



ACIBADEM MEHMET ALI AYDINLAR UNIVERSITY
INSTITUTE OF HEALTH SCIENCES

**COMPARATIVE MOLECULAR DYNAMICS ANALYSIS OF A
FUNGAL HYDROPHOBIN RODA BEHAVIOR AT INTERFACIAL
ENVIRONMENT**

İREM KARA
M.Sc. THESIS

DEPARTMENT OF BIOSTATISTICS AND BIOINFORMATICS

SUPERVISOR
Assoc. Prof. Günseli Bayram Akçapınar

İSTANBUL-2023



ACIBADEM MEHMET ALI AYDINLAR UNIVERSITY
INSTITUTE OF HEALTH SCIENCES

**COMPARATIVE MOLECULAR DYNAMICS ANALYSIS OF A
FUNGAL HYDROPHOBIN RODA BEHAVIOR AT
INTERFACIAL ENVIRONMENT**

İREM KARA
M.Sc. THESIS

DEPARTMENT OF BIOSTATISTICS AND BIOINFORMATICS

SUPERVISOR
Assoc. Prof. Günseli Bayram Akçapınar

İSTANBUL-2023

DECLARATION

I declare that this thesis work is my own work, I had no unethical behavior at any stages from the planning to the writing of the thesis, I obtained all the information in this thesis in accordance with academic and ethical rules, I cited all the information and comments that were not obtained with this thesis work, and I provided resources in the list of references. I also declare that there was no violation of any patents and copyrights during the study and writing of this thesis.

02/02/2023

İrem Kara

PREFACE AND ACKNOWLEDGEMENT

The most critical moment in the academic life of a graduate student is probably the assignment of a supervisor. It is of great importance to find a common ground together and to communicate effectively to carry out the studies. I believe that I am among those who are lucky in this matter. It is without doubt that my supervisor, Assoc. Prof. Günseli Bayram Akçapınar, is one the supervisors who can manage to reach out to her students. During my thesis studies, she has helped me look at things from different perspectives. It was always a great pleasure to discuss new ideas or methods together for our research. I cannot thank her enough for always being a supportive and inspiring mentor, and a friendly academic in this stressful time of my life.

I would also like to express my deepest gratitude to Assoc. Prof. Emel Timuçin and Asst. Prof. Sefer Baday for their time, comments, and suggestions before and during my thesis defense.

I must say that every member of our research group, PEBL, has helped me, more or less, both emotionally and academically, overcome the challenges that I encountered during my studies. Especially, Yunus Emre Dilek and Hüseyin Okan Soykam, with whom we study in similar research areas, have given me valuable advice on key points. I am happy that we had the chance to study together with my dear friend, Sena Kıvrak, in graduate school, just as we did in undergraduate. Having her by my side was a relief against many obstacles that we have faced during these years.

Last but not least, I am grateful for the endless support of my family. They were always there for me in any circumstance. There were times that I felt hopeless and desperate, but they encouraged and motivated me with their faith and trust in me.

Computing sources were provided by Turkish National e-Science e-Infrastructure (TRUBA) to perform Molecular Dynamics simulations and analyze trajectory files.

TABLE OF CONTENTS

DECLARATION.....	iv
PREFACE AND ACKNOWLEDGEMENT	v
TABLE OF CONTENTS.....	vi
LIST OF FIGURES	ix
ÖZET.....	1
ABSTRACT	2
1 INTRODUCTION AND AIM	3
2 BACKGROUND.....	6
2.1 Functions of Hydrophobins	6
2.2 The Role of Hydrophobins in Fungal Pathogenesis	8
2.3 The Role of RodA in Fungal Pathogenesis	8
2.4 Properties of Hydrophobin Structures	9
2.5 Self-Assembly and Layer Formation	13
2.6 Molecular Dynamics Simulations.....	20
2.6.1 Trajectory analyses	23
3 MATERIALS AND METHODS.....	25
3.1 Prediction of the Key Structural Regions in Self-Assembly	25
3.2 Molecular Dynamics Simulations.....	26
3.2.1 Trajectory analyses	30
3.2.2 Determination of clustering states and identification of intermolecular interaction regions.....	30
3.2.3 Measurements of intermolecular interactions	31
3.2.4 Monomer stability	32
4 RESULTS.....	33
4.1 General Analyses	33
4.2 Single Monomer and Two-Monomer MD Simulations.....	35
4.3 Four-Monomer MD Simulations.....	53
4.3.1 Simulation systems with positionally unrestrained water molecules	53
4.3.1.1 Orientation 1	53
4.3.1.2 Orientation 2	62
4.3.1.3 Orientation 3	71
4.3.1.4 Orientation 4	81
4.3.2 Simulation systems with positionally RWM	90
4.3.2.1 Orientation 1	90

4.3.2.2 Orientation 2	99
4.3.2.3 Orientation 3	109
4.3.2.4 Orientation 4	119
5 DISCUSSION.....	130
5.1 RodA Monomers were More Mobile in Bulk Solution Depending on the Initial Orientation	130
5.2 Compact Small Clusters were Formed Regardless of Initial Orientation	131
5.3 Initial Orientation and RWM Affect Cluster Size.....	132
5.4 N-termini and L2 Loops were Highly Involved in Intermolecular Interactions	133
5.5 Interactions Contributing to the Cluster Stability Differ Between the Two Simulation Environments.....	134
5.6 RodA Monomers Adsorbed to the Interface Mostly Through L2 Loops.	135
5.7 There is No Common Structural Region Involved in Class I Hydrophobin Assembly	136
5.8 Limitations	137
6 CONCLUSION	139
7 REFERENCES	140
8 APPENDIX	147
Appendix 1 Intermolecular distance measurements by “dist_bw_chains.tcl”.	147
Appendix 2 Atomic contact measurements by “contacts_for_chains.tcl”.	148
Appendix 3 Hydrophobic contact measurements by “hydrophobic_contacts.tcl”. 149	
Appendix 4. Hydrogen bond measurements by “hbonds.tcl”.....	150
Appendix 5 RMSD measurements by “RMSD_protein.tcl”.....	151
Appendix 6 RMSF measurements by “RMSF_protein.tcl”.....	152
Appendix 8 Contact maps of four monomers in Orientation 2 simulations....	154
Appendix 9 Contact maps of four monomers in Orientation 3 simulations....	155
Appendix 10 Contact maps of four monomers in Orientation 4 simulations...	156
Appendix 11 Contact maps of four monomers in Orientation 1 simulations with positionally RWM.	157
Appendix 12 Contact maps of four monomers in Orientation 2 simulations with positionally RWM.	158
Appendix 13 Contact maps of four monomers in Orientation 3 simulations with positionally RWM.	159
Appendix 14 Contact maps of four monomers in Orientation 4 simulations with positionally RWM.	160
9 CURRICULUM VITAE	161

LIST OF ABBREVIATIONS AND SYMBOLS

α-helix	Alpha helix
Å	Angstrom
AFM	Atomic force microscopy
β-sheet	Beta sheet
CA	Carbon alpha
COM	Center of mass
EM	Electron microscopy
fs	Femtosecond
HOPG	Highly ordered pyrolytic graphite
μs	Microsecond
MD	Molecular dynamics
NAMD	Nanoscale Molecular Dynamics
ns	Nanosecond
NMR	Nuclear magnetic resonance
PME	Particle Mesh Ewald
PBC	Periodic boundary conditions
ps	Picosecond
PDB	Protein Data Bank
REMD	Replica exchange molecular dynamics
RWM	Restrained water molecules
RMSD	Root mean square deviation
RMSF	Root mean square fluctuation
SASA	Solvent accessible surface area
3D	Three dimensional
TEM	Transmission electron microscopy
2D	Two dimensional
UWM	Unrestrained water molecules
VMD	Visual Molecular Dynamics

LIST OF FIGURES

Figure 1. Schematic representation of RodA self-assembly on the surface of <i>A. fumigatus</i> conidia and at air/water interface.	4
Figure 2. Scaled hydropathy plots of representative hydrophobins.	10
Figure 3. Multiple sequence alignment of Class I and Class II hydrophobins.	11
Figure 4. 3D structures of representative Class I and Class II hydrophobins.	12
Figure 5. Four different initial orientations of two and four RodA monomers simulation systems.	27
Figure 6. Representative solution (left) and interface (right) simulation systems with 2 Å thick water walls.	30
Figure 7. Snapshots of conformations at three different time points in monomer solution and interface simulation systems.	35
Figure 8. Snapshots of water boxes at three different time points in monomer interface simulation systems.	36
Figure 9. RMSD and RMSF measurements of single chains in monomer simulation systems in both environments.	37
Figure 10. Salt bridges observed in the monomer simulations.	38
Figure 11. Cluster analysis of solution simulations performed with two monomers.	40
Figure 12. Snapshots of conformations at three different time points in solution simulations performed with two monomers.	41
Figure 13. The COM-COM distances between chains in solution simulations performed with two monomers.	42
Figure 14. Contact maps of two-monomer solution and interface simulations.	43
Figure 15. Measurements of the number of hydrogen bonds between chain pairs in solution simulations performed with two monomers.	44
Figure 16. Hydrophobic contact maps of two-monomer solution simulations.	45
Figure 17. Cluster analysis of two-monomer interface simulations.	46
Figure 18. Snapshots of conformations at three different time points in interface simulations performed with two monomers.	47
Figure 19. The COM-COM distances between chains in interface simulations performed with two monomers.	48

Figure 20. Contact maps of two-monomer solution and interface simulations.	49
Figure 21. Measurements of the number of hydrogen bonds between chain pairs in solution simulations performed with two monomers.	50
Figure 22. Hydrophobic contact maps of two-monomer interface simulations.	51
Figure 23. RMSD (left) and RMSF (right) measurements of individual chains in simulations performed with two monomers in both environments.	52
Figure 24. Cluster analysis of Orientation 1 simulations.	54
Figure 25. Snapshots of conformations at three different time points and COM-COM distances between chains of Orientation 1 solution simulations.	55
Figure 26. Snapshots of conformations at three different time points and COM-COM distances between chains of Orientation 1 interface simulations.	56
Figure 27. Contact maps of Orientation 1 solution simulations.	57
Figure 28. Contact maps of Orientation 1 interface simulations.	58
Figure 29. Hydrophobic contact maps of both Orientation 1 solution and interface simulations.	59
Figure 30. Measurements of the number of hydrogen bonds between chain pairs in Orientation 1 simulation systems in both environments.	60
Figure 31. Measurements of conserved interchain salt bridges throughout the trajectories in Orientation 1 simulation systems for both environments.	61
Figure 32. RMSD and RMSF measurements of individual chains in Orientation 1 simulation systems in both environments.	62
Figure 33. Cluster analysis of Orientation 2 simulations.	63
Figure 34. Snapshots of conformations at three different time points and COM-COM distances between chains of Orientation 2 solution simulations.	64
Figure 35. Snapshots of conformations at three different time points and COM-COM distances between chains of Orientation 2 interface simulations.	65
Figure 36. Contact maps of Orientation 2 solution simulations.	66
Figure 37. Contact maps of Orientation 2 interface simulations.	67
Figure 38. Hydrophobic contact maps of both Orientation 2 solution and interface simulations.	68
Figure 39. Measurements of the number of hydrogen bonds between chain pairs in Orientation 2 simulation systems in both environments.	69

Figure 40. Measurements of conserved interchain salt bridges throughout the trajectories in Orientation 2 simulation systems for both environments.....	70
Figure 41. RMSD and RMSF measurements of individual chains in Orientation 2 simulation systems in both environments.	71
Figure 42. Cluster analysis of Orientation 3 simulations.....	73
Figure 43. Snapshots of conformations at three different time points and COM–COM distances between chains of Orientation 3 solution simulations.....	74
Figure 44. Snapshots of conformations at three different time points and COM–COM distances between chains of Orientation 3 interface simulations.....	75
Figure 45. Contact maps of Orientation 3 solution simulations.....	76
Figure 46. Contact maps of Orientation 3 interface simulations.	77
Figure 47. Hydrophobic contact maps of both Orientation 3 solution and interface simulations.	78
Figure 48. Measurements of the number of hydrogen bonds between chain pairs in Orientation 3 simulation systems in both environments.....	79
Figure 49. Measurements of conserved interchain salt bridges throughout the trajectories in Orientation 3 simulation systems for both environments.....	80
Figure 50. RMSD and RMSF measurements of individual chains in Orientation 3 simulation systems in both environments.	81
Figure 51. Cluster analysis of Orientation 4 simulations.....	82
Figure 52. Snapshots of conformations at three different time points and COM–COM distances between chains of Orientation 4 solution simulations.....	83
Figure 53. Snapshots of conformations at three different time points and COM–COM distances between chains of Orientation 4 interface simulations.....	84
Figure 54. Contact maps of Orientation 4 solution simulations.....	85
Figure 55. Contact maps of Orientation 4 interface simulations.	86
Figure 56. Hydrophobic contact maps of both Orientation 4 solution and interface simulations.	87
Figure 57. Measurements of the number of hydrogen bonds between chain pairs in Orientation 4 simulation systems in both environments.....	88
Figure 58. RMSD and RMSF measurements of individual chains in Orientation 4 simulation systems in both environments.	89

Figure 59. Cluster analysis of Orientation 1 simulation systems with positionally RWM.....	91
Figure 60. Snapshots of conformations at three different time points and COM–COM distances between chains of Orientation 1 solution simulation systems with positionally RWM.....	92
Figure 61. Snapshots of conformations at three different time points and COM–COM distances between chains of Orientation 1 interface simulation systems with positionally RWM.....	93
Figure 62. Contact maps of Orientation 1 solution simulation systems with positionally RWM.....	94
Figure 63. Contact maps of Orientation 1 interface simulation systems with positionally RWM.....	95
Figure 64. Hydrophobic contact maps of both Orientation 1 solution and interface simulation systems with positionally RWM.	96
Figure 65. Measurements of conserved interchain salt bridges throughout the trajectories in Orientation 1 simulation systems with positionally RWM for both environments.....	97
Figure 66. Measurements of the number of hydrogen bonds between chain pairs in Orientation 1 simulation systems with positionally RWM in both environments.....	98
Figure 67. RMSD and RMSF measurements of chains in Orientation 1 simulations with positionally RWM in both environments.....	99
Figure 68. Cluster analysis of Orientation 2 simulation systems with positionally RWM.....	101
Figure 69. Snapshots of conformations at three different time points and COM–COM distances between chains of Orientation 2 solution simulation systems with positionally RWM.....	102
Figure 70. Snapshots of conformations at three different time points and COM–COM distances between chains of Orientation 2 interface simulation systems with positionally RWM.....	103
Figure 71. Contact maps of Orientation 2 solution simulation systems with positionally RWM.....	104

Figure 72. Contact maps of Orientation 2 interface simulation systems with positionally RWM.....	105
Figure 73. Hydrophobic contact maps of both Orientation 2 solution and interface simulation systems with positionally RWM.	106
Figure 74. Measurements of conserved interchain salt bridges throughout the trajectories in Orientation 2 simulation systems with positionally RWM for both environments	107
Figure 75. Measurements of the number of hydrogen bonds between chain pairs in Orientation 2 simulation systems with positionally RWM in both environments...	108
Figure 76. RMSD and RMSF measurements of chains in Orientation 2 simulations with positionally RWM in both environments.....	109
Figure 77. Cluster analysis of Orientation 3 simulation systems with positionally RWM.....	111
Figure 78. Snapshots of conformations at three different time points and COM–COM distances between chains of Orientation 2 solution simulation systems with positionally RWM.....	112
Figure 79. Snapshots of conformations at three different time points and COM–COM distances between chains of Orientation 3 interface simulation systems with positionally RWM.....	113
Figure 80. Contact maps of Orientation 3 solution simulation systems with positionally RWM.....	114
Figure 81. Contact maps of Orientation 3 interface simulation systems with positionally RWM.....	115
Figure 82. Hydrophobic contact maps of both Orientation 3 solution and interface simulation systems with positionally RWM.	116
Figure 83. Measurements of conserved interchain salt bridges throughout the trajectories in Orientation 3 simulation systems with positionally RWM for both environments	117
Figure 84. Measurements of the number of hydrogen bonds between chain pairs in Orientation 3 simulation systems with positionally RWM in both environments...	118
Figure 85. RMSD and RMSF measurements of chains in Orientation 3 simulations with positionally RWM in both environments.....	119

Figure 86. Cluster analysis of Orientation 4 simulation systems with positionally RWM.....	121
Figure 87. Snapshots of conformations at three different time points and COM–COM distances between chains of Orientation 4 solution simulation systems with positionally RWM.....	122
Figure 88. Snapshots of conformations at three different time points and COM–COM distances between chains of Orientation 4 interface simulation systems with positionally RWM.....	123
Figure 89. Contact maps of Orientation 4 solution simulation systems with positionally RWM.....	124
Figure 90. Contact maps of Orientation 4 interface simulation systems with positionally RWM.....	125
Figure 91. Hydrophobic contact maps of both Orientation 4 solution and interface simulation systems with positionally RWM.	126
Figure 92. Measurements of conserved interchain salt bridges throughout the trajectories in Orientation 4 simulation systems with positionally RWM for both environments.....	127
Figure 93. Measurements of the number of hydrogen bonds between chain pairs in Orientation 4 simulation systems with positionally RWM in both environments...	128
Figure 94. RMSD and RMSF measurements of chains in Orientation 4 simulations with positionally RWM in both environments.....	129

LIST OF TABLES

Table 1. Number of atoms and box dimensions of each simulation system.	27
Table 2. Duration of each run performed in this thesis.	29
Table 3. Regions predicted as amyloidogenic and residues predicted as to be located at protein-protein interfaces.	34



ÖZET

Arayüz Ortamında Mantar Hidrofobin RodA Davranışının Karşılaştırmalı Moleküler Dinamik Analizi

Aspergillus fumigatus fırsatçı bir insan patojenidir ve bağışıklık yetersizliği olan konakçılarda ciddi solunum yolu hastalıklarına neden olur. Enfeksiyona eşeysiz sporlar (konidia) aracılık eder ve konidia yüzeyi hidrofobin adı verilen özel amfipatik proteinlerle kaplıdır. Bu proteinler, hücre duvarı bileşenlerini, özellikle de patojene karşı bağışıklık tepkisini tetikleyen polisakkaritleri maskeler. *A. fumigatus* Sınıf I RodA hidrofobini, çubuksu yapılarının paralel düzenlenmesinden kaynaklanan karakteristik bir morfolojiye sahip, oldukça düzenli amfipatik katmanlar halinde birleşir. Çubuklar, nörodejeneratif ve metabolik hastalıklara neden olan amiloid proteinlere benzer şekilde β -yaprakları açısından zengin amiloid benzeri yapılardan oluşur. Ayrıca, hidrofobinler hidrofilik/hidrofobik arayüzlerde katmanlar oluşturabilir. Aynı sınıftaki hidrofobinler benzer morfolojik karaktere sahip katmanlar halinde düzenlense de kendiliğinden düzenlenmede görev alan yapısal bölgeler değişiklik göstermektedir. RodA hidrofobininin kendiliğinden düzenlenmesindeki kilit bölgeler ve katmanların olası yapı taşları daha önceki çalışmalarda önerilmiştir. Bununla birlikte, kendiliğinden düzenlenme mekanizmalarının atomik detayları henüz bilinmemektedir. Bu tezde, kendiliğinden düzenlenmenin ilk adımları hakkında fikir edinmek için hava/su arayüzlerinde tüm atom MD simülasyonları gerçekleştirilmiştir. Düzenlenmeyi hızlandırmak ve periyodik resimlerle etkileşimi engellemek için su kutularının kenarlarına yakın su moleküllerine kısmi konum kısıtlamaları uygulanmıştır. Kümelenme davranışının büyük ölçüde ilk yönelimlere bağlı olduğu gözlemlenmiştir ve hem L2 hem de L3 döngülerinin arayüzde havaya doğru dönerek üst üste binmekten ziyade yan yana hizalandığı bir kendiliğinden düzenlenme modeli önerilmiştir.

Anahtar Sözcükler: *Aspergillus fumigatus*, Kendiliğinden düzenlenme, RodA, Arayüz, Moleküler dinamik simülasyonları

ABSTRACT

Comparative Molecular Dynamics Analysis of a Fungal Hydrophobin RodA Behavior at Interfacial Environment

Aspergillus fumigatus is an opportunistic human pathogen and causes severe respiratory diseases in immune-deficient hosts. The infection is mediated by the asexual spores (conidia) and the surface of the conidia is covered with special amphipathic proteins called hydrophobins. These proteins mask the cell wall components, especially polysaccharides that trigger an immune response against the pathogen. Class I RodA hydrophobin from *A. fumigatus* assembles into highly ordered amphipathic layers with a characteristic morphology resulting from the parallel arrangement of rodlet structures. The rodlets consist of amyloid-like structures rich in β -sheets as a similar amyloid aggregate of proteins resulted in neurodegenerative and metabolic diseases. Furthermore, hydrophobins can form layers at hydrophilic/hydrophobic interfaces. Although hydrophobins form the same classes assembled into layers with similar morphological character, the structural regions involve in the self-assembly vary. The key regions in RodA hydrophobin self-assembly and the possible building units of the layers were already proposed in earlier studies. However, the atomic details of the self-assembling mechanism are not known yet. In this thesis, all-atom MD simulations were performed at air/water interface to get insight into the initial steps of self-assembly. Partial positional restraints on the water molecules near the edges of water boxes to fasten aggregation and prevent interactions with periodic images. It is observed that clustering behavior highly depends on initial orientations, and a self-assembly model is suggested in which both L2 and L3 loops are aligned side-by-side rather than overlapping, turning air at the interface.

Keywords: *Aspergillus fumigatus*, Self-assembly, RodA, Interface, Molecular dynamics simulations

1 INTRODUCTION AND AIM

Hydrophobins are small, secreted proteins produced by filamentous fungi to function in varied roles and conditions in the life cycle of the organism. Reducing the surface tension of the liquid environment, attaching to the host surface, protecting the surface of the organism from the physical and chemical treats are some of the functions of hydrophobins (1–3). Hydrophobins are grouped into two classes based on the distribution of the hydrophobic residues across the protein sequence and behaviors of the proteins in the aqueous environment. Both monomeric and multimeric structural states of these proteins were found on the surface aerial structures of fungi and in the growth medium (1). The assembled structures may have different levels of strength and forms depending on the hydrophobin class. Furthermore, the assembling behavior of the hydrophobins can be varied depending on the surfaces that they interact. After secretion to the extracellular environment, hydrophobins may exist in different structural forms (e.g., monomeric, dimeric, and tetrameric) based on the protein concentration, pH and temperature. (4).

Including RodA, Class I hydrophobins assembles into highly ordered and stable layers on the aerial structures, such as conidia (asexual spores), aerial hyphae and/or fruiting bodies depending on the role of the hydrophobin in growth or reproduction (1). In several studies, it was shown that these layers consist of parallel rodlets which are rich in β -sheet structures. Furthermore, increase in this secondary structural content was also detected in the assembled structures at hydrophilic/hydrophobic interfaces (Figure 1) (5–12). In most cases of the Class I assembly, these interfacial environments were found to be necessary in layer formation. It was strongly suggested that the interfaces trigger the conformational changes in some hydrophobins by exposing the aggregation-prone regions located mostly on the loop regions. It is important to note that the loop responsible for self-assembly may differ between hydrophobins (12–15). On the other hand, Class II hydrophobins are able to form porous amphipathic monolayers at hydrophilic/hydrophobic interfaces. The monolayers of this class consist of repeating hexameric units of hydrophobin multimers that oligomerized without significant conformational rearrangements (16–20).

