(=S)Nc4ccc(cc4)OC</chem>	G7
<chem>FC(F)(F)Oc(cc1)cc(c12)/C(C(=O)N2)=N/NC(=S)NCC=C</chem>	D2
<chem>C1CSCCN1CN2C(=O)C(\c(c23)cc(cc3)OC(F)(F)F)=N/NC(=S)Nc4cc(Cl)ccc4</chem>	D11-tm
<chem>C=CCNC(=S)N\N=C(\c(c12)cc(cc1)OC(F)(F)F)C(=O)N2CN3CCOCC3</chem>	D28
<chem>FC(F)(F)Oc(cc1)cc(c12)/C(C(=O)N2C)=N/NC(=S)NC=C</chem>	D16
<chem>c1cccc(c12)/C(C(=O)N2)=N/N=c3/sc(C(=O)[O-])c(C)n3CC=C</chem>	I7
<chem>c1cccc(c12)/C(C(=O)N2)=N/N=c3/sc(C(=O)[O-])c(C)n3-c4ccc(Cl)cc4</chem>	I10
<chem>C1CCCN1CN2C(=O)C(\c(c23)cc([N+])([O-])=O)cc3)=N/NC(=S)Nc4ccc(cc4)OC</chem>	G2

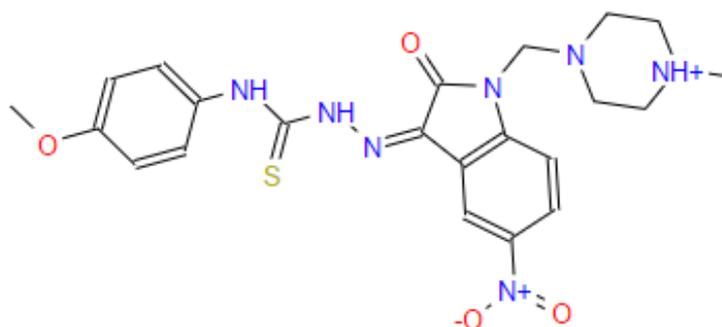


Figure D.1. The chemical structure of G7