A. fumigatus is a filamentous fungus known as one of the opportunistic human pathogens. It produces conidia in its life cycle to reproduce. Mostly, inhaled conidia of the pathogen are eliminated by innate immune system of healthy individuals. However, in the patients with existing lung conditions or immunodeficiency, *A. fumigatus* can cause serious respiratory conditions and in some cases even death (21–23). The cell wall of the conidia is rich in α - and β - glucans that induce immune response of the host. However, the outer surface of the conidia is covered by the Class I RodA hydrophobins which form amphipathic rodlet layers that mask the pathogen glucans against immune recognition and response similar to a Trojan horse (Figure 1) (11,24). This was one of the earliest functions associated with these multifaceted proteins during human infection.

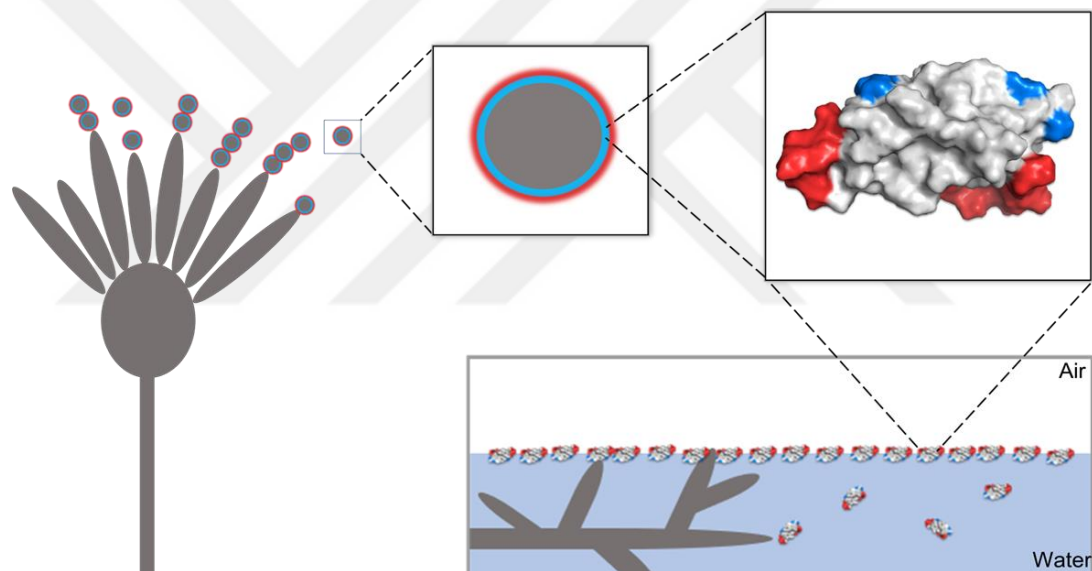


Figure 1. Schematic representation of RodA self-assembly on the surface of *A. fumigatus* conidia and at air/water interface. The hydrophilic and hydrophobic face of the RodA rodlet layer were shown as blue and red circles, respectively. Hydrophobic patches and charged residues were depicted as red and blue surfaces on RodA monomer. A representative mycelium and secreted RodA monomers were shown in water. The figure is modified from the study of Wösten and Scholtmeijer (25).

As with Class I hydrophobins, other proteins may also form amyloid aggregates to varying degrees and shapes with a core structure rich in β -sheets and ability to bind amyloid-binding dyes (e.g., thioflavin T (ThT)). These protein aggregates may have a function similar to hydrophobins or may result in pathogenic structures and cause neurodegenerative or non-neurodegenerative diseases. For example, amyloid β peptide

(A β), α -Synuclein and islet amyloid polypeptide are known to form amyloid fibrils associated with Alzheimer's and Parkinson's diseases, and type II diabetes, respectively (26). Due to this similar protein aggregation behavior, finding clues about the molecular mechanism of the self-assembly of hydrophobin layers and key structural features playing a role in aggregation may deepen the understanding also about the diseases caused by these types of protein aggregates.

To date, both Class I and Class II hydrophobins have been studied by many research groups to understand the self-assembly and surface adsorption behavior. All-atom or coarse-grained models of monomeric hydrophobins were generated and full-length folded, unfolded, truncated hydrophobins or only the loop regions that predicted to play a role in surface adsorption or self-assembly was monitored in bulk solution and at a wide range of hydrophobic/hydrophilic interfaces (7,13,19,24,27–32). The length, thickness, and repeating units of the rodlet layers formed on the hydrophobic surfaces or on conidia of several fungi were determined. Key protein sequence regions were predicted and/or studied by using mutagenesis in several Class I hydrophobins such as SC3, EAS, DewA, Vmh2, SC16 and RodA (11,14,27,33–35). A number of self-assembly models in solution (SC3, MPG1) at hydrophilic/hydrophobic interface (SC3, EAS, MPG1, SC16) or on conidia surface (RodA) were proposed for different hydrophobins with the help of experimental data, bioinformatics predictions and/or by considering only structural resemblance of these key regions (13,15,24,27,28,33,35,36). Although similar morphology and secondary structural content were observed, the region responsible for layer formation varies in Class I hydrophobins. However, the molecular mechanisms of rodlet formation and self-assembly of *A. fumigatus* RodA on conidia or at hydrophilic/hydrophobic interfacial environments are not fully elucidated yet. In this thesis, to gain insight into the initial steps of RodA self-assembly at air/water interfaces, all-atom Molecular Dynamics (MD) simulations have been conducted to monitor the effect of number of monomers, initial monomer orientation dependent clustering behavior, intermolecular interactions, and nonbonding interactions.

2 BACKGROUND

2.1 Functions of Hydrophobins

Filamentous fungi produce mesh-like structures called mycelium by branching cell wall-covered hyphae during the growth phase over different substrates. Aerial hyphae may differentiate, and form varied types of structures at the tips (sporangiophore, conidiophore, or oidiospores) or fruiting bodies containing spores to employ in asexual and sexual reproduction, respectively. Ascomycetes mostly produce conidia (asexual spores) on the tips of the branched structures (sterigmata) of aerial hyphae unlike sporangium (sac-like structure at tips containing sporangiospores) of Zygomycetes. On the other hand, fruiting bodies are mostly observed in Basidiomycetes. Hydrophobin expression is observed in these fungal phyla in diverse stages of the fungal life cycle to be employed in multiple functions. These special proteins possess surface hydrophobicity and have been found to be significantly expressed during the formation of aerial hyphae and detected on the hyphal cell walls in addition to fruiting bodies and, in some cases, gas channels. Moreover, most hydrophobins cover the surface of the conidia by forming amphipathic layers to ease aerial dispersion and mask fungal cell wall components (α - and β -glucans, chitins) that trigger the host immune system, aggregate at air/water interface to reduce surface tension for growth and mediate attachment to a plethora of surfaces (1,37).

SC3 hydrophobin from *Schizophyllum commune* is found to be expressed during aerial hyphae formation (37,38). Aerial hyphae and fruiting bodies of *S. commune* was shown to be covered with SC3 and the secreted hydrophobin also forms layers at the air/water interface and facilitate attachment to diverse surfaces for fungal development (1,37,39,40). The hydrophobin DewA from *Aspergillus nidulans* has been detected only on conidia but there is no evidence of its existence as rodlets on the surface of other fungal structures (41). EAS hydrophobin from *Neurospora crassa*, and MPG1 from *Magnaporthe grisea* were also found to cover the surface of conidia by forming amphipathic layers (42,43). Class II hydrophobins HFBI and HFBII have been found in cell walls, on the spore surfaces of *Trichoderma reesei* and in the growth medium

(44). *RODA* gene have been found to be expressed before the conidia formation; the transcript detected on the branches at the tips of conidiophores of *A. nidulans* and the protein on the conidia forming rodlet layers on *A. fumigatus* (11,37,41).

Post-transcriptional modifications (glycosylation) or possible modification sites on the protein sequence (for myristylation) were observed in SC3, SC16 and EAS, respectively (5,35,45). It was proposed that glycosylation of SC3 play a role in solid surface adsorption due to decrease in surface interaction after deglycosylation and α -helical structure which forms after surface adsorption (5). Additionally, it was hypothesized that addition of the lipids may play a role in the layer formation on the surface of fungal structures (45).

In all fungi species that express hydrophobins, hydrophobin monomers are secreted to the aqueous environment from the tips (apices) of the hyphae and aggregate into layers at the medium/air interface and provide the fungus a way to branch through hydrophobic environments by lowering the surface tension of the medium. As the amphiphilic hydrophobins orient themselves, the hydrophobic surface of the layer faces toward the environment and the hydrophilic side covers the hydrophilic cell wall (1,37,45). Models have been proposed for aerial and hydrophobic growth of the hyphae facilitated by the hydrophobin layers (46). According to the models concerning aerial growth, growing hyphae may be covered with the preformed hydrophobin layer when it reaches the air/water interface. The newly secreted hydrophobins extend this layer resulting in full coverage of the hyphal surface or the hyphal surface exposed to air lacking the hydrophobic character is subsequently filled with secreted hydrophobins at the interface (1). After the initial attachment to a solid surface, secreted hydrophobins from the tip of the hyphae may help the formation of a continuous hydrophobin layer between the hyphae and the surface. In the case of the presence of an aqueous medium, secreted hydrophobins may self-assemble at the hyphae/solid surface interface (maybe host surface) to facilitate growth without a need for fully coated hyphae (37).

2.2 The Role of Hydrophobins in Fungal Pathogenesis

It has been demonstrated in previous studies that hydrophobins are implicated in the hydrophobic interactions of fungal spores with animals and plants (37). Class I hydrophobin MPG1 is produced by the rice plant pathogen *M. grisea*. The hydrophobin was found to function in signaling the formation of infection structures (appressoria). To introduce a signal for the development of appressorium in the fungi, attachment to a host surface was determined to be required. The layer formation by MPG1 hydrophobin on the spore surface was suggested to facilitate this host-pathogen interaction in addition to the role of preventing fungi become wet (15,43).

O. ulmi (non-aggressive) and *O. novo-ulmi* (aggressive) are the two members of Ascomycota *Ophiostoma* and are known as pathogens of elm trees. Cerato-ulmin (CU) is a Class II hydrophobin and was found to be expressed in spore formation and highly correlated with aerial hyphae formation in these species (47). Although CU hydrophobin was determined to be toxic for elm trees in early studies, it was observed that even *O. novo-ulmi* lacks cu expression can cause pathogenicity. Targeted gene disruption studies revealed that the virulence of *O. novo-ulmi* was not affected by the absence of the cu gene products. Moreover, changing the production amount of CU in nonaggressive fungi *O. ulmi* was found to have no significant effect on pathogenic behavior. Therefore, it was proposed that CU is one of the virulence factors of these pathogenic fungi (1,47).

2.3 The Role of RodA in Fungal Pathogenesis

Apart from plant infection, it is well known that various genera of fungi (*Aspergillus*, *Candida*, *Cryptococcus*, and *Pneumocystis*) are human pathogens and can cause severe diseases or even death by mostly infecting the lungs, skin, and brain. Most fungi-related lung diseases, such as chronic pulmonary aspergillosis, resulted from infection by the opportunistic pathogen *Aspergillus fumigatus* and *Aspergillus flavors* of immune deficient and organ transplant patients and patients already suffered from non-fungal lung diseases (asthma, chronic obstructive pulmonary disease) (48).

Hundreds of thousands of conidia are inhaled by humans throughout their lifetime. However, most inhaled conidia of the pathogenic fungi are inactivated by the host immune system, except for patients with the aforementioned conditions. In *Aspergillus* species where RodA protein is produced, it is known that the hydrophobin self-assembles on the conidia surface and forms rodlet layers that completely cover the surface and mask the cell wall components (polysaccharides). Single or multiple cysteine residues were mutated individually or together, and the surface coverage behavior of these mutants was monitored in the study of Valsecchi *et al.* The mutant RodA hydrophobins were not detected among conidial components in the cytoplasm by using antibodies. However, the exposed pattern of cell wall polysaccharides of the mutant conidia, although rodlets were not formed. In the case of *RODA* deletion, and β -(1,3)-glucans were more exposed in contrast to wild type and cysteine mutants. Conidia cell wall polysaccharides (α -(1,3)-glucans and β -(1,3)-glucans) were found to be able to induce an immune response and cytokine secretion by differently affecting the level of dendritic cell activation due to dissociated (during germination) or disrupted RodA rodlet layers (49). Recently, RodA was found to interact with a component of the humoral immune system. The C3 protein of the humoral immune system was directly involved in the immune response to conidia and the activation of the C3 protein was observed after interaction of the protein with RodA in addition to β -(1,3)-glucans and galactomannan. The inactivation of the pathogen by the immune system was facilitated by the phagocytosis of the conidia. The activated C3 interacts with the CR3 and CR4 receptors which are employed in the phagocytosis. The inactivation of these receptors was found to affect the killing process of the conidia negatively (3).

2.4 Properties of Hydrophobin Structures

Fungal hydrophobins have less than 20 kDa molecular weight (2). The hydropathy pattern (Figure 2) and solubility of the hydrophobins divide the proteins into two classes. Ascomycetes produce only Class I unlike Basidiomycetes, which species express both Class I and Class II hydrophobins (1). Although the general fold is similar among hydrophobins, the amino acid sequences varied especially between eight

conserved cysteine residues. These sequence regions between the third and fourth (C3-C4), fourth and fifth (C4-C5), and seventh and eighth (C7-C8) cysteine residues were usually named L1, L2, and L3 loops, respectively, for simplicity in most studies (14,34). However, the sequence divergence is less significant within the loops of Class II hydrophobins (Figure 3) (45). Eight cysteine residues in the hydrophobin sequence form disulfide bridges in a specific pattern observed among all hydrophobins and two of these linkages are observed between core β -strands in both classes (Figure 4) (36). The cysteine pairs form disulfide bonds as follows: C1-C6, C2-C5, C3-C4, and C7-C8 (11). The core of the hydrophobin structure consists of characteristic four β -strands which are supported by these disulfide bonds (2,36). Negatively charged residues are dominant in the hydrophobins (NC2, DewA, EAS, MPG1 and RodA), and most of the charged residues in the hydrophobins are located on a single surface, except and SC16.

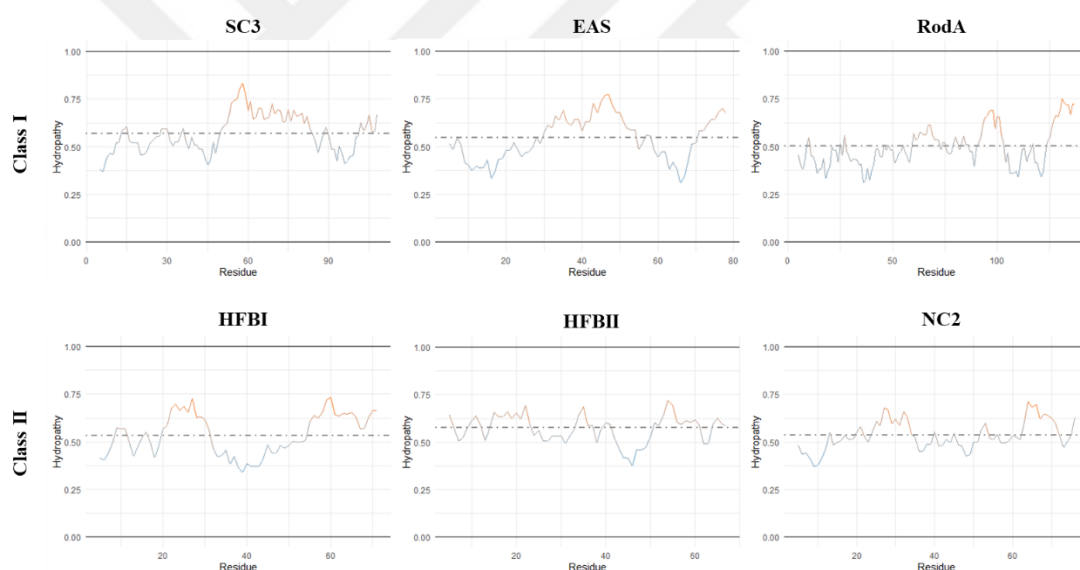
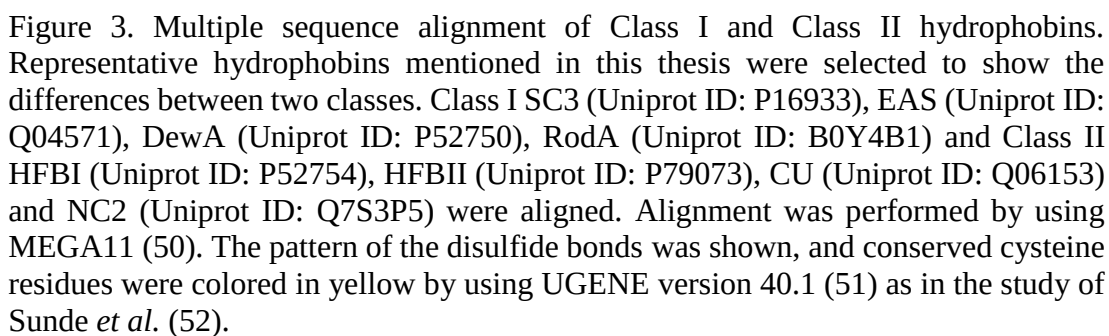


Figure 2. Scaled hydropathy plots of representative hydrophobins. Class I SC3 (Uniprot ID: P16933), EAS (Uniprot ID: Q04571), RodA (Uniprot ID: B0Y4B1) and Class II HFBI (Uniprot ID: P52754), HFBII (Uniprot ID: P79073) and NC2 (Uniprot ID: Q7S3P5) were plotted. Plots was generated by using Bioconductor by setting window size 9. Dashed line depicts the average value.



The NMR solution structure of EAS hydrophobin contains two unstructured loops (L1 and L3) and a shorter L2 loop (Figure 4). The core and L1 loop are determined to be significantly hydrophobic. Two significant hydrophobic and hydrophilic surfaces were observed. The charged residues are mostly concentrated on a face making it hydrophilic (36). In the structure of DewA, L1 and L2 loops contain secondary structural elements (two helices in L1 and a β -sheet and a small helix in L2) and L3 loop is shorter than these two loops and does not form any ordered structure (Figure 4). The charged residues are in proximity to each other forming the hydrophilic face of the hydrophobin (14). Unlike other Class I hydrophobins mentioned in this thesis, the core of the MPG1 hydrophobin consists of an open β -barrel of four β -strands. The hydrophobic and hydrophilic surfaces can easily be observable on the protein surface. A large portion of L1 loop is determined to be flexible (15). The longest loop is L1 loop in SC16 and the helical structure is located at the center of the loop. The hydrophobic region of SC16 is found to be buried between the central β -barrel and L1 loop. The surface of the protein is covered with many charged residues in addition to polar, uncharged residues (Figure 4) (35).

Class II hydrophobins also have β -sheets at their center, but loop regions are significantly shorter (Figure 4). The hydrophobic patches of HFBI and HFBII consist of residues from both L1 and L3 and most of these loops are the strands of the β -core and β -hairpins (53,54). As in MPG1, NC2 has a β -sheet at its core in an open form. L1 and L3 loops are unstructured, unlike L2 loop. A small number of residues from N-

termini, C1-C2 region, L2 and L3 loops all contributed to the hydrophobic surface of the hydrophobin (18).

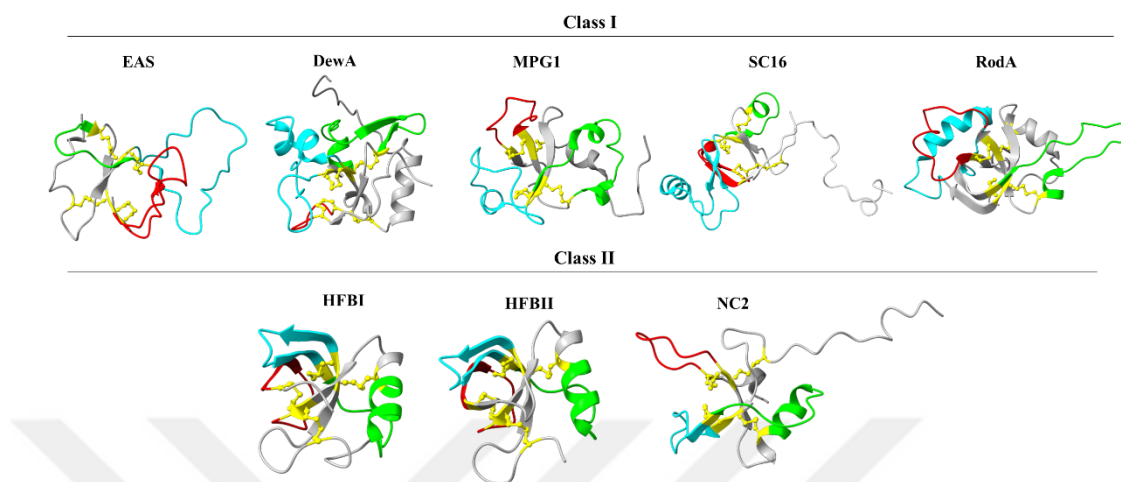


Figure 4. 3D structures of representative Class I and Class II hydrophobins. Class I EAS (PDB ID: 2FMC), DewA (PDB ID: 2LSH), MPG1 (PDB ID: 2N4O), SC16 (PDB ID: 7S86), RodA (PDB ID: 6GCJ), and Class II HFBI (PDB ID: 2FZ6), HFBII (PDB ID: 2B97), NC2 (PDB ID: 4AOG) were visualized. L1, L2 and L3 loops and disulfide bonds were colored in cyan, green, red, and yellow, respectively. For RodA, Regions L99-Q104 and N66-Y70 were not considered as L1 loop since these regions form part of the central β -sheet. Hydrophobin conformations were visualized by using ChimeraX version 1.4 (55).

An intermediate class of hydrophobins was identified in addition to Class I and Class II hydrophobins. Currently, the knowledge about these Class III hydrophobins is that they also have characteristic cysteine residues, but the distinct hydrophobic patches observed in Class I and II are not detected in Class III (11).

The NMR solution structure of RodA lacking N-terminal (residue positions 1-38) was determined in 2019 (11). The core of the monomeric hydrophobin consists of β -strands which form a closed barrel-like structure as already seen in other hydrophobins. Helical structures with varied sizes are adopted in L1 loop, after the C-terminal of L2 loop, the region near the first cysteine, and the C-terminal of the monomer (Figure 4). Two disulfide bonds stabilize the β -barrel and others are formed between the barrel to the first cysteine positioned on the C-terminal of a helical structure and to the cysteine positioned on the other helix at the C-terminal of L2 loop. In NMR measurements, significant portions of the loop regions between cysteine

residues were found to be dynamic, especially L2 and L3 loops. Hydrophobic sequence regions were observed in the loops L2 (114-120) and L3 (136-148). The hydrophobin surface is simply divided into two: (1) a hydrophobic surface that contains L2 and L3 loops and charged residues that neutralize each other in addition to a few isolated positively charged residues and (2) a surface with the dominance of negatively charged residues. The total number of charged residues in full length RodA was significantly high (ASP:10, LYS:8, GLU:3, ARG:1) compared to the other hydrophobins mentioned in this study (Figure 4) (11).

2.5 Self-Assembly and Layer Formation

Hydrophobin layers were found to be formed on the spores of Ascomycetes, Basidiomycetes, and Zygomycetes. Hyphae of the fungi are also covered with these layers except the species from Ascomycetes (37). Both Class I and Class II hydrophobins form amphipathic layers in varied forms and sizes on the surfaces of aerial structures of fungi and at hydrophilic/hydrophobic surfaces. In Class I, β -sheet rich amyloid-like amphipathic rodlet layers are formed by mostly less ordered soluble hydrophobins after conformational changes. These highly stable layers can only dissolve in strong acids (formic acid, trifluoroacetic acid) by the disruption of hydrophobic interactions. On the other hand, Class II hydrophobins do not assemble into rodlets, and layers of this class can be solved in organic solvents and detergents (2,45).

The self-assembly behavior of the hydrophobins has been observed firstly in the studies carried out with purified Class I hydrophobin SC3 (45). It was observed that the interaction of SC3 with air/water, oil/water, and solid hydrophobic surface/water interfaces resulted in the formation of layers of rodlet which have similar morphology to amyloid fibrils (9,45,56). Highly ordered amphipathic rodlet structures after conformational changes at the air/water interface and solid surfaces were observed and the thickness of layers and the distance between individual rodlets were measured as 10 nm (5,28,57). It was proposed that SC3 self-assemble into multilayers on solid surfaces (5) and non-covalent interactions (most probably hydrophobic interactions)

proposed to play a role in these hydrophobin layers (28,57). In addition, SC3 layers with varied permeability degrees were obtained and this behavior is related to the morphology of the layers in different hydrophobin concentrations (56).

SC3 is found to be present in the aqueous environment as monomers, dimers, and tetramers or higher state oligomers (nanorods) in solution depending on the pH, ion, and protein concentrations. It was claimed that dimers further aggregated and resulted in longer oligomers (10,28,30). Basic pH and low protein concentration provide an environment for SC3 to assemble in solution (10) and the hydrophobin was found to be unassembled at the interface if the protein concentration is below 3.7 g/mL (28). It was proposed that the solubility of the monomeric protein is provided by intermolecular disulfide bonds (58). In the study of Scholtmeijer *et al.*, it was claimed that some cell wall components, especially β -glucans, secreted into the aqueous environment and facilitate spontaneous hydrophobin assembly at both air/water and hydrophobic solid surface/water interfaces (30).

The secondary structural change adopted by SC3 monomers differs depending on the adsorbed interface and surface hydrophobicity affects the degree of hydrophobin self-assembly (30,39). The intermediate α -helical state is dominant on the solid hydrophobic surface, unlike air/water interfaces where β -sheet content is observed to increase progressively in time due to β -sheet stacking (detected with ThT and Congo red which binds only these structures) after initial helical structure formed (5,9,56). Additional chemical treatment is required for this intermediate conformational state to reach the final β -sheet rich structure (6). In addition, MD simulation studies performed to monitor the adsorption behavior of both unfolded and homology model of SC3, which was predicted by using experimentally solved structure of HFBII, supported the results that increased β -sheet structures formed in SC3 at the hydrophilic/hydrophobic interfaces (7,27).

It was suggested that the L1 loop (C3-C4) of SC3 is responsible for the surface adsorption and was found to bind Teflon with high affinity, despite the loop was also proposed in the MD simulation studies to be at the protein-protein interface of dimeric

hydrophobin in solution (27,59). At the interface MD simulations, a portion of L1 loop was used to monitor dimerization. However, the dimer dissociated after hydrophobin adsorption to a solid surface (27).

SC16 from *S. commune* was also determined to be a Class I hydrophobin and requires a hydrophobic/hydrophilic interface to self-assemble and probably form monolayers since the height of the layer was measured as ~2 nm (35,60). The L1 loop is predicted as an amyloidogenic region by a prediction algorithm and a possible conformational change was suggested prior to oligomerization (35).

EAS hydrophobin is monomeric in solution due to highly disordered L1 loops (12,29). Coarse-grained MD simulations reached μ s long simulation time performed with different lengths of multiple monomeric hydrophobins, in which the L1 loop was shortened or completely deleted. Replica Exchange MD simulations (REMD) showed that relatively large L1 loop provided solubility to the hydrophobin in varied temperature conditions. EAS aggregation favored lower temperatures (29). As in all Class I hydrophobins, it was shown with spectroscopy measurements, and staining methods that β -sheet structures increased in EAS rodlets as in amyloid fibers and hydrophobic/hydrophilic interface mediates rodlet formation (8,12,36). Unlike SC3, it was observed both in experimental studies and solution and air/water interface MD simulations performed with full length EAS and shortened L1 loop (15 amino acids) that the loop was not a crucial player for rodlet formation and adsorption to surfaces and it was proposed that it may involve in lateral interactions (8,36). Moreover, surface adsorption of EAS monomer was studied by using all-atom MD simulations, and it was observed that there were hydrogen bonding interactions between L1 and L3 loops, and water-mediated adsorption to a solid surface through the L1 loop made L3 loop available for hydrophobin self-assembly (4). The first model was proposed for the EAS rodlet formation at the air/water interface. It was claimed that β -barrels form the core of the rodlets by getting arranged in an order in which hydrogen bonds formed between adjacent β -barrels of monomers and the hydrophobic surface of the monomers faces towards air as hydrophilic (charged) surface is oriented to water (36). However, studies conducted with mutated EAS hydrophobin revealed that a portion of the L3

loop is crucial in hydrophobin self-assembly where it forms amyloid-like β -cores of the EAS rodlets. It was claimed that L3 loops stacked in antiparallel monomers to avoid steric clashes in the EAS oligomer. A model of EAS oligomerization was generated by using molecular docking techniques (33). Additionally, a three-level layer formation model was suggested by Morris *et al.* in which conformational changes are required after adsorption to the surface to increase intermolecular interactions and form a highly ordered β -sheet-rich layer. Protein concentration does not affect the self-assembly of EAS; there was no higher state EAS hydrophobins observed in the solution other than monomers despite higher concentrations. On the other hand, EAS aggregation and layer formation are observed to be negatively affected in varied solvent conditions that decrease surface tension (13).

The self-assembly of Class I DewA is also mediated by hydrophilic/hydrophobic interfaces. In solution, DewA was found to be dimerized in high concentrations. However, there were two different monomer conformations observed for DewA in which N-terminal and L1 loop adopt different structural states and only one of them was able to form dimers. It was demonstrated that both monomer conformations can form rodlet layers in a range of hydrophilic/hydrophobic interfaces. The key region(s) responsible for the self-assembly is not experimentally determined yet but, predicted as a portion of the L2 loop (K69-V82) by prediction algorithms and claimed that the L2 loop extends and forms β -sheets as in the case of the L3 loop of EAS (14).

In the studies conducted with Vmh2 regarding hydrophobin self-assembly, it was concluded that a hydrophobic/hydrophilic interface is not necessary to form amyloid-like structures. By using dynamic light scattering (DLS) and transmission electron microscopy (TEM), it was observed that, at neutral pH, the protein formed oligomers with different lengths and short but thicker structures varying depending on hydrophobin concentration. The presence of ions (Ca^{2+}) and temperature increase have a positive effect on Vmh2 self-assembly. It was claimed that the L1 loop is destabilized by various factors (pH, ion), and this conformational change expose the initiation point of self-assembly. Additionally, it was observed that β -sheet rich structures were formed over time, independent of the hydrophobic/hydrophilic interface, and the

Vmh2 hydrophobin, unlike other Class I hydrophobins, showed a twisting behavior similar to other amyloid structures (34,61). A model of Vmh2 oligomer that was suggested to form at the air/water interface was predicted by molecular docking with the guidance of the earlier study of Macindoe *et al.* for EAS oligomerization (33,34).

As in the case of most Class I hydrophobins, a hydrophobic/hydrophilic interface is necessary for MPG1 to form rodlets. Atomic force microscopy (AFM) measurements of dried hydrophobin on highly ordered pyrolytic graphite (HOPG) revealed that rodlet layers with ~2.5-15 nm thickness formed on the surface. A model for self-assembly at the air/water interface is suggested. In this model, MPG1 monomers were claimed to be more compact in solution, and adsorption to the air/water interface provides a way for hydrophobin to adopt an elongated structure. After the incorporation of the monomer into the preformed oligomer, a second conformational rearrangement occurs to form a proper fibril-like structure. L3 loop was proposed to be the region responsible for the amyloid formation in the rodlets since its conformation was found to be similar to the L3 loop in EAS oligomer (15).

Models for HFBI and HFBII hydrophobin films were proposed by using molecular docking methods with the help of electron microscopy (EM) measurements (19). It was determined that these hydrophobins form ordered porous monolayers at the air/water interface and the subunits of these layers were proposed to be symmetrical hexameric structures (16,17,19). There was no difference observed between the solution and at water/octane or air/water interface structures of the hydrophobin monomer (31,62). Tetrameric structures were observed in high protein concentrations in solution. Oligomer size decreases in low protein concentrations and both monomers and dimers are presented in solution (16,62–64). Brownian dynamics simulations were performed with a large amount of HFBI by neglecting the hydrophobin-water interactions (using implicit solvent) and μ s long simulation trajectories were generated. Non-polar interactions mediated tetramer formation was observed. Hydrophobic surfaces of HFBI were found to be located at the interface of the hydrophobin tetramers (32). Although the hydrophobic surface was also found to be buried in the oligomers of HFBII, it was proposed that these surfaces probably face

toward the air at the air/water interface to form an amphipathic layer (65). Coarse-grained representation of HFBII with varied particle properties (hydrophilic, hydrophobic, or uniform) leads to getting an insight into the role of the hydrophobicity pattern of protein surface in surface adsorption (31). Site-directed mutagenesis on the charged residues of HFBI revealed that these residues involved in oligomerization and interaction with both solid and liquid surfaces. It was shown that each substitution, somehow, affected a function of this multifunctional hydrophobin (66). In recent studies, mutations introduced in the cysteine residues were found to result in the disruption of the integrity of the hydrophobic patch of HFBI and affect the protein self-assembly both at the interface and in solution negatively (67,68). The aggregation mechanism of HFBI and HFBII hydrophobins at the interface (water/air) has been investigated with the help of mutations introduced in the charged residue side chains. In this study, it was concluded that side-to-side interactions of proteins are effective players in self-assembly at the interface and the rate may vary according to electrostatic interactions (69).

Monolayer formation of Class II NC2 from *N. crassa* on the hydrophobic surface was determined. These monolayers were found to resemble HFBI and HFBII layers which show a layered morphology formed by hexamers (18). MD simulations and molecular docking performed with NC2 have shown that the hydrophobin was highly surface active and the building block of the monomeric layer was hexamer units. L3 loop was suggested to be positioned at the hydrophobin-hydrophobin interface in the layers (20). Unlike HFBI and HFBII, the hydrophobin was found to be monomeric in solution (18).

On the conidia surface, RodA hydrophobins were detected as ordered multilayers with the help of TEM and AFM measurements. These layers were proposed to be constructed from bilayers (~5 nm) and these bilayers contain parallel rodlets of four RodA monomers. In solution, it was suggested that oligomers of RodA (most probably dimers formed by stacked β -sheets) may also form as SC3, and these oligomers may be the building blocks of the parallel rodlets (The average distance between two rodlets was measured as 11 nm) which are progressively formed through the overlap of the

first and second monolayer in the final bilayer on a conidium. It was proposed that hydrophilic surface of the monomers in layers was buried in the inner part of the bilayers since the outer surface of the rodlet layers on the conidia was found to be hydrophobic. In this case, RodA in the inner first monolayer also pointing its hydrophobic surface to conidia cell wall and non-bonded interactions was claimed to occur between hydrophobin residues and cell wall components (24).

The AFM method was used in an attempt to image nanodomain formations in the rodlet layers on the *A. fumigatus* conidia. A significant angle was observed to present between parallel rodlets of two adjacent domains. The reason for nanodomain formation was claimed to be resulted from the water layers trapped between nanodomains and negatively charged residues on the previously formed layers on the conidia surface. Monte Carlo simulations were conducted to understand the mechanism of the coverage of the conidia surface as a single RodA layer by permitting only side-to-side arrangements of the hydrophobin oligomers on 2D plate surface (24).

Unlike other Class I hydrophobins, two regions were determined to be involved in the cross- β structures formed in RodA self-assembly. Two stretches in both L2 and L3 loops (amino acid positions 112-117 and 143-147, respectively) were found to have a key role in rodlet formation probably by positioning in the stacked β -sheets on the conidia surface. Disruption of these predicted amyloidogenic stretches by introducing mutations on the hydrophobin sequence that alter the physicochemical properties of these loops together resulted in a lack of full coverage of conidia surface by hydrophobin layers and disability to mask the cell wall components that trigger an immune response. On the hydrophobic surface (HOPG), rodlet formation was not detected for mutant RodA, which contains two mutations in L2 and L3 loops, at room temperature but, changing incubation conditions facilitate the formation of rodlets as in the case of wildtype RodA. The tendency for self-assembly of the stretches from L2 and L3 loops were studied *in vitro*, and it was observed by TEM images that both peptides form fibrils with different morphologies and sizes. On the other hand, N-terminal of RodA (residues up to the 39th position) was found to be not necessary for the rodlet formation on the conidia surface unlike cysteine residues. Mutagenesis

studies of *RODA* gene have revealed that cysteine residues are mandatory for rodlet formation on the conidia surface (11).

2.6 Molecular Dynamics Simulations

It is well known that in some proteins, large-amplitude motions result in conformational rearrangements which are induced by diverse reasons, such as protein-protein interactions, presence of ions, temperature change, protein concentration, and surface interactions. These changes are crucial for their function (70). For example, voltage-gated sodium channels which have open and close conformations to mediate or block ion transport or binding pocket of G proteins in G protein-coupled receptors are relatively buried and helical of the protein opens to bind GDP or GTP depending on the interaction with receptor proteins (71). In the case of Class I hydrophobins, it is known that hydrophobic surface adsorption results in conformational change or secondary structural changes that occur in self-assembly. In experimental methods, atomic-level details of these processes cannot be monitored. MD simulations are computational methods employed to mimic the behavior of molecular systems in this wide range of conditions as a function of time. To obtain a trajectory of the protein motion step-by-step, Newton's equation of motion is solved by using algorithms (one of the widely used algorithms is Verlet algorithm (72)) developed by implementing Taylor Series Expansion (Equation 2.6.1.1) (73).

$$m_i \frac{d^2 r_i}{dt^2} = f_i = - \frac{\partial}{\partial r_i} U(r_1, r_2, \dots, r_N) \quad (2.6.1)$$

In each time step, forces acting on each atom and, accelerations and velocities are calculated, and the new positions of atoms are determined. The initial positions of the atoms (cartesian coordinates) in the biomolecules were obtained from experimental methods (X-ray diffraction, NMR) or quantum mechanics, and the velocities were assigned to these atoms by considering Maxwell distribution at the given temperature (73). The forces are derived from the functional expression of the total potential energy of the system (force field) which consists of two components: bonded and non-bonded

energy terms (Equation 2.6.1.2). The parameters used in the force field are derived from the experiments (72).

$$U = \sum_{bonds} \frac{1}{2} k_b (r - r_0)^2 + \sum_{angles} \frac{1}{2} k_a (\theta - \theta_0)^2 + \sum_{torsions} \frac{V_n}{2} [1 + \cos(n\phi - \delta)] + \sum_{improper} V_{imp} + \sum_{LJ} 4\epsilon_{ij} \left(\frac{\sigma_{ij}^{12}}{r_{ij}^{12}} - \frac{\sigma_{ij}^6}{r_{ij}^6} \right) + \sum_{elec} \frac{q_i q_j}{r_{ij}} \quad (2.6.2)$$

In Molecular Mechanics, bonded atoms are represented as masses connected to each other with springs and atomic vibrations (harmonic oscillations) occur between bonded atoms in the proteins as in all other types of molecules. These vibrations, such as bond stretching between two adjacent atoms, angle bending between three connected atoms, and torsional motion due to bond rotations, requires the incorporation of different empirical energy terms to the force field (the first four-term in Equation 2.6.1.2) (70,73). The energy terms non-bonded interactions are described by using Coulomb's law for electrostatic interactions and Lennard Jones 12-6 potential (the last two terms in Equation 2.6.1.2) (73). The force fields are widely used in the current studies are CHARMM (74), AMBER (75), and GROMOS (76) which are updated periodically. The potential energy caused by water molecules in the simulation systems is also considered separately by using water models developed for varied purposes.

Prior to the MD simulations, the simulation system should be prepared in accordance with the purpose of the simulation regardless of the MD software to be used. Here, the steps of preparation for simulations performed with NAMD (77) is explained. In the first step, the naming of the atoms and residues should be set to a proper name readable by the program, and topological information of each residue should also be defined, and missing atomic coordinates are predicted. In the case of cellular or secreted proteins, solvation is required to study them for varied purposes, and it was suggested in the simulations performed for DNA that the padding thickness should be at least 10 Å to prevent interactions with periodic images (78). Ions may be added to neutralize the system (Protonation states of the residues, ion concentrations,

or type of the solvent may vary depending on the aim of the study.) Therefore, the files of the solvated and ionized system are generated at the end of these processes. A configuration file is also needed to specify the input or parameter files and define the parameters and run time.

In a simulation box, only a limited number of solvent molecules can be used. Solving the equations of motion in each time step requires a significant computational source and time. Thus, keeping the simulation system as small as possible is efficient to obtain a trajectory in a reasonable time. However, it is possible to create an infinite solution environment by applying periodic boundaries in all directions of the simulation box. Periodic boundary conditions (PBC) create replicas surrounding the original box and molecules leaving the box can reenter with the same velocity.

Most MD simulations are performed in four steps: energy minimization, heating, equilibration, and production. The high potential energy of the system can be lowered by using various types of iterative methods, e.g., conjugate gradient and steepest descent algorithms in the minimization step (70). The number of steps to minimize the energy may vary depending on the system. However, the minimization algorithm should be converged at the end of the run. Heating is performed by gradually increasing the temperature of the system and in equilibration, both structure of the protein and the energy of the system is equilibrated. Harmonic restraints can also be applied to protein or solvent to minimize and heat the system to prevent large fluctuations in the energy. Production runs are then performed to explore the motion of the system in a predefined simulation time by calculating the position of each atom in each time step. MD simulations can be performed in both NVT or NPT ensembles by using proper algorithms to control the temperature and pressure of the system and one of the most used methods is Langevin dynamics for temperature and pressure control.

To reduce computational cost and explore the conformational space longer, a range of methods are used. The non-bonded interactions between three sets of atoms are neglected in energy calculations. Bond lengths are fixed to prevent high-frequency

vibrational motions, especially in water molecules (e.g., SHAKE algorithm). In addition, coarse-grained representation of the atom groups in protein (or solvent) can be used to perform longer simulations (73).

2.6.1 Trajectory analyses

Trajectories generated from MD simulation runs can be analyzed by considering various aspects depending on the purpose of the study. The most common analyses are the investigation of the stability of the entire protein structure, protein compactness, and per-residue fluctuations by calculating Root Mean Square Deviation (RMSD) (Equation 2.6.2.1), Radius of Gyration (Rg) (Equation 2.6.2.2) and RMSF (Equation 2.6.2.3), respectively.

$$RMSD = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_{1,i} - x_{0,i})^2 + (y_{1,i} - y_{0,i})^2 + (z_{1,i} - z_{0,i})^2} \quad (2.6.1.1)$$

N denotes the number of atoms, and the average of the vector distance between the coordinates of the atoms in reference structure (x_0, y_0, z_0) and the structure at the specified time step (x_1, y_1, z_1) is calculated in RMSD (79).

$$R_g = \sqrt{\sum_{i=1}^N (r_i - R_c)^2 / N} \quad (2.6.1.2)$$

The center of mass (R_c) of the protein structure and atomic positions (r_i) is required to calculate Rg (80).

$$RMSF_i = \sqrt{\frac{1}{T} \sum_{j=1}^T \left(\vec{r}_{j,i} - \frac{1}{T} \sum_{k=1}^T \vec{r}_{k,i} \right)^2} \quad (2.6.1.3)$$

RMSF of a single atom i with the position vector $\vec{r}$ is calculated by considering the total number of structures in the trajectory (T) (79).

The secondary structural content of the proteins may vary during the simulations and these changes can be monitored as non-bonded interactions by using a wide range of algorithms. In the case of analysis of the systems with multiple proteins, protein-protein contact maps or oligomerization can be determined by defining contact cutoffs and measuring the distance between specified residue regions. The most widely used trajectory analysis and visualization programs are VMD (81) and Chimera (55).



3 MATERIALS AND METHODS

3.1 Prediction of the Key Structural Regions in Self-Assembly

To date, aggregation-prone regions in amino acid sequence of RodA were determined via experimental (by introducing point mutations in *RODA* gene or protein sequence) and computational approaches (predicting the regions by using web-based tools). The regions that play a role in hydrophobin self-assembly were predicted by using bioinformatics methods and/or by introducing mutations into the corresponding region in earlier studies as in the case of RodA (11). In order to come a consensus and validate the previously predicted and experimentally determined regions, Class I hydrophobins with available data for their key structural players in self-assembly, in addition to RodA, were used to perform the new predictions. The web-based prediction tools that were commonly used for these Class I hydrophobins were used to re-predict these regions and validate the previous results.

The first prediction tool was Amylpred2 (82) (http://thalis.biol.uoa.gr/AMYL_PRED2/) which outputs the sequences with a high amyloidogenic propensity of a given protein sequence by a consensus of eleven different algorithms. The predicted regions are called hits and determined by finding the overlapping regions predicted by at least half of the used algorithms. In the prediction processes of this thesis, each algorithm available in the tool was used to predict the aggregation propensity of the hydrophobin sequences.

TANGO (83) (<http://tango.crg.es/protected/academic/calculation.jsp>) performs predictions by making physicochemical assumptions about the protein aggregates and predicting the secondary structural states of the given regions of the protein sequence by considering the statistical approaches. The output of the run is simply the tabulation of the probabilities of secondary structural states (β -sheet, β -turn, α -helix) and aggregation types (β - and helix aggregation) adopted by each residue in a peptide or denatured protein. A region with high aggregation probability of at least 5% and including 5-6 residues is assumed as highly prone to aggregation. These regions were

identified in the selected Class I hydrophobins to identify a pattern or differences. The effects of change in pH, temperature and ionic strength were included in the predictions.

In addition to tendency predictions for amyloid formation of hydrophobin sequences, regions that may involve in protein-protein interactions were also predicted by using CPORT (84) (<https://alcazar.science.uu.nl/services/CPORT/>) since these aggregation-prone regions have high probability to be at the protein-protein interfaces. A score-based consensus of the six different web tools, aimed to predict interface regions in proteins, was considered in CPORT algorithm. Predictions were performed by providing only the files that contain 3D structural information of the hydrophobins. The result page of a CPORT run includes the active and passive residues that are predicted as residues with high propensity interacting with the partner protein(s) and residues that are likely to play a role in interacting, respectively. However, only the residues that were predicted as active due to their higher probabilities in protein-protein interactions were considered.