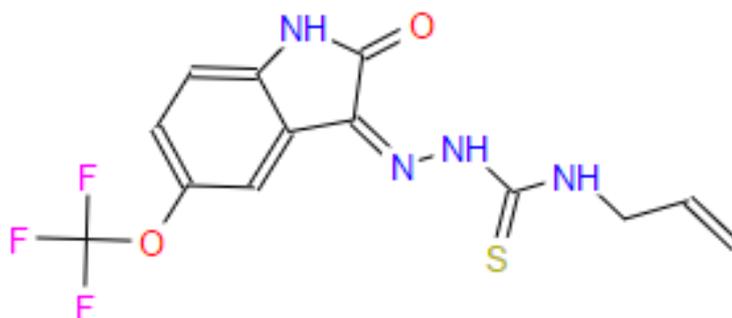


Figure D.2. The chemical structure of D2

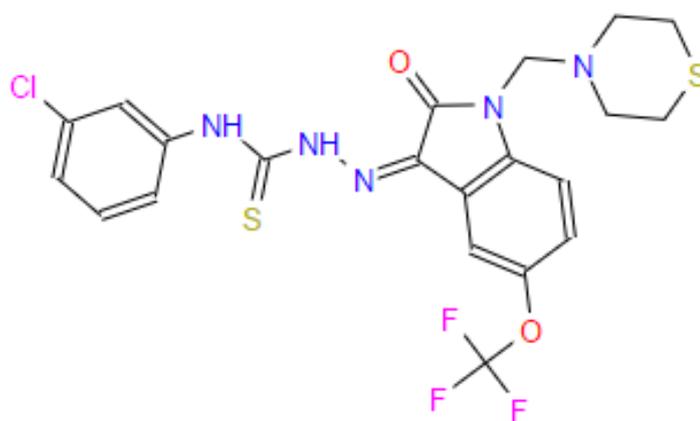


Figure D.3. The chemical structure of D11-tm

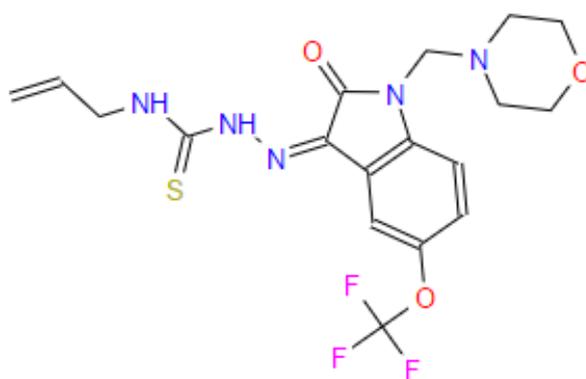


Figure D.4. The chemical structure of D28

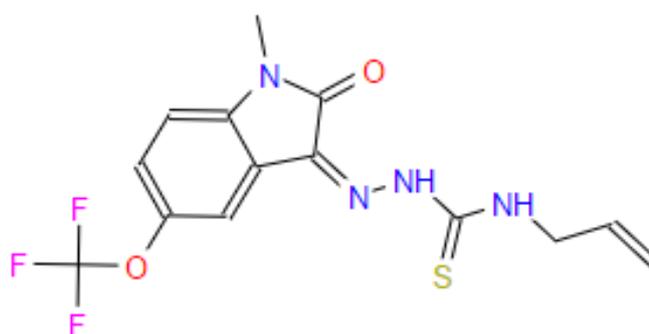


Figure D.5. The chemical structure of D16

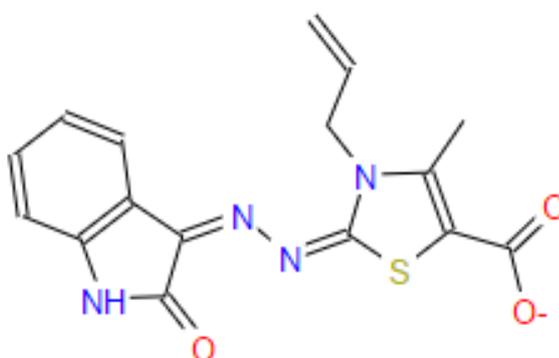


Figure D.6. The chemical structure of I7

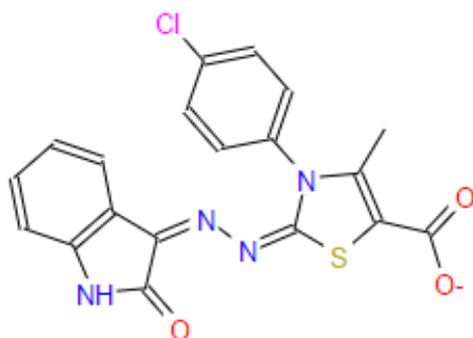


Figure D.7. The chemical structure of I10

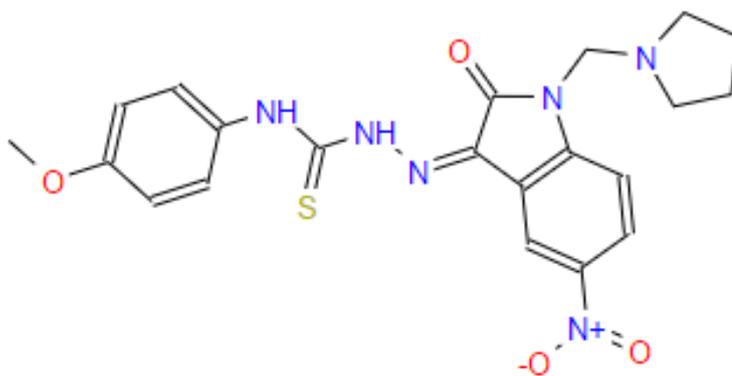


Figure D.8. The chemical structure of G2