3.2 Molecular Dynamics Simulations

Four different simulation systems were prepared with a single and multiple monomers by orienting multiple monomers in different initial orientations to increase the conformational sampling: Monomers were placed randomly in the box simulation box (Orientation 1), tips of the L2 loops were pointed to the center of mass (COM) of monomers (Orientation 2 for four-monomer systems) or located in close proximity to each other through their long axis (the axis) (Orientation 2 for two-monomer systems), L3 loops were pointed to the COM of monomers (Orientation 3), and loops were distantly placed from each other and the distance between monomers were increased compared to the other systems. (Orientation 4) (Figure 5). Both bulk solution and air/water interface simulations were performed for each simulation system to compare the behavior of the hydrophobins in these environments.

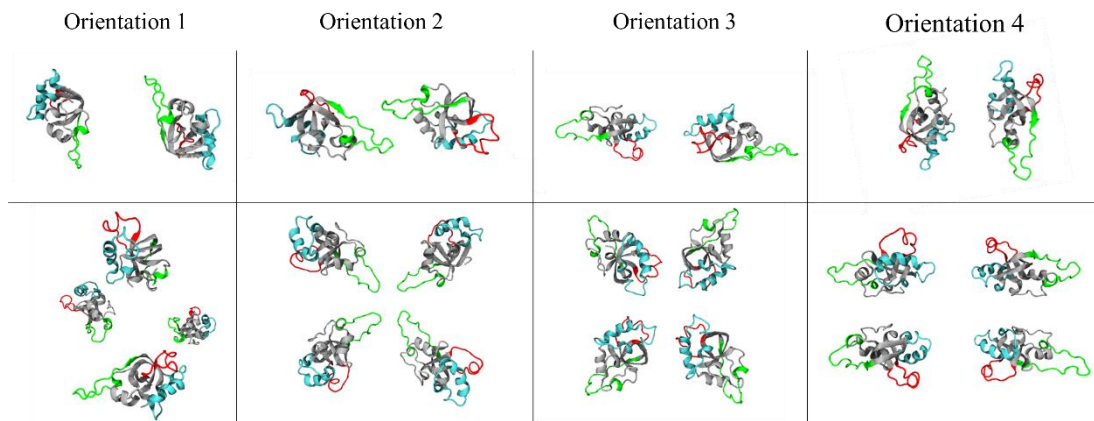


Figure 5. Four different initial orientations of two and four RodA monomers simulation systems. L1, L2, and L3 loops and non-loop regions were denoted in cyan, green, red, and gray cartoon representations, respectively.

MD simulations were conducted by using NAMD 2.14 (77) (NVIDIA CUDA acceleration used for equilibration and production runs) with CHARMM36m forcefield (74), and TIP3P (85) water model. Prior to MD simulations, the atomic coordinate file of RodA three-dimensional structure (PDB ID: 6GCJ) was downloaded from the Protein Data Bank (PDB) as a pdb file format (86). Protein structure files (PSF) were prepared by using psfgen package in VMD version 1.9.4 (81). RodA monomers were solvated with varied amount of water molecules by immersing the macromolecule into water boxes by using Solvate Plugin 1.7 and sodium ions were added to neutralize the system by using Autoionize Plugin 1.5. The details about the simulation boxes were given in Table 1.

Table 1. Number of atoms and box dimensions of each simulation system.

Orientation	Number of atoms	Initial Box Dimensions (Å)
Single Monomer		
1	35365	$72 \times 72 \times 72$
Two Monomers		
1	50117	$66 \times 113 \times 71$
2	49640	$68 \times 124 \times 63$
3	56186	$63 \times 145 \times 66$
4	48050	$64 \times 97 \times 83$
Four Monomers		
1	111202	$103 \times 110 \times 104$
2	112159	$130 \times 128 \times 72$
3	106462	$125 \times 137 \times 66$
4	108196	$102 \times 75 \times 150$

Monomer simulation systems were prepared by using CHARMM-GUI Input Generator (87) (<https://www.charmm-gui.org/>). Hydrophobins were immersed into the water with the padding thickness of 15 Å from the minimum/maximum coordinates of the protein(s). In the case of multiple monomers, proteins were rotated in the simulation box to minimize the volume of the water. The configuration file was generated by optimizing the parameters of the physical conditions of each simulation system. All simulations were performed at 298 K (ambient temperature) and 1 atm. The integration timestep was set to 2 fs. The trajectory files of production runs were saved in every 1000 steps. The potential energy of the simulation systems was calculated by restraining covalent bonds using SHAKE algorithm (88). Non-bonded interactions were calculated in each timestep and full electrostatic interactions in 2 timesteps intervals. Electrostatic and van der Waals interactions were considered in the 12 Å cutoff. Periodic boundary conditions were applied in all directions of the simulation system and particle mesh Ewald (PME) summation was used to calculate long-range electrostatic interactions. The simulation systems were equilibrated in NVT ensemble. Temperature control was performed by using Langevin dynamics by setting coupling constant to 1 ps⁻¹. Nose-Hoover Langevin piston was employed to control the system pressure by setting oscillation and damping time constants to 100 fs and 50 fs, respectively. Energy minimization was performed in different time steps depending on the simulation system with conjugate gradient algorithm, followed by heating, gradually from 98 K to 298 K by 1 K at 0.2 ps intervals. Positional restraints were applied to monomer backbone in minimization and heating runs and were gradually removed in 2 ns and the monomers was set as fully flexible in the last 2 ns of the equilibration runs in multiple monomers simulations. In the case of single monomer, restraints were removed gradually in 3.6 ns. Solution MD simulations were performed in NPT ensemble. NVT production simulations of air/water environments were conducted after short NPT runs to decrease large fluctuations both in water molecules and the monomers. Air/water interfaces were created by doubling the length of simulation box in z-dimension after short NPT runs. The details about the conditions of the simulations were given in the Table 2.

Table 2. Duration of each run performed in this thesis.

Simulation Environment	Minimization (Steps)	Heating (ns)	Equilibration (ns)	Short NPT (ns)	Production (ns)
Single Monomer					
Solution	40000	0.04	3.6	-	50
Interface	40000	0.04	3.6	-	50
Two Monomers					
Solution	40000	0.04	4	-	50
Interface	40000	0.04	4	-	50
Four Monomers					
Solution	50000	0.04	4	1 (RWM)	100
Interface	50000	0.04	4	10 (UWM) 1 (RWM)	100

Since the monomers were independent and mobile in the simulation box, they may leave the main simulation box during simulations and make contacts with their periodic images. To prevent the RodA monomers from this behavior, positional restraints were applied to the oxygen atoms of the water molecules. Prior to the production simulations, short NPT runs were performed to equilibrate the system pressure and water molecules. Positionally restrained water molecules (RWM) located within the 2 Å of the edges of the simulation boxes were selected in the last timestep of the trajectories of short NPT runs. These artificial water walls were created in each dimension of the simulation boxes in both simulation environments except water molecules on the xy-plane air/water interface simulations (Figure 6). Force constants used in the calculations of the constraint potentials were set to 1 kcal/mol/Å² for the restrained atoms and given in the beta factor column of the pdb files obtained from the last frames of the trajectories of short NPT runs.

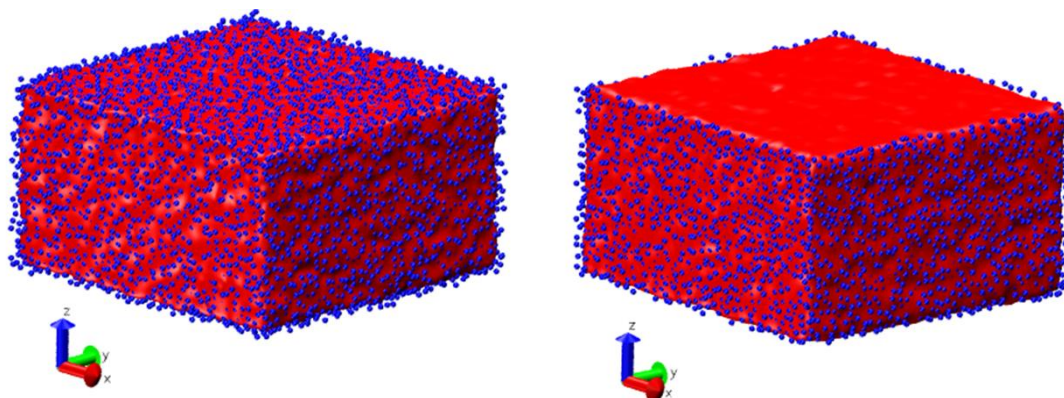


Figure 6. Representative solution (left) and interface (right) simulation systems with 2 Å thick water walls. The water box and positionally RWM were depicted in red surface and blue points, respectively.

3.2.1 Trajectory analyses

3.2.2 Determination of clustering states and identification of intermolecular interaction regions

To detect the cluster formation in simulations, calculate the probabilities of the intermolecular interactions in the simulations of multiple monomers and plot the results, python scripts used in the study (89) were employed after slight modifications. To examine the behavior of the monomers relative to each other and identify the initial time steps of the clustering, COM of each monomer was measured by using the tcl script “dist_bw_chains.tcl” (Appendix 1) in two and four-monomer simulations throughout the trajectories. The interacting regions of the chains were investigated by measuring the number of atomic contacts between each possible residue pairs of two chains in a 4 Å cutoff by using “contacts_for_chains.tcl” (Appendix 2). The measurements were performed for each chain pair (chains A & B, A & C, A & D, B & C, B & D, C & D). For contact measurements of each chain pair, heat maps were generated. Loop regions L1, L2 and L3 were separated with horizontal lines and colored as cyan, green and red, respectively. L1, L2 and L3 loops of RodA were assumed as residues 71-98, 106-126 and 136-151, respectively; the remaining residues between the corresponding inter-cysteine regions were excluded since they are involved in the formation of the central β -barrel. The regions before L1 and after L3 were assumed as N- and C-termini to minimize the number of regions in the analyses.

To detect whether there was an interaction between L2 loops, contact maps were created by considering only the residues of these loops and given below the whole chain-chain contacts. All measurements were performed by employing VMD Tcl Console and graphics were generated via RStudio version 4.2 (RStudio 2022) (90).

3.2.3 Measurements of intermolecular interactions

Once the clustering behavior was observed (close COM-COM distances between chains, significant number of atomic contacts) in the trajectories, the cluster stabilities were investigated by checking whether the chain-chain contacts last more than 5 ns in the time course of the simulations. Subsequently, non-bonded interactions (ionic interactions, hydrophobic interactions, and hydrogen bonds) were measured to detect the individual contributions of these interactions to the cluster stability. Salt bridges between charged residues were measured by using Salt Bridges Plugin of VMD. The oxygen-nitrogen distance cutoff was set to 3.2 Å and transient ionic interactions between chains and intrachain interactions were neglected in the analyses; residue pairs with conserved atomic distance up to 4 Å in at least last 1 ns of the trajectory were considered as a salt bridge. The change in oxygen-nitrogen distances between residue pairs in the course of simulation time were plotted.

Hydrophobic residues tend to avoid interacting with water and approximate to each other to decrease the solvent accessible surface area of the protein and promote intra-chain and multimeric structure stability. In this study, the duration and regional information about hydrophobic interactions were obtained by using a novel method. The hydrophobic residues in RodA sequence were selected and atomic contacts were measured only between the hydrophobic residues of two monomers by running the script “hydrophobic_contacts.tcl” (Appendix 3) in VMD Tcl Console (the residues selected as hydrophobic in VMD: Alanine, Leucine, Isoleucine, Valine, Proline, Phenylalanine, Methionine and Tryptophan). It was assumed that there is a hydrophobic interaction between two residues if their atoms were within a 4 Å cutoff distance. Similar contact maps as in the full-length intermolecular interactions were generated for hydrophobic interactions.

The contribution of hydrogen bonds to the cluster stability was measured by using the script “hbonds.tcl” (Appendix 4). Donor-acceptor distance was set to 3.5 Å and the angle between donor-hydrogen-acceptor atoms was 30°. The number of hydrogen bonds formed between each chain pair was calculated throughout the trajectories of the two- and four-monomer simulations and plotted by using RStudio version 4.2 (RStudio 2022).

3.2.4 Monomer stability

The cluster stability depends on the individual stabilities of RodA monomers. RMSD and RMSF measurements give information about overall and per-residue protein stability. The Tcl scripts “RMSD_protein.tcl” (Appendix 5) and “RMSF_protein.tcl” (Appendix 6) were run in VMD Tcl Console to calculate overall and per-residue stability of the monomers in single, two- and four-monomer MD simulations. The measurements were performed by using the coordinates of the carbon alpha (CA) atoms only. The changes in the RMSD and RMSF measurements over simulation time were plotted by using RStudio version 4.2 (RStudio 2022); the data of the parallel runs were plotted together.

4 RESULTS

4.1 General Analyses

In earlier studies, amyloidogenic or aggregation-prone region predictions were performed only for RodA, EAS, and DewA by using Amylpred2 or TANGO (11,14,33). In this study, similar regions re-predicted with significant propensities as in these previous predictions. The non-loop region of RodA were also predicted as amyloidogenic. However, this non-loop region corresponded to a strand of the central β -barrel of RodA (Table 3).

The effect of the changes in protein and ion concentration, pH, and temperature were tested on the predictions performed by using TANGO. In the most runs, the predictions were not affected from these changes in the prediction conditions. In our results, similar regions with the previous studies were predicted as having a high tendency for β -aggregation, except RodA. There were no regions predicted to be prone to β -aggregation, except at temperature 283 K. The probability of being an aggregation-prone region increased to 7% for the region 63I-67T (N-terminal) in the predictions at 283 K (Table 3). Portions from each loop (L1, L2, and L3) were predicted as hits by Amylpred2 runs, except DewA; only the L2 loop was predicted as an amyloidogenic from the loops (Table 3).

New predictions were performed by using Amylpred2 and TANGO to validate the hydrophobin regions that were not predicted but were suggested to be involved in the self-assembly (L1 loop of SC3 and Vmh2) by using the findings obtained from experimental methods and MD simulations (27,34). Although our predictions are in line with the previous studies, there were additional regions with high probabilities outside of these previously predicted regions (Table 3).

In addition to aggregation propensity predictions, CPORT was employed to predict the residues that may have a role in hydrophobin self-assembly by considering the fact that aggregation-prone sequence regions are the residues present at the protein-

protein interfaces. Regions that were previously suggested as aggregation-prone or amyloidogenic were also predicted as interface residues in CPORT results, except the L3 loop of RodA and most of L1-L2 region in EAS (Table 3).

Table 3. Regions predicted as amyloidogenic and residues predicted as to be located at protein-protein interfaces.

Hydrophobin (Chain Sequence)	Uniprot ID	PDB ID	Prediction Tools				
			Amylpred2		TANGO		CPORT
			Previous Studies	Current Results	Previous Studies	Current Results	Current Results
RodA (19-159)	B0Y4B1	6GCJ	<u>L2</u> I112-I117 <u>L3</u> G143-G147	G100-Q104 <u>L2</u> I112-I117 <u>L3</u> G143-L148	-	-	V39-R40 P44- T48 K50-Q51 T53- E54 <u>L2</u> L108-K126
EAS (27-108)	Q04571	2FMC	-	<u>L1</u> L32-P36 <u>L1-L2</u> L43-Q53 <u>L3</u> N70-A78	<u>L1-L2</u> S42-G55 <u>L3</u> T68-A82	<u>L1-L2</u> G44-S52 <u>L3</u> F72-N79	<u>L1</u> S21-G27 P29-G44 <u>L3</u> G69-N70 I74, N76-A78
SC3 (25-136)	P16933	-	-	V11-T19 T24-A27 <u>L1</u> A53-S74 <u>L2</u> L76-G82 <u>L3</u> N99-T107	-	V11-T18 <u>L1</u> 51V-S63 V67-S72 <u>L2</u> L76-V81	-
DewA (19-135)	P52750	2LSH	-	V29-C35 K92-C98 <u>L2</u> D74, G77-V82	<u>L2</u> K69-V82	<u>L2</u> L78-V82	K92 <u>L1</u> E40, N42-L50 G52-L55 N57, G64 <u>L2</u> A70-S71 I73-G77 L79-A80 V82-H84
Vmh2 (I22-L111)	Q8WZI1	-	-	<u>L1</u> L24-G37 D40-V45 <u>L2</u> I50-G56 <u>L3</u> C67-G69 N73-G79	-	<u>L1</u> L24-I36 L41-V46 <u>L2</u> I50-V55 <u>L3</u> G74-I79	-

4.2 Single Monomer and Two-Monomer MD Simulations

There was no difference observed between bulk solution and interface monomers in terms of monomer stability (Figure 7). In both solution and interface simulations, loops of the monomers were significantly mobile, unlike central β -barrels, N- and C-termini. In each interface simulation run, monomers adsorbed to the interface equally through L2 (Run 2, Run 3) and L3 (Run 1, Run 3) loops (Figure 8). The fluctuations were the largest and lowest in L3 and L1 loops, respectively (Figure 9).

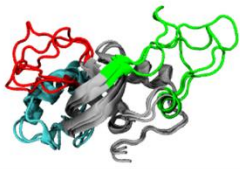
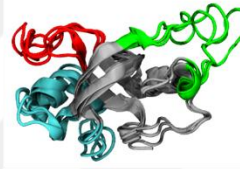
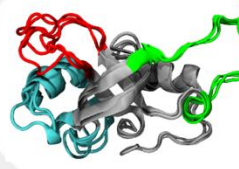
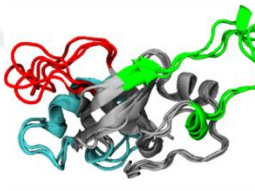
Solution			
Visualized Molecule(s)	Rank of Parallel Simulation		
	1	2	3
Protein Only			
Interface			
Protein Only			

Figure 7. Snapshots of conformations at three different time points in monomer solution and interface simulation systems. Snapshots were taken at the beginning, middle, and end of the trajectory and superimposed. L1, L2, and L3 loops and non-loop regions were denoted in cyan, green, red, and gray cartoon representations, respectively.

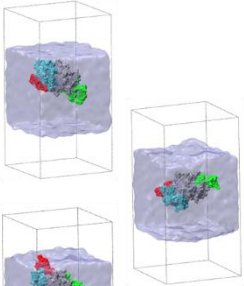
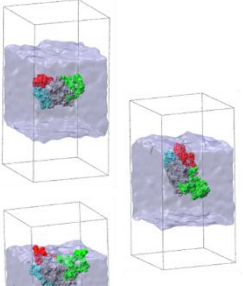
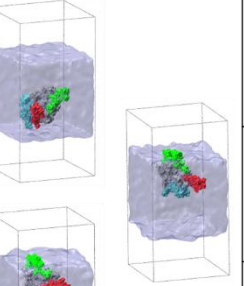
Interface				
Visualized Molecule(s)	Rank of Parallel Simulation			Time (ns)
	1	2	3	
Protein and Water				0
				25
				50

Figure 8. Snapshots of water boxes at three different time points in monomer interface simulation systems. Snapshots were taken at the beginning, middle, and end of the trajectory. L1, L2, and L3 loops and non-loop regions were denoted in cyan, green, red, and gray surface representations, respectively.

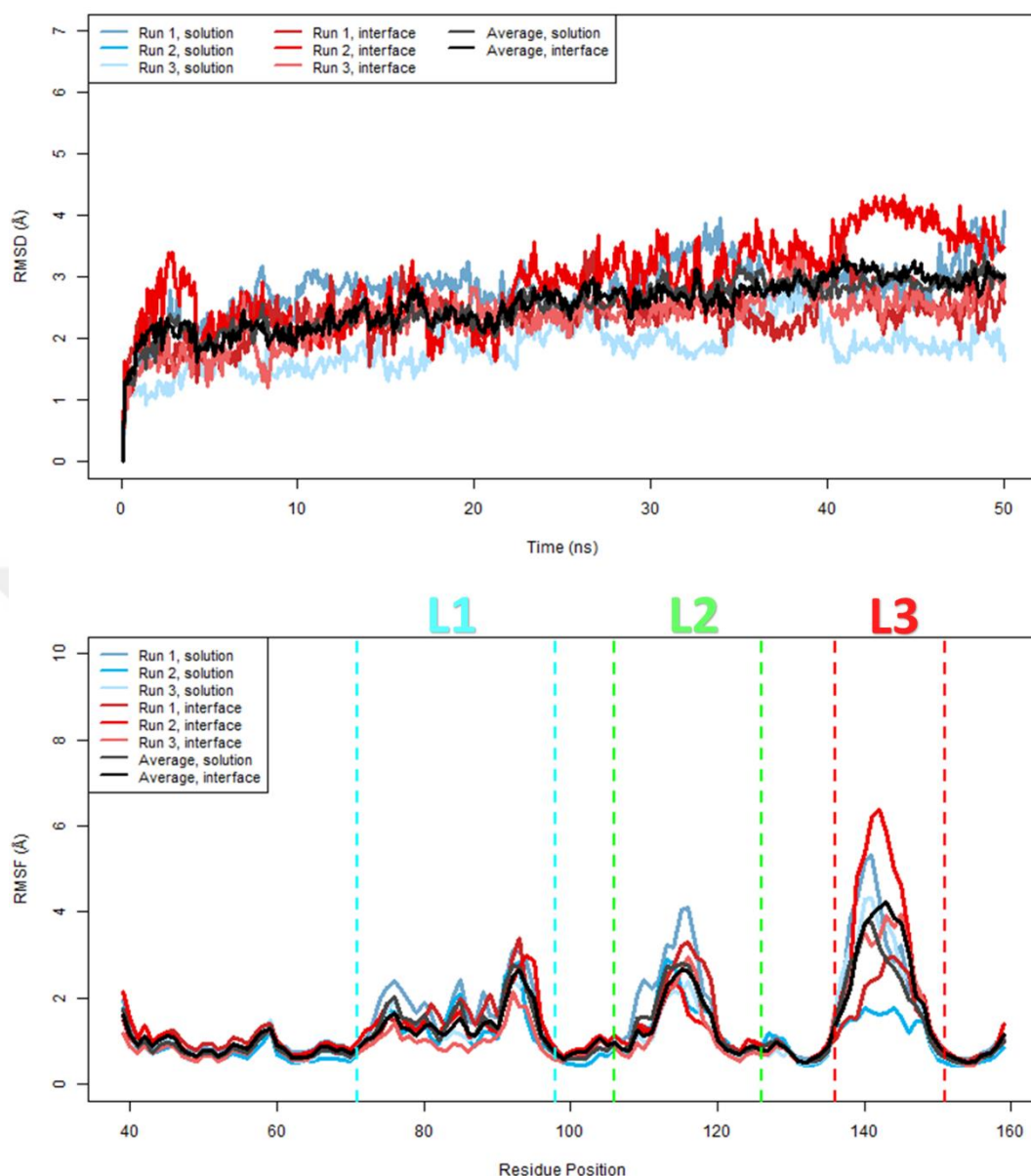


Figure 9. RMSD and RMSF measurements of single chains in monomer simulation systems in both environments. Blue and red color gradients (from light to dark) were used to differentiate parallel solution and interface simulations, respectively. Averages were depicted in gray and black lines respectively. Regions between dashed cyan, green, and red lines show L1, L2, and L3 loops, respectively.

The salt bridges measured were mostly transient in all monomers that were simulated in both environments, except E54-K50 which stabilized N-terminal α -helices. Despite the highest number of salt bridges measured in the monomers in two parallel solution runs, only five of them were common between all simulations and all L2-L2 salt bridges were observed in all parallel simulations (Figure 10). Two ionic interactions were measured at the N- and C-terminal of the L2 loop; the latter

contributed stability of the small α -helix. These two ends of the L2 loop were also transiently interacting via D109 and K126. In most simulations, two negatively charged residues at N- (D73) and C-terminal (E97) of the L1 loop were transiently interacting with R40 at the N-terminal of the monomer. The ionic interaction between K87 (L1) and D140 (L3) was rarely observed when all simulations were considered (Figure 10).

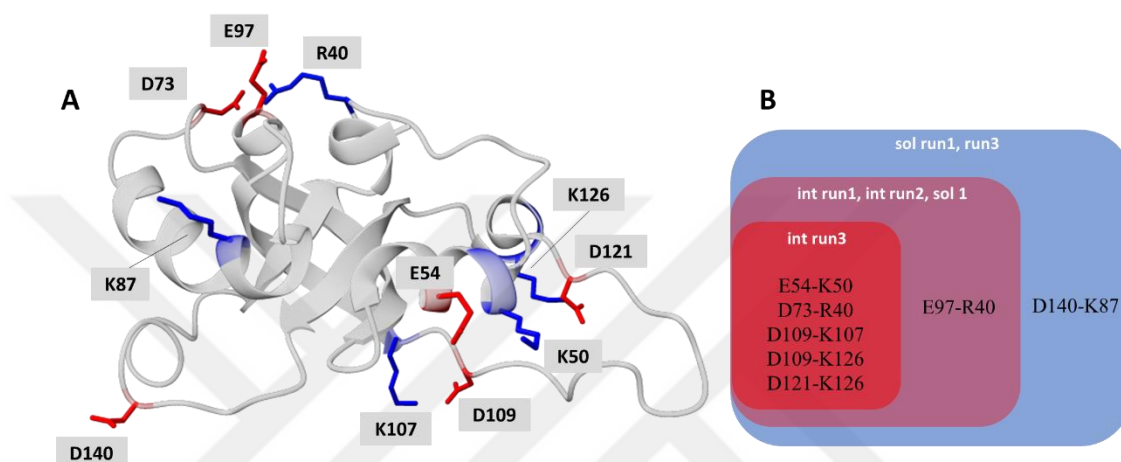


Figure 10. Salt bridges observed in the monomer simulations. A) Positions of the charged residues on the monomeric structure. B) Venn diagram of salt bridges drawn to check common residue pairings. Negatively and positively charged residues were depicted in red and blue sticks, respectively. The structure given was taken directly from NMR solution structure (PDB ID: 6GCJ).

In the solution simulations of the two-monomer systems with four different initial orientations, clusters of two were observed at least for 25 ns in the two parallel simulations of Orientation 1 and 4. However, only one cluster was observed in the simulations of Orientation 2 and 3 (Figure 11-12). The most stable COM-COM distance was measured between monomers of the first simulation run of Orientation 1. Although the most stable cluster with stable oligomer size for the longest simulation time was measured in the third run of Orientation 2, the distance between monomers was not stable compared to the previously mentioned cluster of Orientation 1. The distances between monomers within other conserved clusters either fluctuated around or decreased below 40 Å. However, small fluctuations were observed (Figure 13). The intermolecular interactions with the highest interaction probabilities mostly involved L2 and L3 loops (data not shown). There were no conserved interactions between

monomers established through C-terminals. Intermolecular interaction regions mostly covered more than two regions of a monomer. For example, L3 of chain B were interacting with both N terminal and L2 of chain A in the conserved cluster of the first parallel run of Orientation 4 (Figure 14). The highest and most conserved number of hydrogen bonds (≤ 5) was observed in the first run of Orientation 1. On the other hand, the contribution of the hydrogen bonds to the stability was the lowest in the cluster with the longest stable oligomer size (in the last run of Orientation 2) compared to the other conserved clusters (Figure 15). Conserved hydrophobic intermolecular interactions were mostly detected in heterogenous loop contacts except for the cluster of the first run of Orientation 1 simulations; there were no conserved hydrophobic interactions (Figure 16). On the other hand, only one conserved ionic interaction was observed only in this cluster (data not shown).

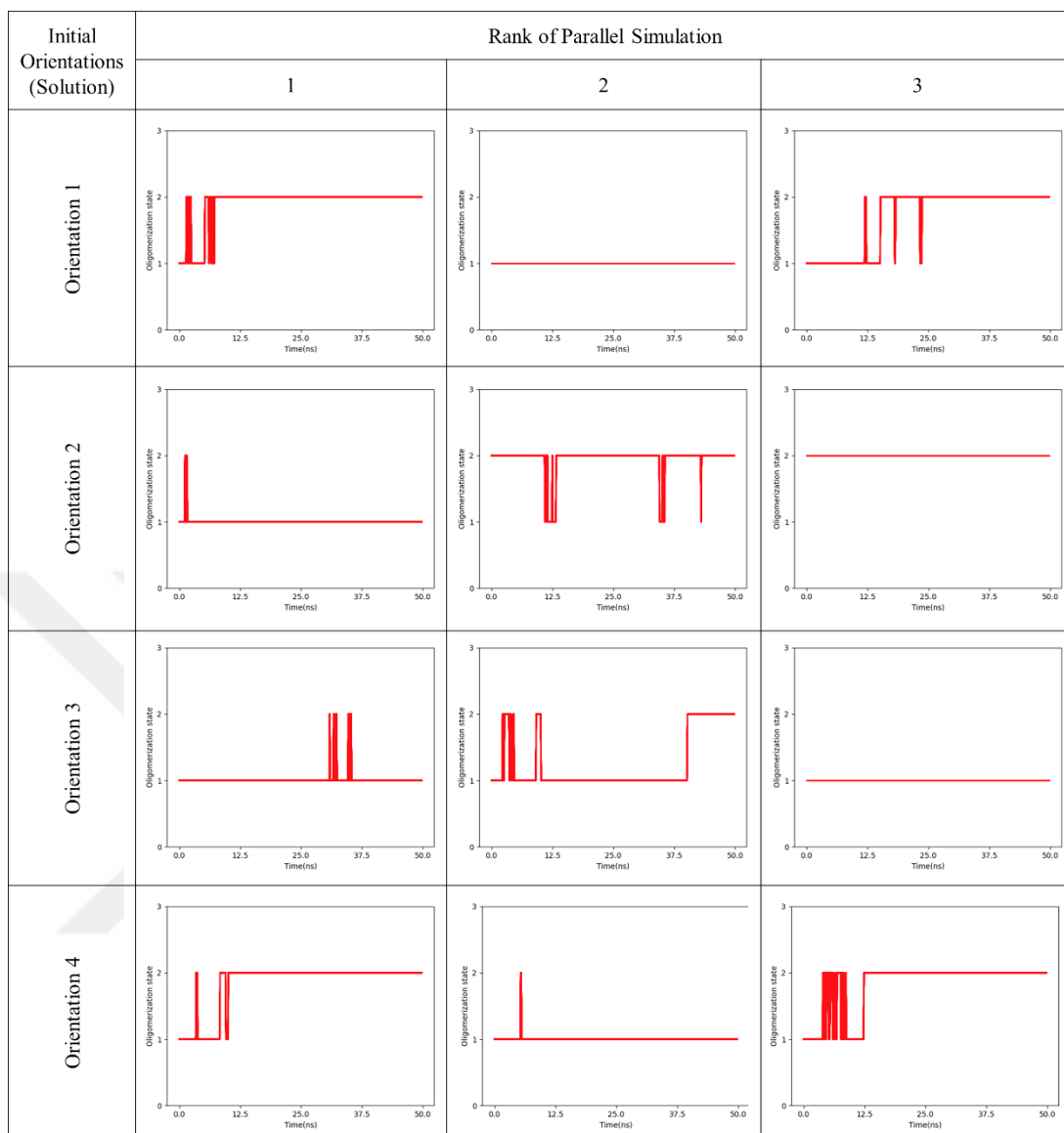


Figure 11. Cluster analysis of solution simulations performed with two monomers. Monomers in a 4 Å distance cutoff assumed as clustered in the simulation systems and the cluster lifetime was measured throughout the trajectories.

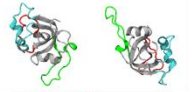
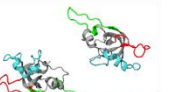
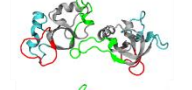
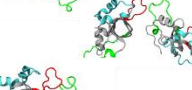
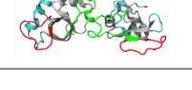
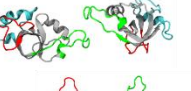
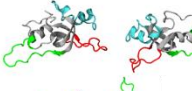
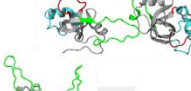
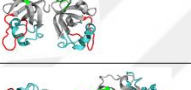
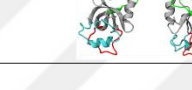
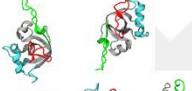
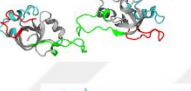
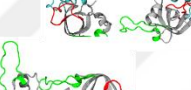
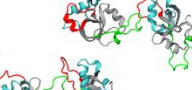
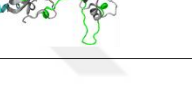
Rank of Parallel Simulation	Initial Orientations (Solution)				Time (ns)
	Orientation 1	Orientation 2	Orientation 3	Orientation 4	
1		No contacts observed.	No contacts observed.		0
					25
					50
2	No contacts observed.			No contacts observed.	0
					25
					50
3			No contacts observed.		0
					25
					50

Figure 12. Snapshots of conformations at three different time points in solution simulations performed with two monomers. Snapshots were taken at the beginning, middle, and end of the trajectory. L1, L2, and L3 loops and non-loop regions were denoted in cyan, green, red, and gray cartoon representations, respectively.

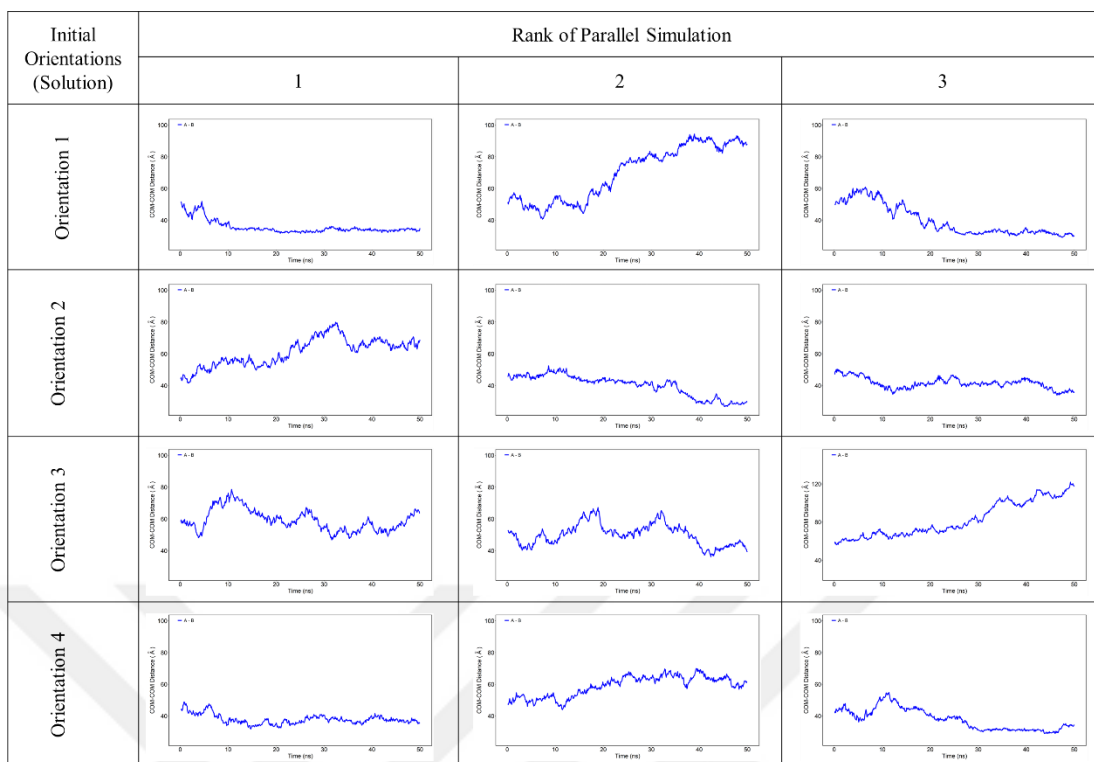


Figure 13. The COM-COM distances between chains in solution simulations performed with two monomers. The distances were measured throughout the trajectories for each parallel run.

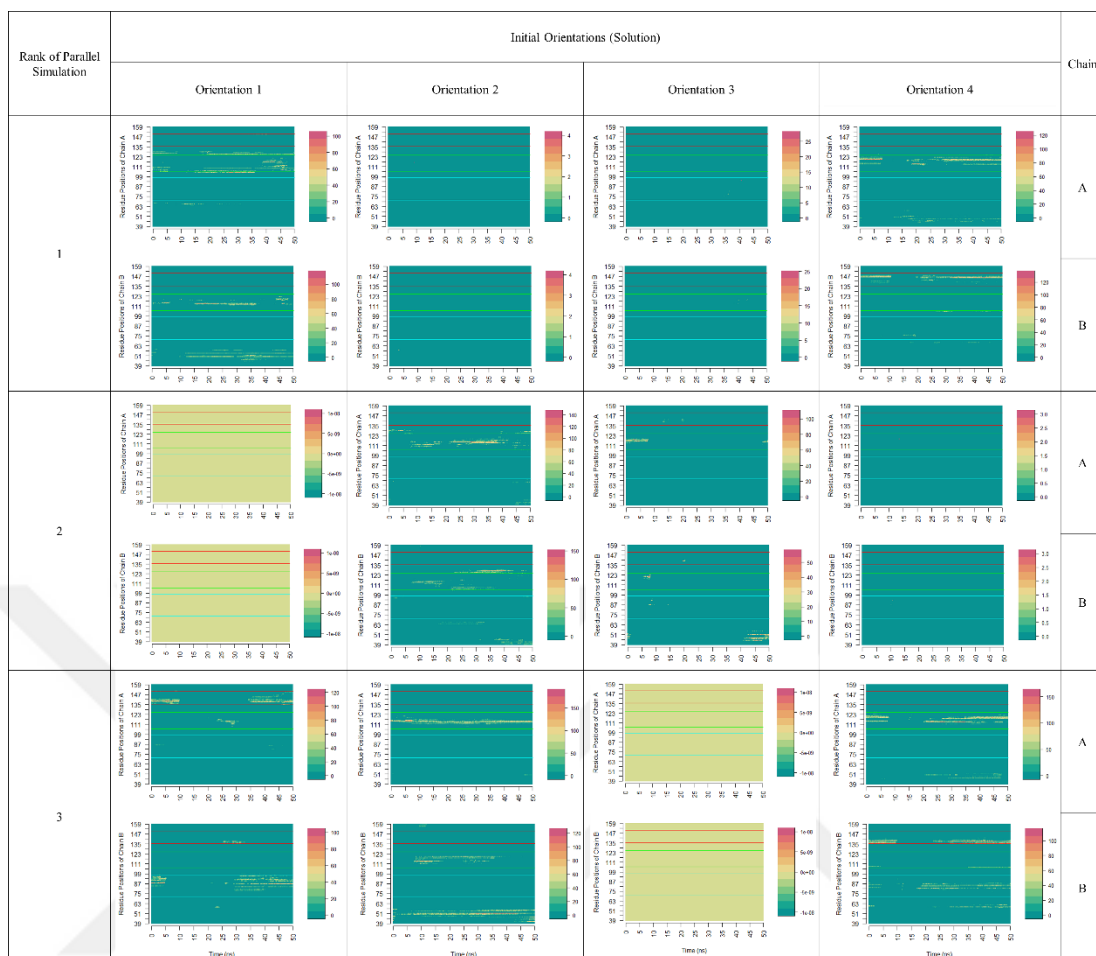


Figure 14. Contact maps of two-monomer solution and interface simulations. Each map shows the interaction of the chain with the other chain in the simulation box. L1, L2, and L3 loops were shown as cyan, green, and red horizontal lines on the maps, respectively. Light yellow background on the maps indicates no atomic interactions measured. The number of atomic contacts was assigned to a color gradient from dark green to red. The distance cutoff was set to 4 Å.

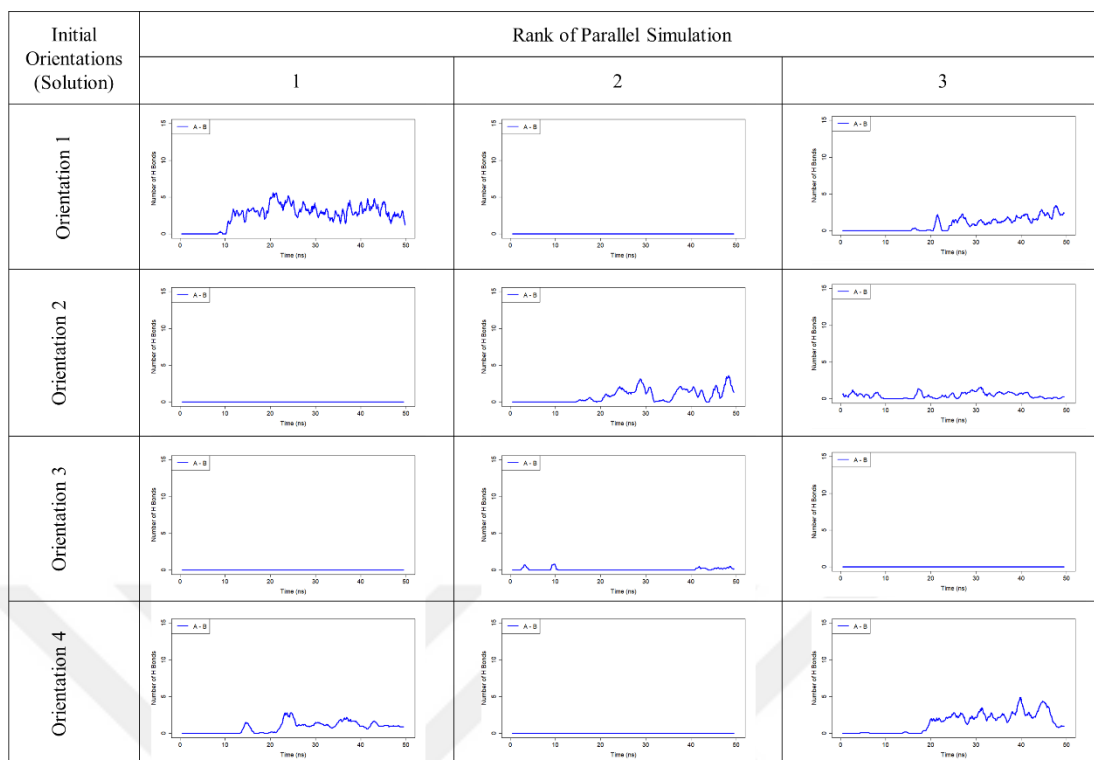


Figure 15. Measurements of the number of hydrogen bonds between chain pairs in solution simulations performed with two monomers. Donor-acceptor distance and angle cutoff were set at 3.5 Å and 30° respectively.

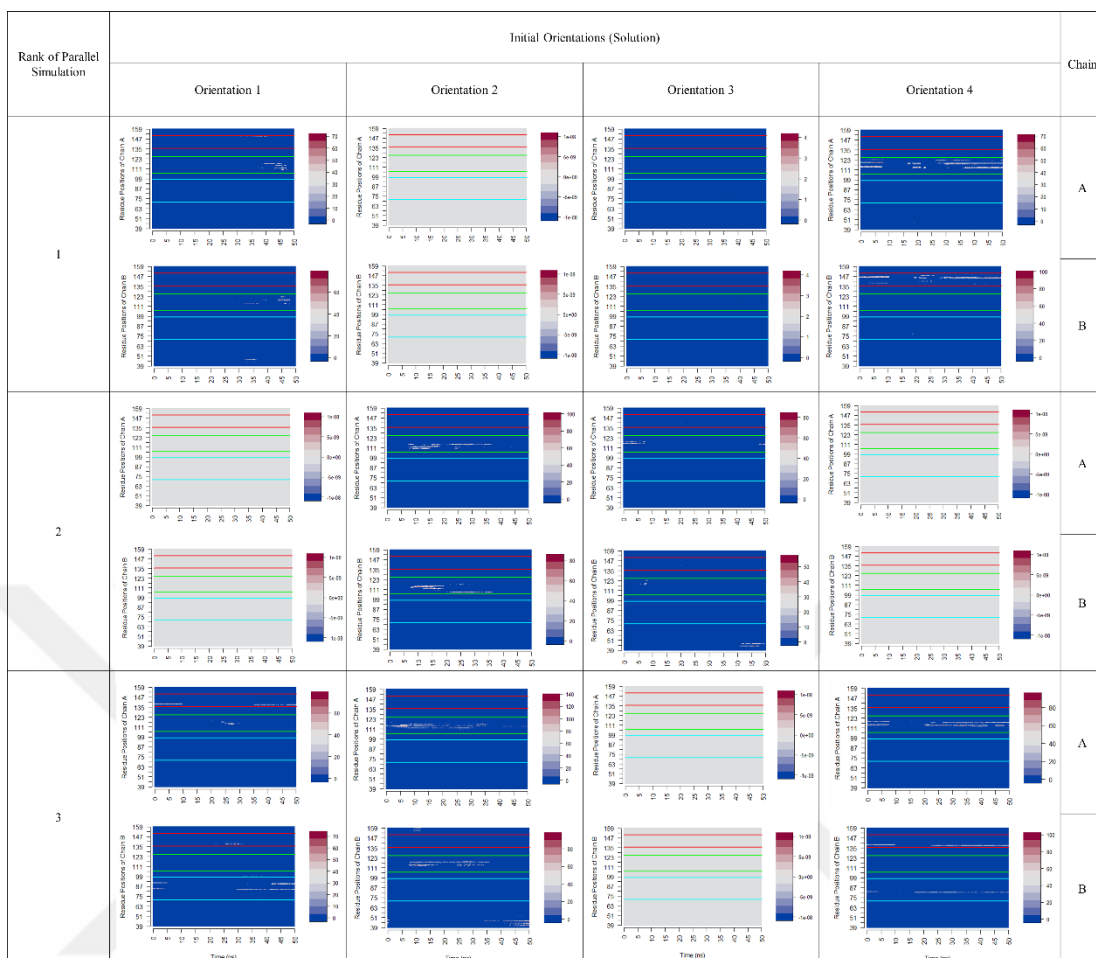


Figure 16. Hydrophobic contact maps of two-monomer solution simulations. Each map shows the interaction of the chain with the other chain in the simulation box. L1, L2, and L3 loops were shown as cyan, green, and red horizontal lines on the maps, respectively. Light gray background on the maps indicates no atomic interactions measured. The number of atomic contacts was assigned to a color gradient from dark blue to red. The distance cutoff was set to 4 Å.

At the interface simulations conducted with two RodA monomers, clusters were conserved mostly in Orientation 1 and Orientation 4 as in solution simulations. In Orientation 1 and 4 simulation trajectories, two clusters were observed for at least 20 ns in 2/3 of the parallel runs (Figure 17-18). In the conserved clusters, the maximum COM-COM distance was measured as 40 Å throughout the trajectory and the minimum distance was 30 Å between monomers in the cluster formed in the first run of Orientation 4 simulations (Figure 19). Monomers mostly interacted through one or more residues from various monomer regions and N-termini, L1, and L3 loops were highly involved in conserved intermolecular interactions. Although the variety of the interaction regions was high, few residues from the same region for a monomer were

measured as contacting (≤ 5) (Figure 20). On the other hand, L2 loops were mostly interacting with air at the air/water interface including the conserved clusters (Figure 18).

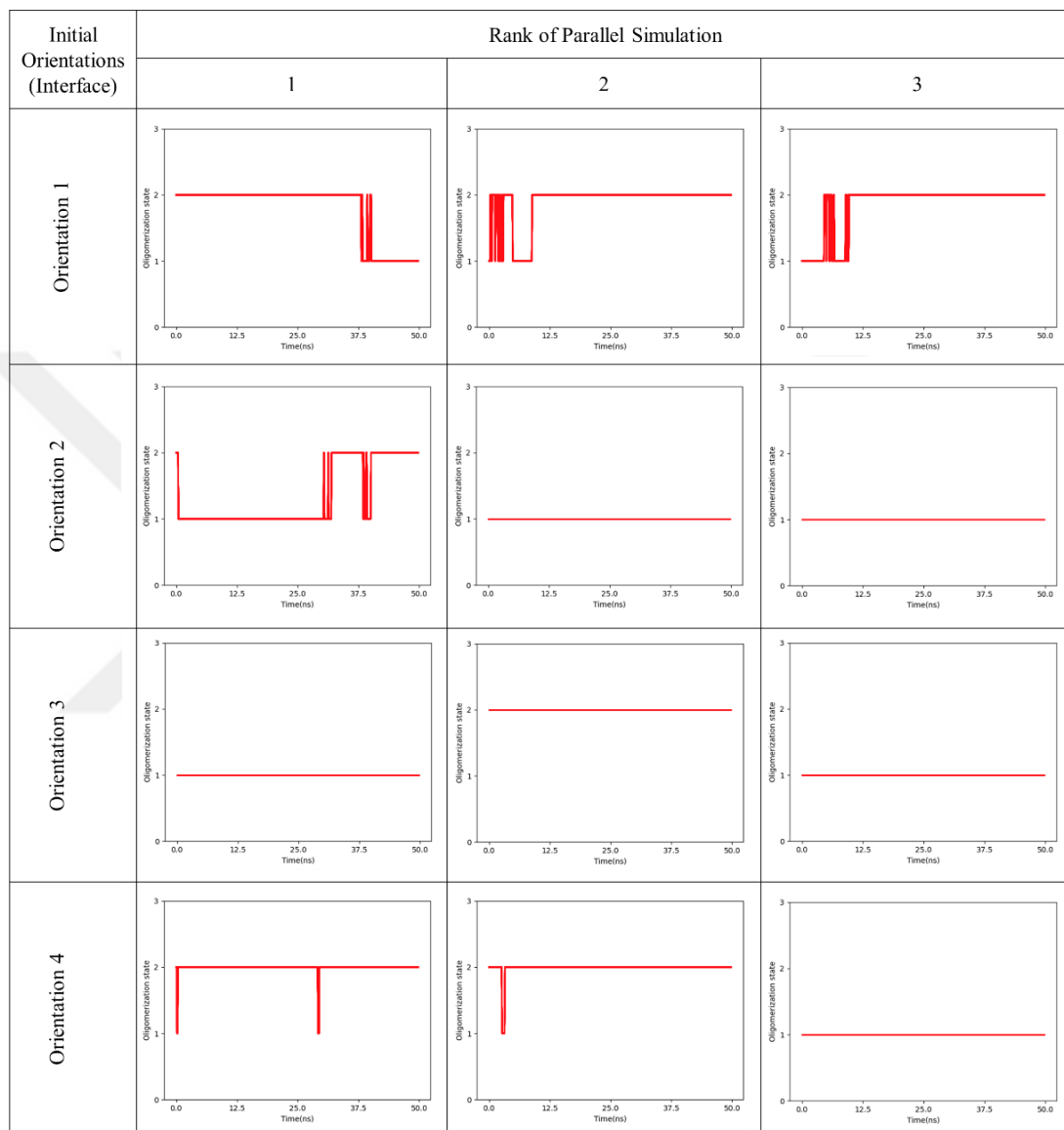


Figure 17. Cluster analysis of two-monomer interface simulations. Monomers in a 4 Å distance cutoff assumed as clustered in the simulation systems and the cluster lifetime was measured throughout the trajectories.

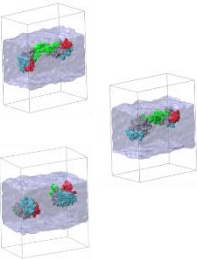
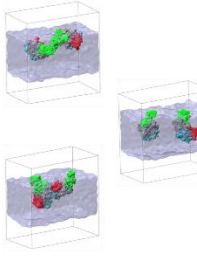
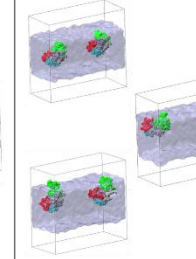
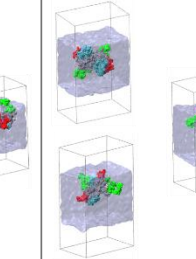
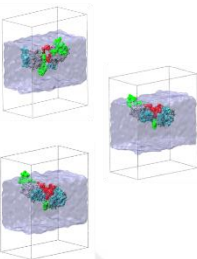
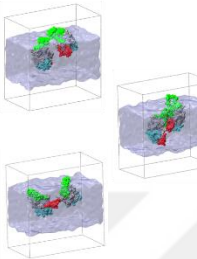
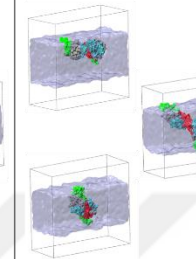
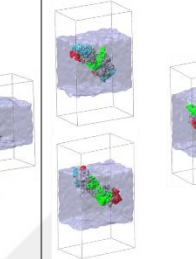
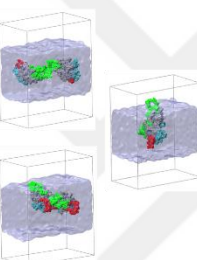
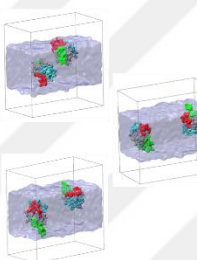
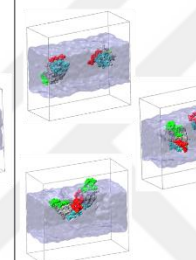
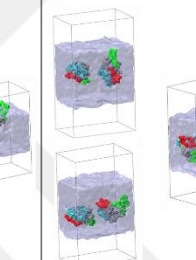
Rank of Parallel Simulation	Initial Orientations (Interface)				Time (ns)
	Orientation 1	Orientation 2	Orientation 3	Orientation 4	
1					0
					25
					50
2					0
					25
					50
3					0
					25
					50

Figure 18. Snapshots of conformations at three different time points in interface simulations performed with two monomers. Snapshots were taken at the beginning, middle, and end of the trajectory. L1, L2, and L3 loops and non-loop regions were denoted in cyan, green, red, and gray surface representations, respectively.

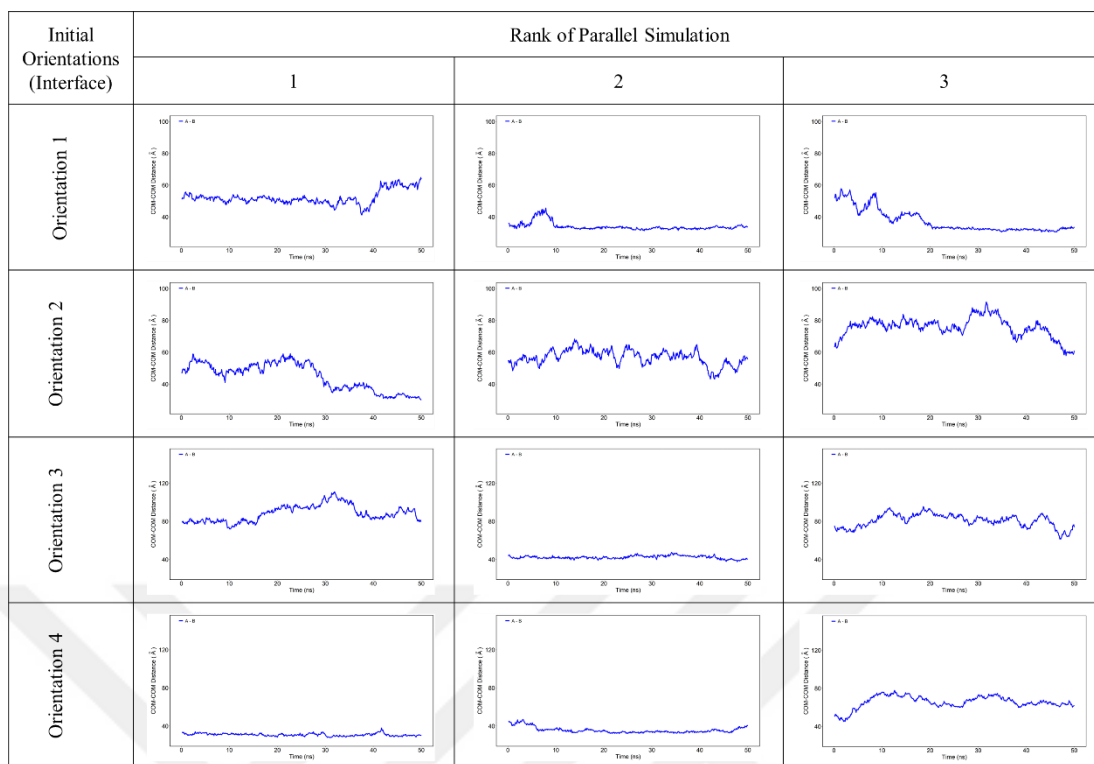


Figure 19. The COM-COM distances between chains in interface simulations performed with two monomers. The distances were measured throughout the trajectories for each parallel run.

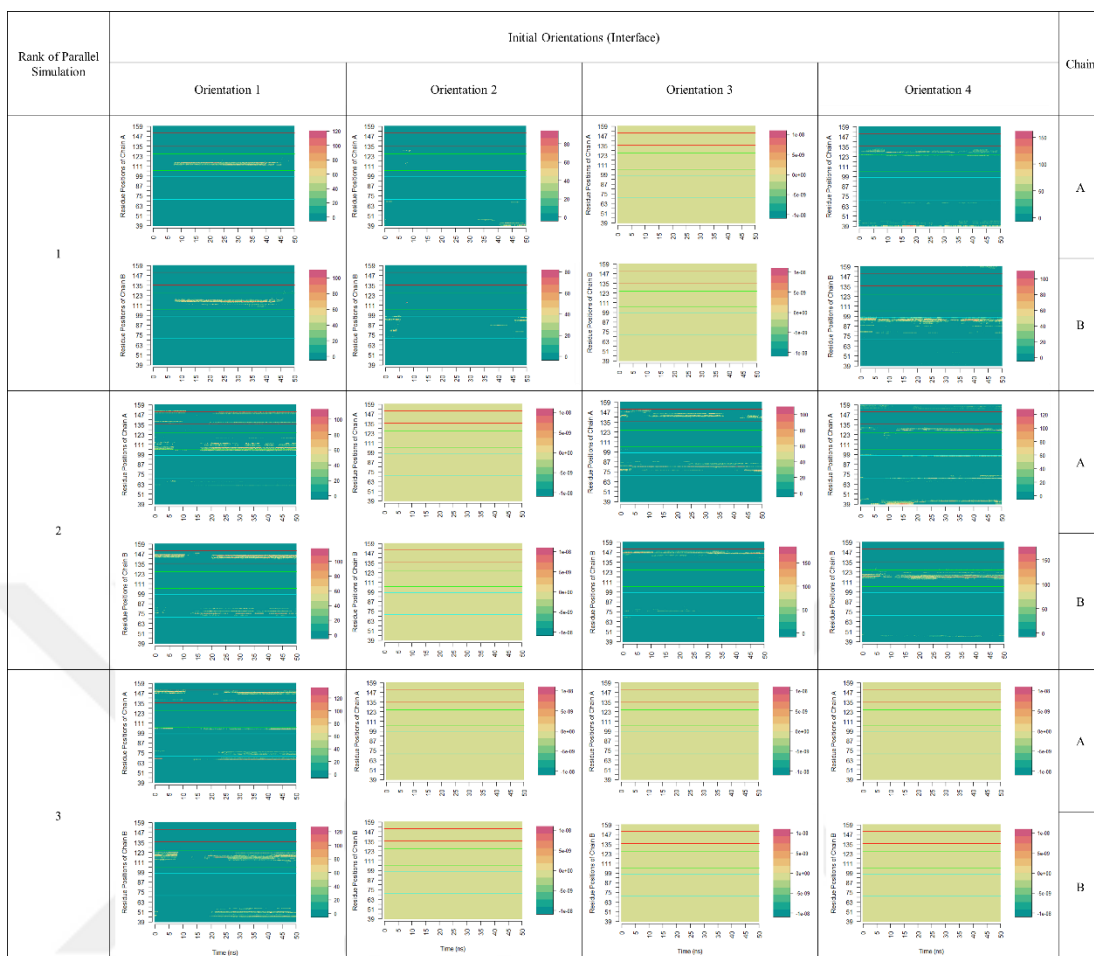


Figure 20. Contact maps of two-monomer solution and interface simulations. Each map shows the interaction of the chain with the other chain in the simulation box. L1, L2, and L3 loops were shown as cyan, green, and red horizontal lines on the maps, respectively. Light yellow background on the maps indicates no atomic interactions measured. The number of atomic contacts was assigned to a color gradient from dark green to red. The distance cutoff was set to 4 Å.

Conserved hydrogen bonds were only measured in the last two simulation runs of Orientation 1; others were fluctuating throughout the simulation time (Figure 21). Hydrophobic interactions contributed to the stability of only three clusters in the simulations of Orientation 1, 3, and 4 in homo- and heterogenous loop and N-terminal-loop interactions (Figure 22). In total, only five conserved salt bridges were observed in the third and the first run of Orientation 1 and 4 simulations, respectively. Most ion pairs were formed between N-termini and L1 loops of the monomers (data not shown). There was no difference between monomers of different orientation simulations in terms of overall structure stability or loop flexibility. Only two monomers in

Orientation 1 solution monomers were highly mobile (higher RMSD) compared to others (both in solution and at the interface simulations) due to the larger fluctuations in L3 loops (Figure 23).

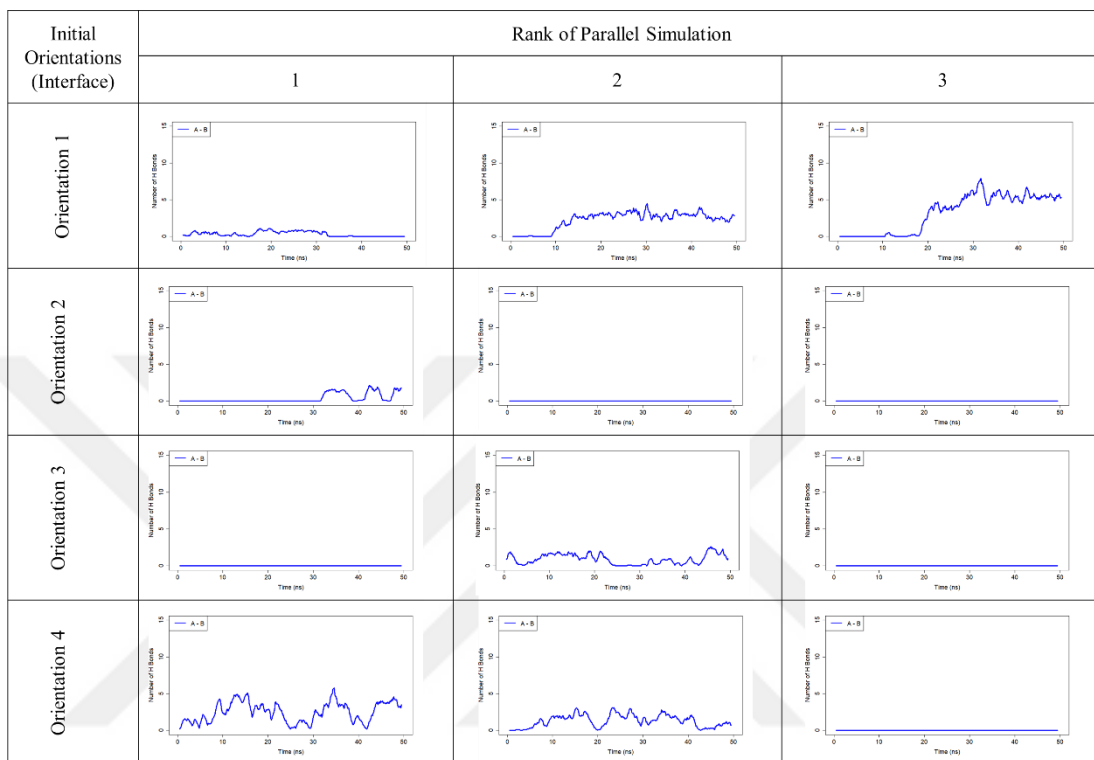


Figure 21. Measurements of the number of hydrogen bonds between chain pairs in solution simulations performed with two monomers. Donor-acceptor distance and angle cutoff were set at 3.5 Å and 30° respectively.

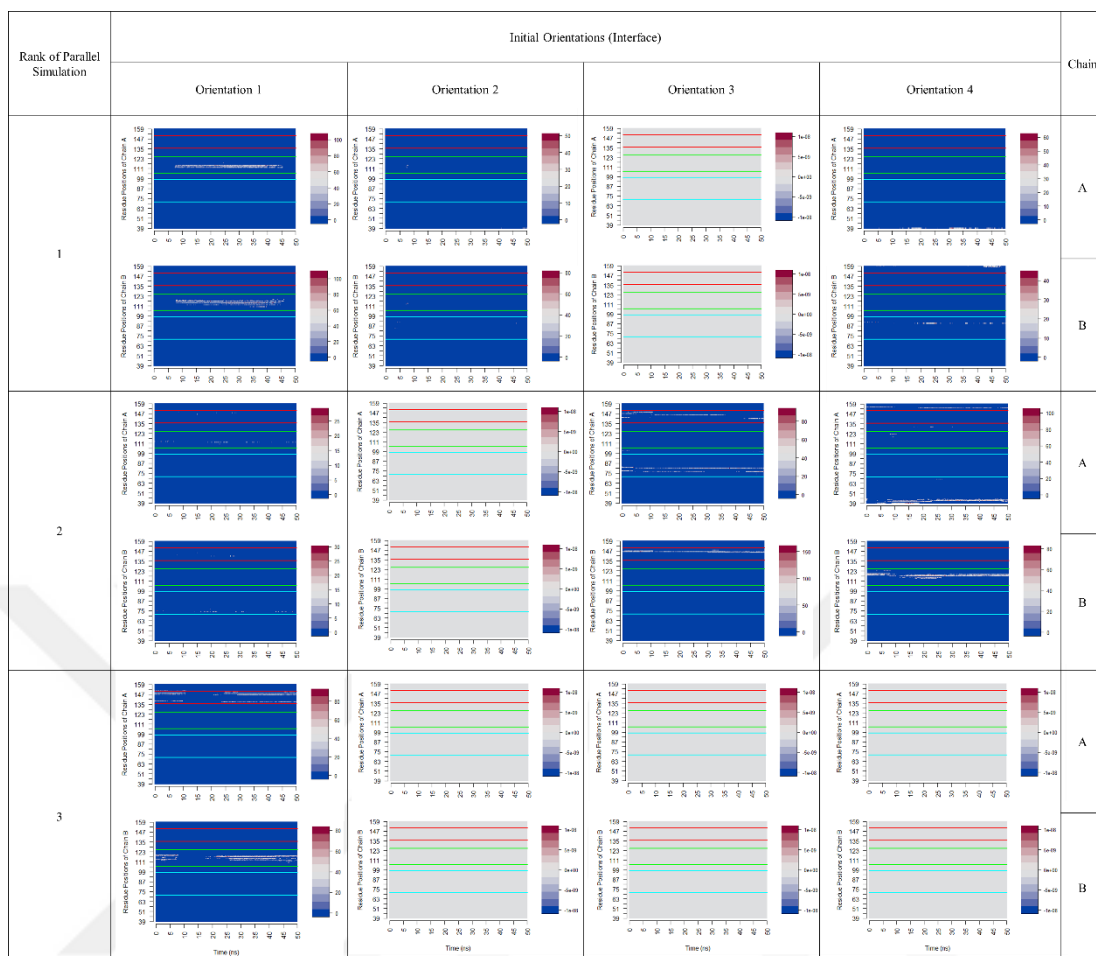


Figure 22. Hydrophobic contact maps of two-monomer interface simulations. Each map shows the interaction of the chain with the other chain in the simulation box. L1, L2, and L3 loops were shown as cyan, green, and red horizontal lines on the maps, respectively. Light gray background on the maps indicates no atomic interactions measured. The number of atomic contacts was assigned to a color gradient from dark blue to red. The distance cutoff was set to 4 Å.

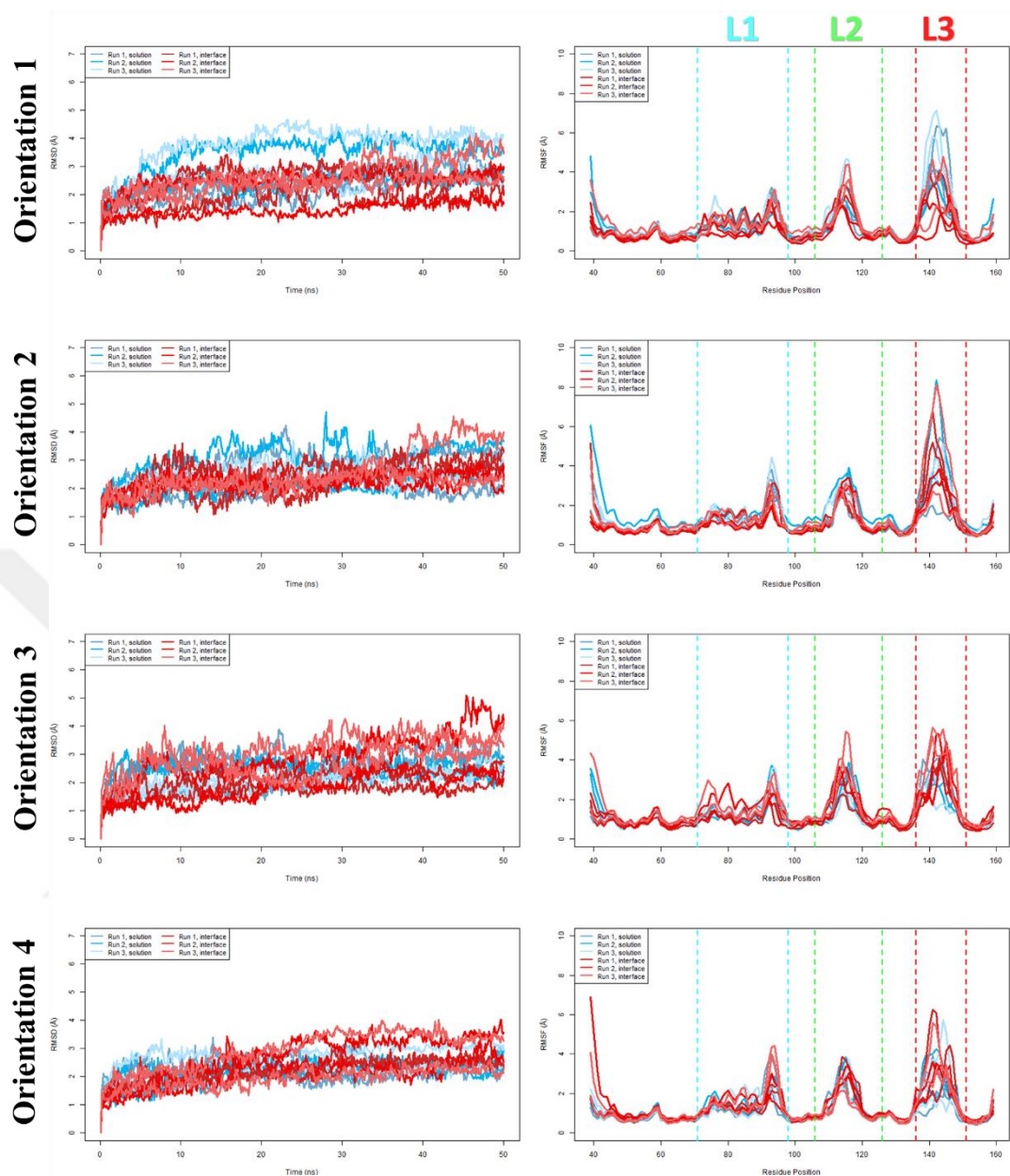


Figure 23. RMSD (left) and RMSF (right) measurements of individual chains in simulations performed with two monomers in both environments. Blue and red color gradients (from light to dark) were used to differentiate parallel solution and interface simulations, respectively. Regions between dashed cyan, green, and red lines show L1, L2, and L3 loops, respectively.

4.3 Four-Monomer MD Simulations

4.3.1 Simulation systems with positionally unrestrained water molecules

4.3.1.1 Orientation 1

Varied sizes of RodA clusters were observed at different time points both in Orientation 1 solution and at interface simulations. There was no common clustering behavior in clustered RodA monomers in terms of cluster form and interacting sites. Clusters of at least two monomers were observed in five simulations (Figure 24). The majority of the clusters were formed by three monomers in both environments, although long-term transient interactions of four monomers were observed in several simulation systems. Two isolated clusters (clusters of two and three) were observed for the last 30 ns of the second and third interface simulations (Figure 25). Chains A, C, and D formed an L-shaped cluster conserved in the late time steps of the trajectory in the first solution simulation run. Chains A and C were stable in the cluster with ~40 Å COM–COM distance during the simulation time (Figure 25). After the first 10 ns, chains A, B, and D clustered in the second parallel run with a relatively compact shape. In the last solution run of Orientation 1, there was no stable cluster, and the monomers became distant from each other in the course of simulation time (Figure 25).

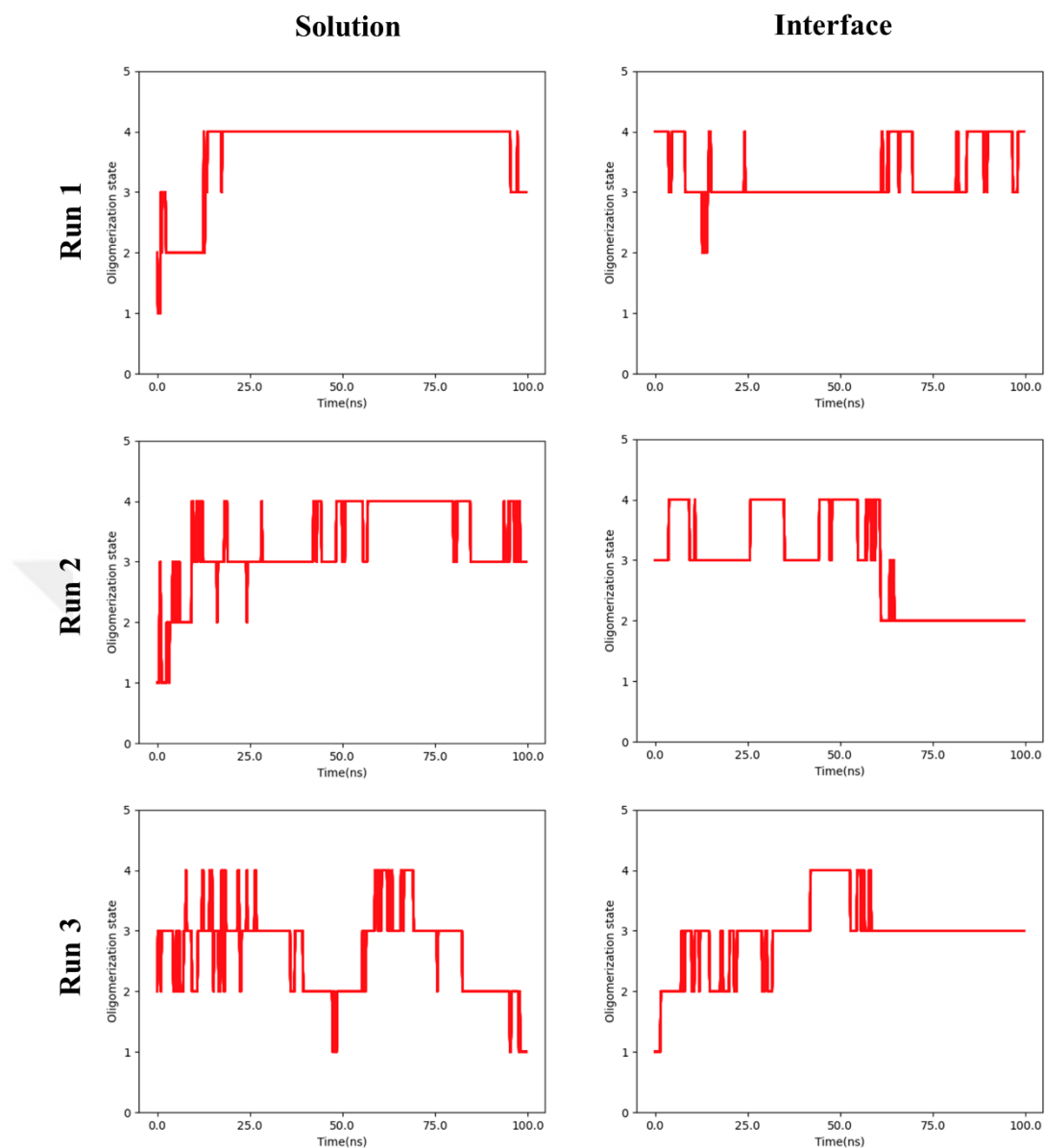


Figure 24. Cluster analysis of Orientation 1 simulations. Monomers in a 4 Å distance cutoff assumed as clustered in the simulation systems and the cluster lifetime was measured throughout the trajectories for both simulation environments.

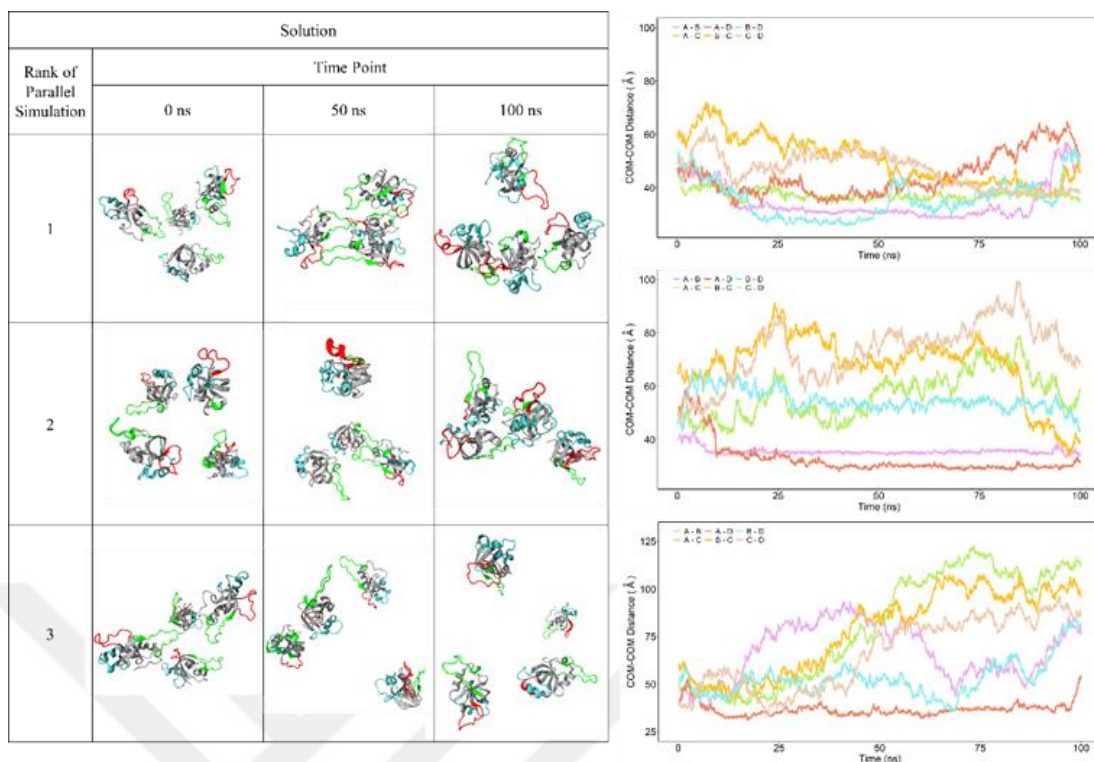


Figure 25. Snapshots of conformations at three different time points and COM-COM distances between chains of Orientation 1 solution simulations. Snapshots were taken at the beginning, middle, and end of the trajectory. L1, L2, and L3 loops and non-loop regions were denoted in cyan, green, red, and gray cartoon representations, respectively. The COM-COM distances between chains A & B, A & C, A & D, B & C, B & D, and C & D were shown in purple, green, orange, yellow, cyan, and beige lines, respectively.

At the air/water interface, clusters in varied shapes formed in the early steps with two or three monomers. Chain-chain distances were stable for 50 ns at least for a chain pair. In the globular RodA cluster observed in the first interface run, the A-B and A-D distances were constantly fluctuating around 30 Å and 35 Å, respectively. In the second run, the C-D chain distance was stabilized at around 40 ns and side-to-end interaction was observed. Clustering of chains B and D was through their loop regions in the early time points of the third interface run and the number of monomers which formed a conserved cluster increased to three after 25 ns and a planar-shaped cluster which formed with the largest distance between chains was around 55 Å. Air contact was observed in L2 and/or L3 loops of isolated monomers and only L2 loops of a monomer in the clusters (Figure 26).

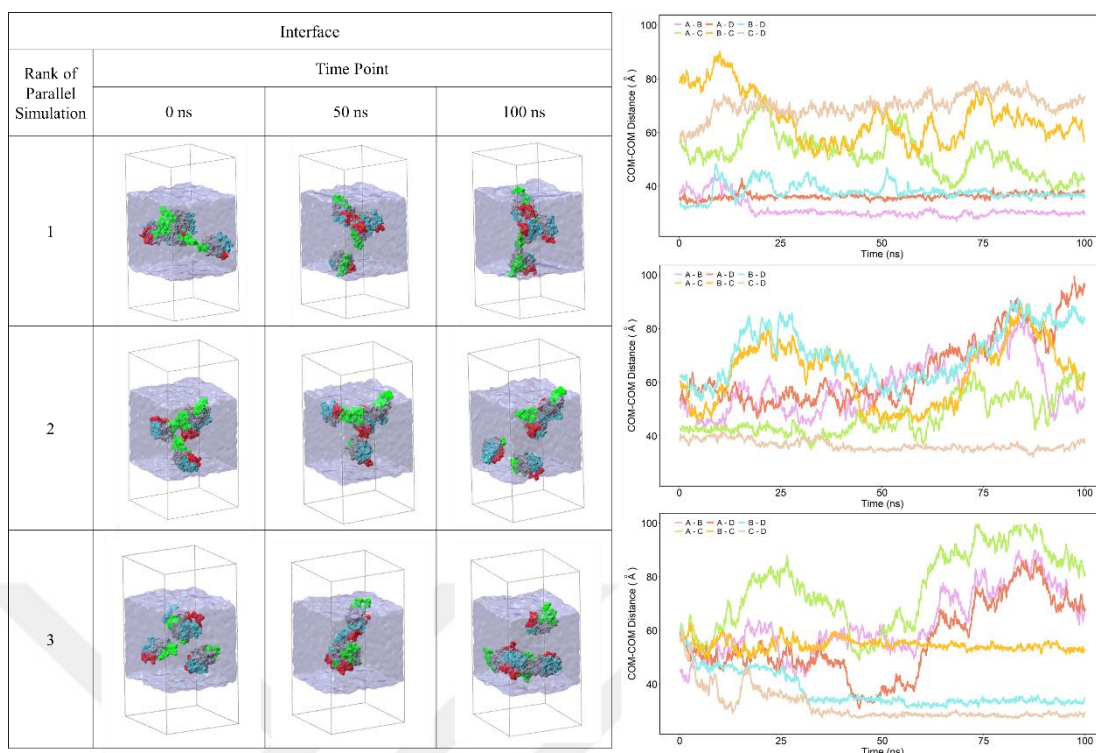


Figure 26. Snapshots of conformations at three different time points and COM-COM distances between chains of Orientation 1 interface simulations. Snapshots were taken at the beginning, middle, and end of the trajectory. L1, L2, and L3 loops and non-loop regions were denoted in cyan, green, red, and gray surface representations, respectively. The COM-COM distances between chains A and B, A and C, A and D, B and C, B and D, and C and D were shown in purple, green, orange, yellow, cyan, and beige lines, respectively.

The number of chain–chain atomic contacts was measured for each chain. In both simulation environments, both N-termini and L2 regions played a key role in the inter-chain interactions in the clusters for significant time periods (at least 5 ns) from a time point to the end of the trajectories. The number of conserved atomic contacts was higher in L2 loops and L2-L3 region compared to other interacting regions in solution and at interface simulations, respectively. N-termini was interacting with relatively high probabilities during the simulations by contacting either N-termini or L2 loops of other chains in solution simulations (Figure 27). On the other hand, there were no dominantly conserved pairs of interacting regions at the interface, except L2-L2 with high interaction probability in the last parallel run (Appendix 7). The size of the

regions involved in interactions was larger at the interface simulations although the variety was similar to the solution (Figure 28).

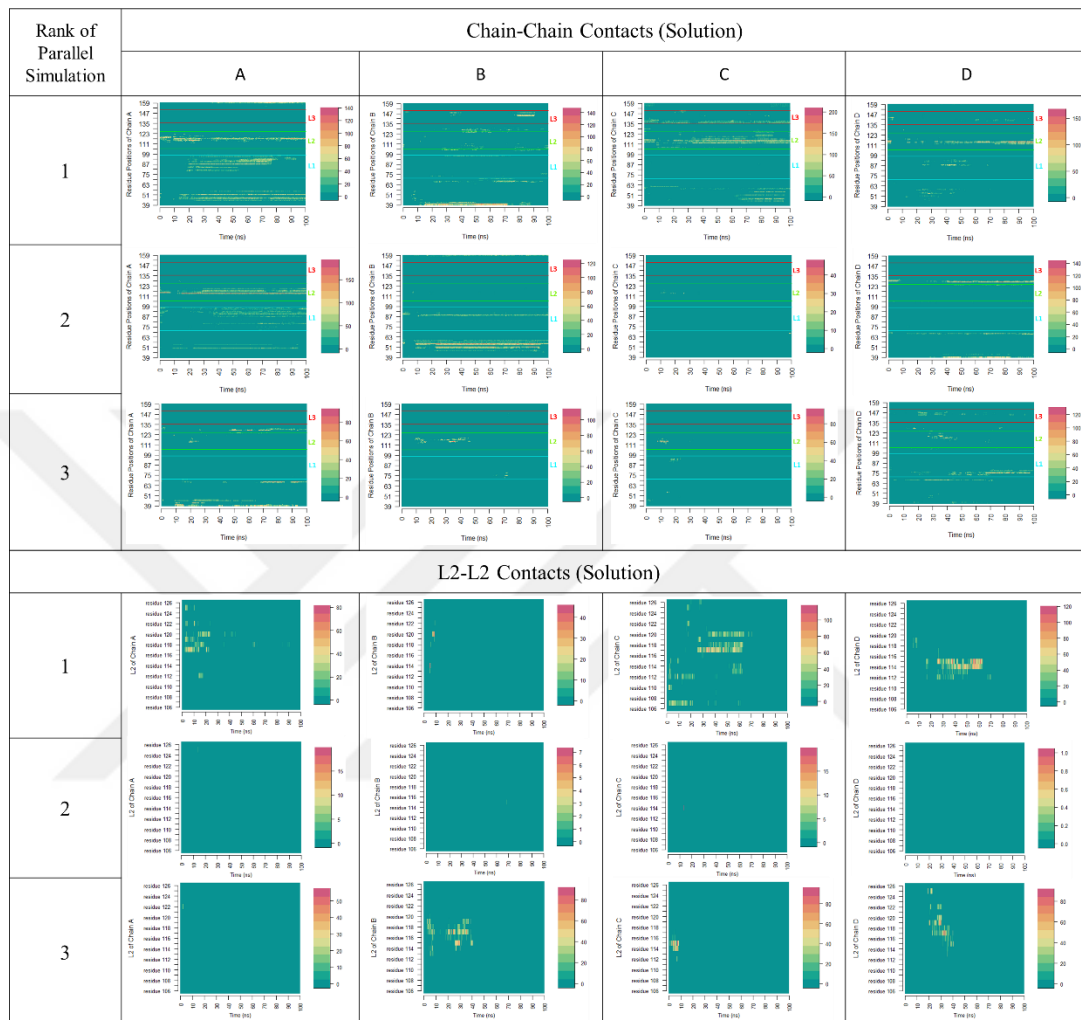


Figure 27. Contact maps of Orientation 1 solution simulations. Each map shows the interaction of the chain column with the remaining chains of the simulation system. L1, L2, and L3 loops were highlighted on the maps. L2-L2 interactions were measured by selecting only L2 loops. Light yellow background on the maps indicates no atomic interactions measured. The number of atomic contacts was assigned to a color gradient from dark green to red. The distance cutoff was set to 4 Å.

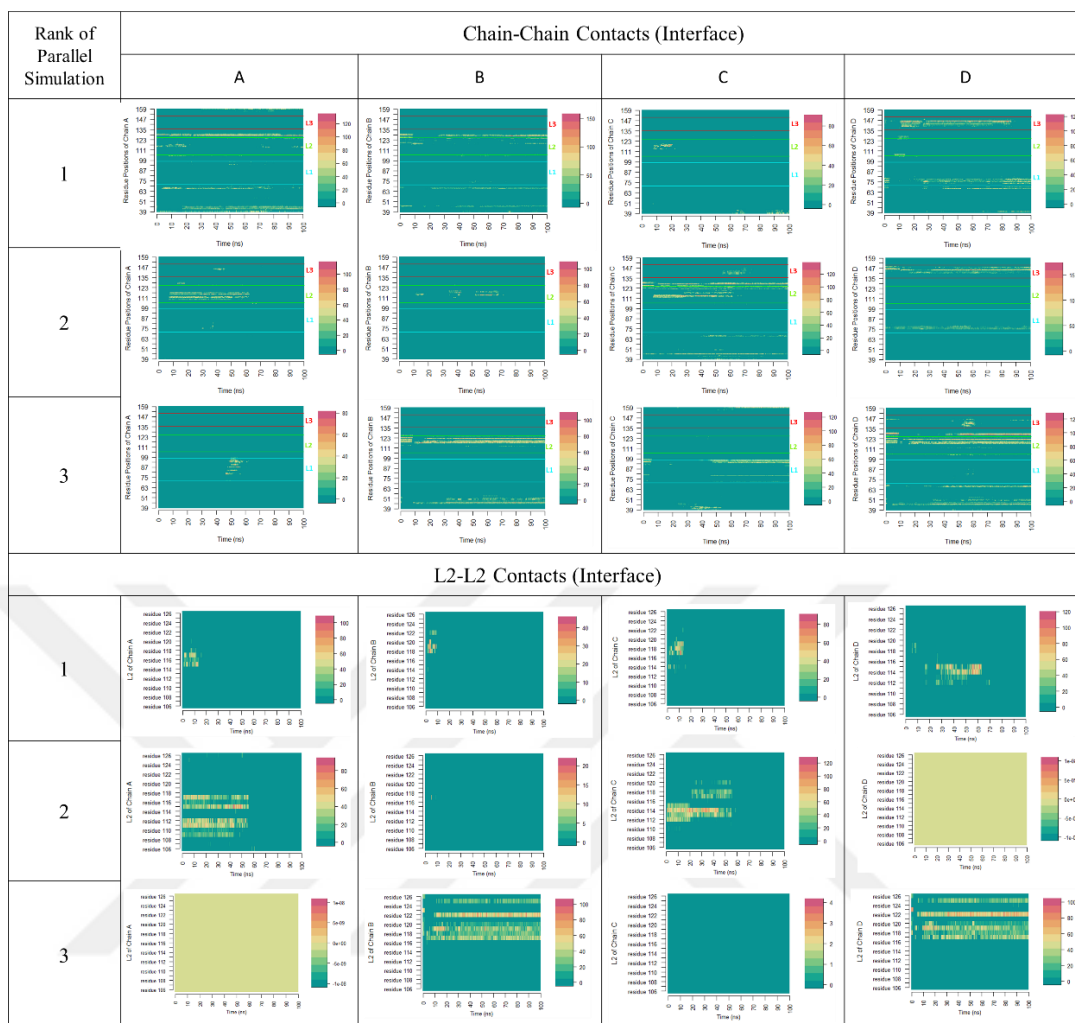


Figure 28. Contact maps of Orientation 1 interface simulations. Each map shows the interaction of the chain column with the remaining chains of the simulation system. L1, L2, and L3 loops were highlighted on the maps. L2-L2 interactions were measured by selecting only L2 loops. Light yellow background on the maps indicates no atomic interactions measured. The number of atomic contacts was assigned to a color gradient from dark green to red. The distance cutoff was set to 4 Å.

Conserved hydrophobic interactions were detected in most regions in solution and at interface simulations; hydrophobic regions of L2 loops were higher in number in the monomer interactions in both environments and C-termini of monomers in solution simulations. It was observed that the number of interacting atoms was higher, and the duration of the interactions was longer at the interface clusters (Figure 29).

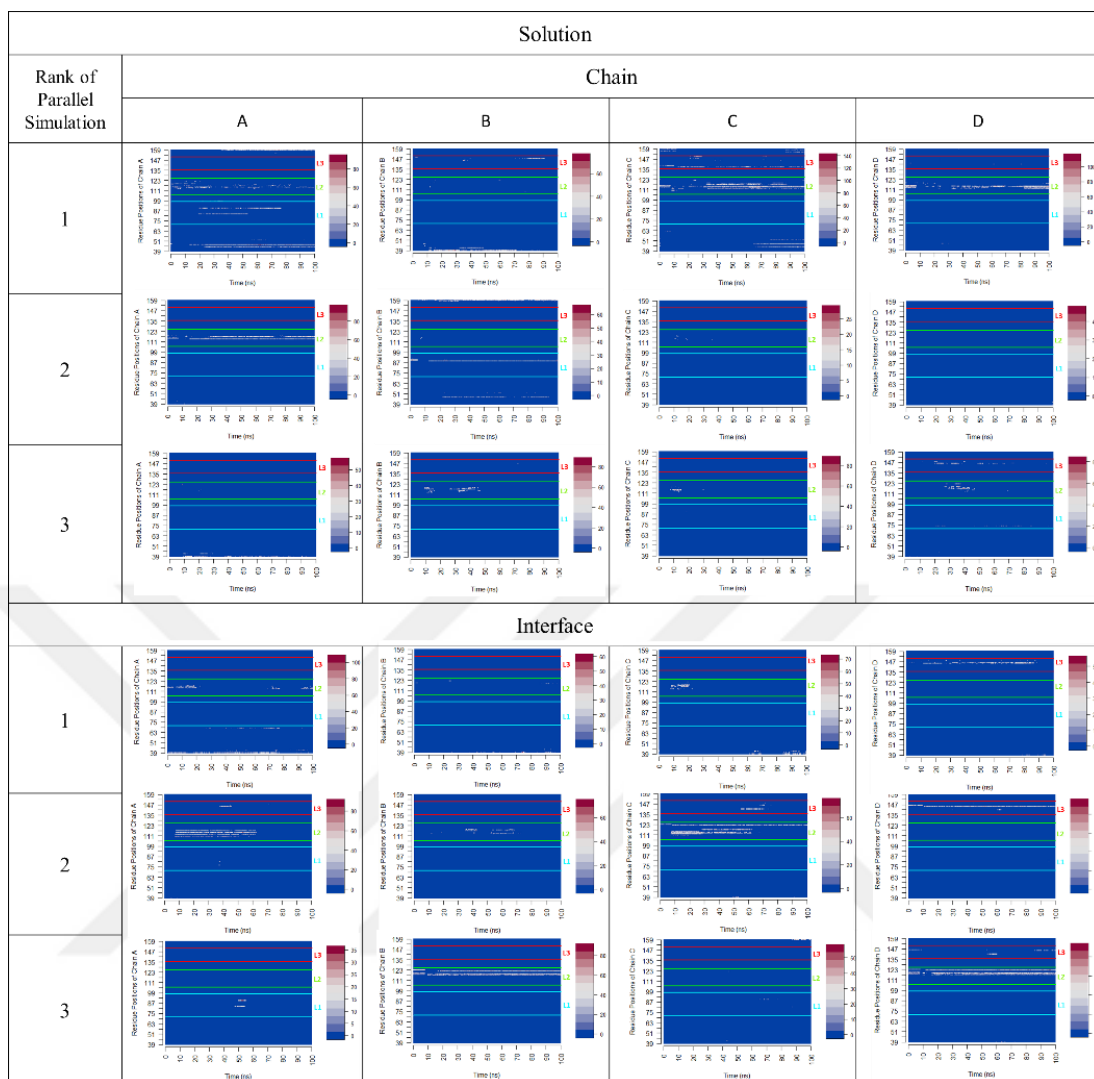


Figure 29. Hydrophobic contact maps of both Orientation 1 solution and interface simulations. Each map shows the interaction of the chain column with the remaining chains of the simulation system. L1, L2, and L3 loops were highlighted on the maps. The number of atomic contacts was assigned to a color gradient from dark blue to red. The distance cutoff was set to 4 Å.

The number of hydrogen bonds was also measured between clustered chains of Orientation 1. A conserved and relatively stable number of hydrogen bonds were detected between chain pairs at the interface clusters; larger fluctuations were observed in solution clusters (Figure 30).

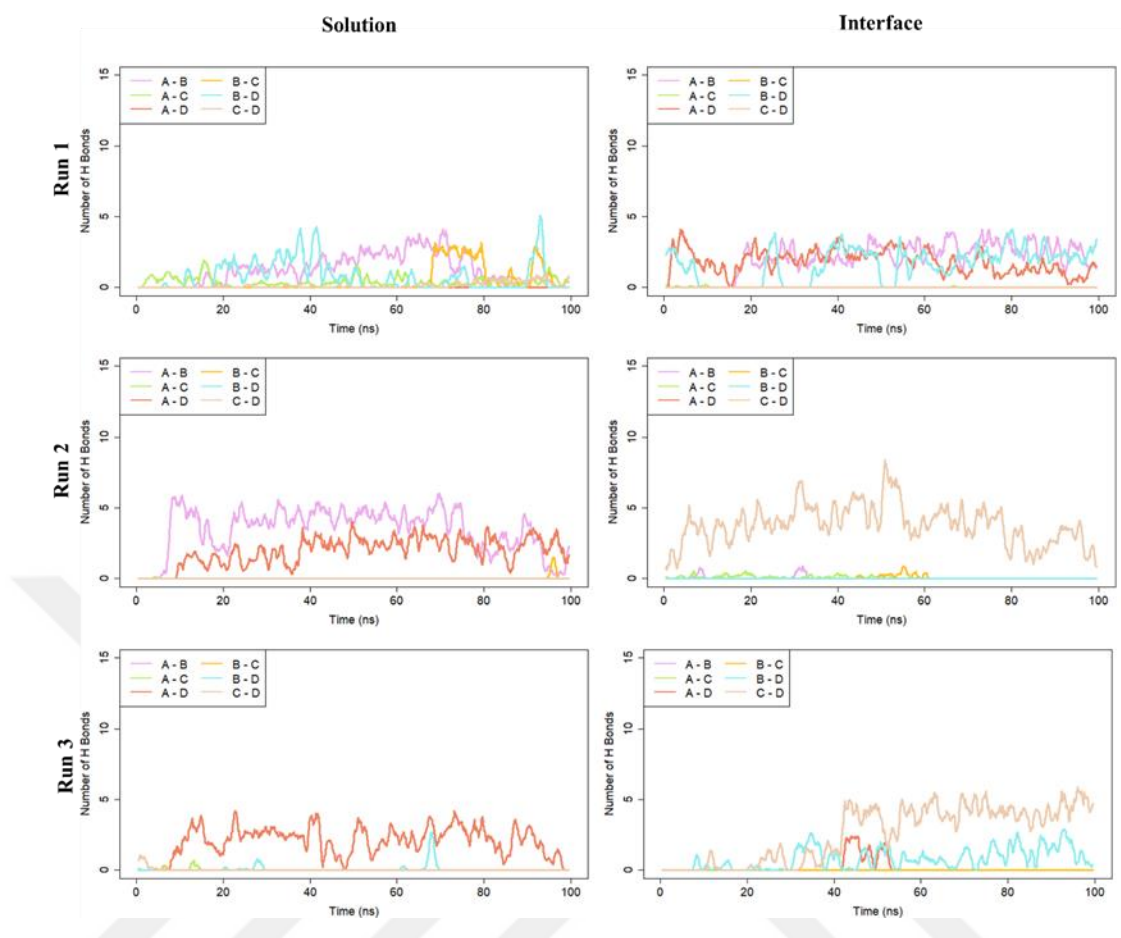


Figure 30. Measurements of the number of hydrogen bonds between chain pairs in Orientation 1 simulation systems in both environments. The chain pairs A & B, A & C, A & D, B & C, B & D, and C & D were shown in purple, green, orange, yellow, cyan, and beige lines in the graphs, respectively. Donor-acceptor distance and angle cutoff were set at 3.5 Å and 30° respectively.

The number of conserved salt bridges was low in both environments; only one salt bridge was stable in the second parallel solution run and the interacting pairs were in L1 (GLU97, chain A) and non-loop (LYS67, chain D) regions, respectively. At the interface, two conserved salt bridges were observed in two independent simulations: between non-loop regions of chains B (ASP46) and D (LYS67) and between L1 loop and non-loop region of chains C (GLU79) and D (LYS128), respectively. The majority of the salt bridges were transient (Figure 31).

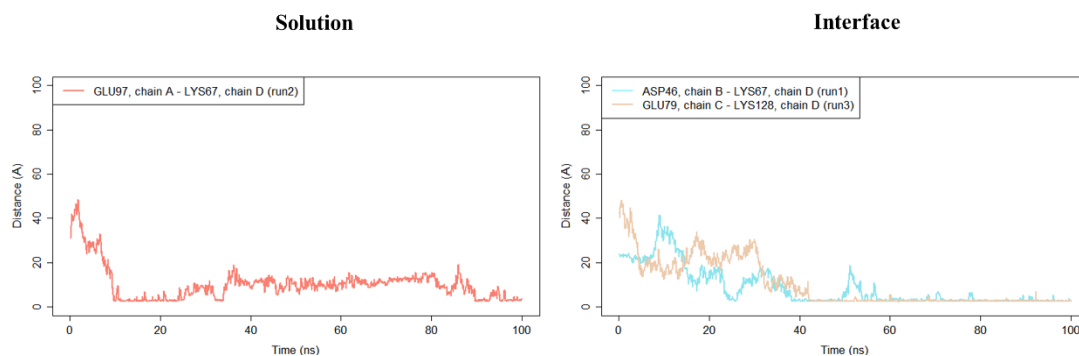


Figure 31. Measurements of conserved interchain salt bridges throughout the trajectories in Orientation 1 simulation systems for both environments. The distance between residues that form bridges was measured in the time course of the simulations. The chain pairs A & B, A & C, A & D, B & C, B & D, and C & D were shown in purple, green, orange, yellow, cyan, and beige lines in the graphs, respectively. Oxygen-nitrogen distance cutoff was set to 3.2 Å. However, salt bridge distance up to 4 Å was acceptable as conserved interaction.

Monomers were more stable at air/water interface simulations. In solution, higher RMSD values were measured for chain C in the last parallel run. L1 and L3 loops were fluctuating on a large scale in solution relative to loops at the interface (Figure 32).

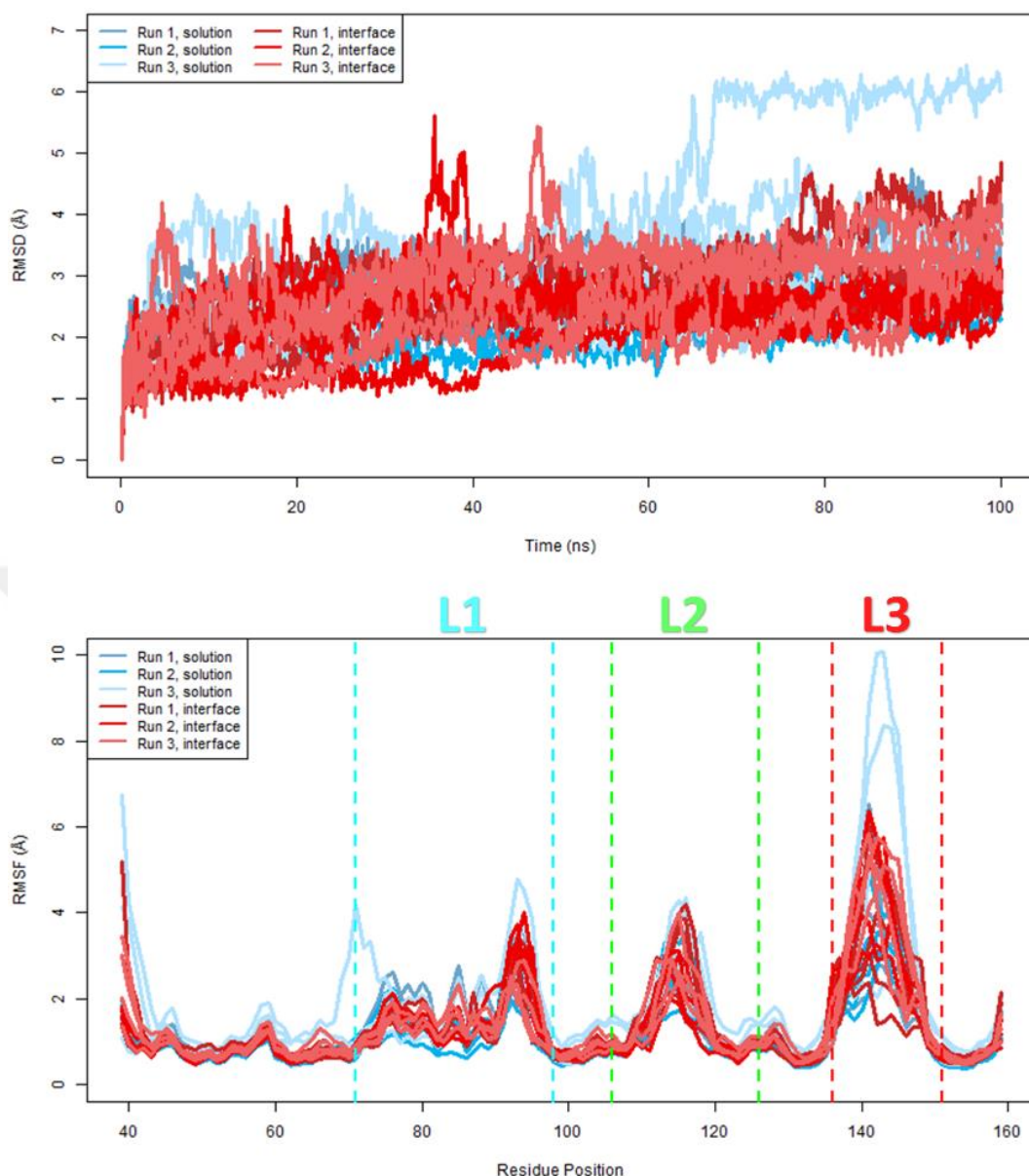


Figure 32. RMSD and RMSF measurements of individual chains in Orientation 1 simulation systems in both environments. Blue and red color gradients (from light to dark) were used to differentiate parallel solution and interface simulations, respectively. Regions between dashed cyan, green, and red lines show L1, L2, and L3 loops, respectively.

4.3.1.2 Orientation 2

Clusters of at least two monomers were observed in both simulation environments (Figure 33). In the first solution run, Chains A and C formed a cluster by interacting through their loop regions in the early steps of the trajectory with fluctuating COM-COM distance between 30 Å and 40 Å (Figure 34). After 20 ns, chains A, B, and D

were transiently interacting, and three monomers formed an L-shaped cluster in which chains A and D connected via loop regions and B was interacting with D with a shorter chain–chain distance in the last 20 ns. In the last parallel run of solution systems, chain–chain distances were not stable and transient interactions observed; clusters of chains C and D remained in the last 10 ns (Figure 34).

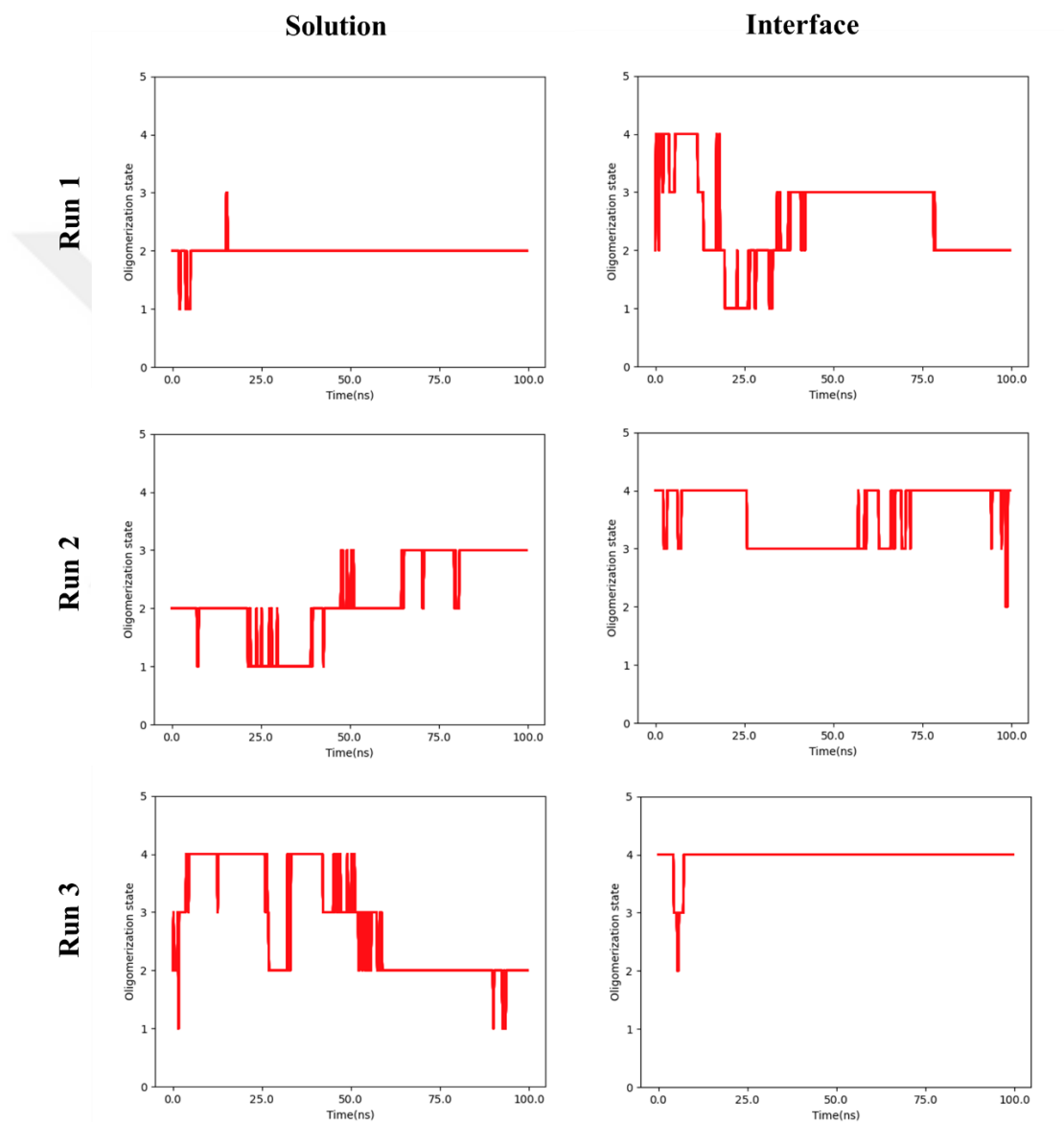


Figure 33. Cluster analysis of Orientation 2 simulations. Monomers in a 4 Å distance cutoff assumed as clustered in the simulation systems and the cluster lifetime was measured throughout the trajectories for both simulation environments.

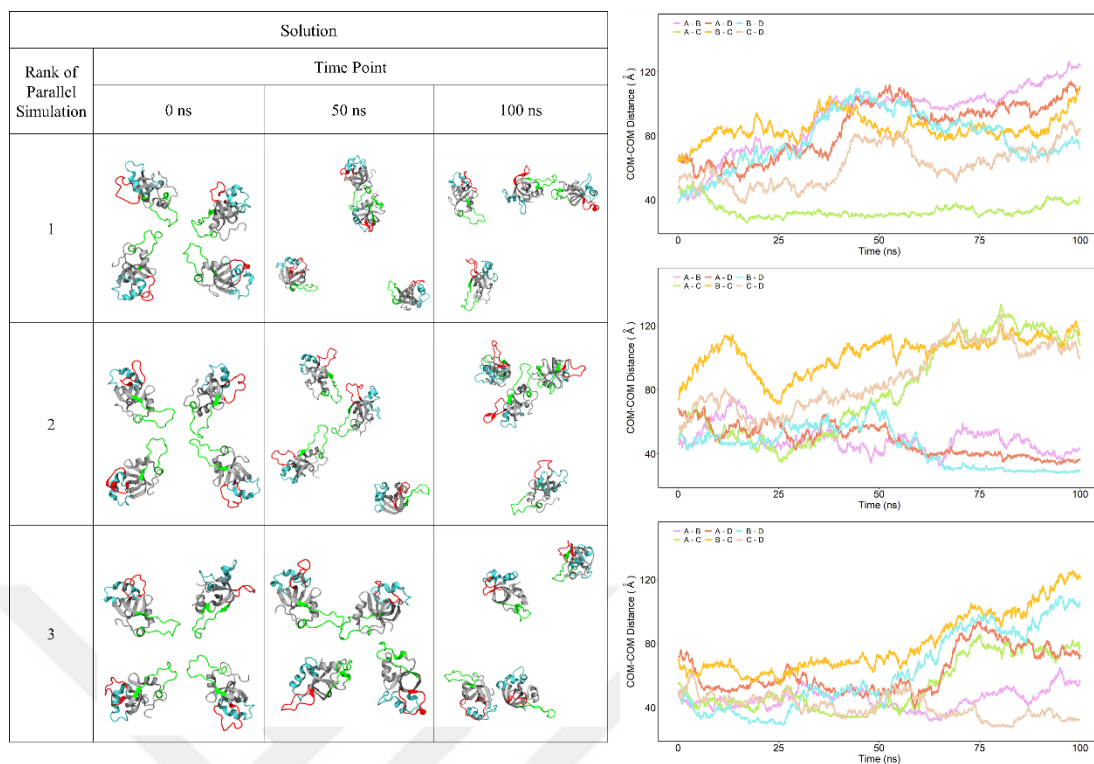


Figure 34. Snapshots of conformations at three different time points and COM-COM distances between chains of Orientation 2 solution simulations. Snapshots were taken at the beginning, middle, and end of the trajectory. L1, L2, and L3 loops and non-loop regions were denoted in cyan, green, red, and gray cartoon representations, respectively. The COM-COM distances between chains A & B, A & C, A & D, B & C, B & D, and C & D were shown in purple, green, orange, yellow, cyan, and beige lines, respectively.

At interface simulations, initial interactions of four monomers were disrupted and an L-shaped cluster of three monomers formed after 35 ns in the trajectory of the first run. In the last 20 ns, a cluster of chains A and C was conserved. L2 loops of chains A, B, and D were interacting with the air at the upper air/water interface. In the second parallel run, chains A, C, and D were in proximity of each other. However, a stable cluster was only formed between two monomers. Each monomer was interacting with the air through their L2 or L3 loops at the upper interface. In the last parallel interface simulation, four monomers were near each other from the early time points interacting via their L2 loops. After 40 ns, the COM-COM distance between chains B and D decreased and resulted in a slightly more compact cluster. Both L2 and L3 loops of chains B, C, and D were aligned at the interface; chain A were interacting with the air only through its L2 loop (Figure 35).

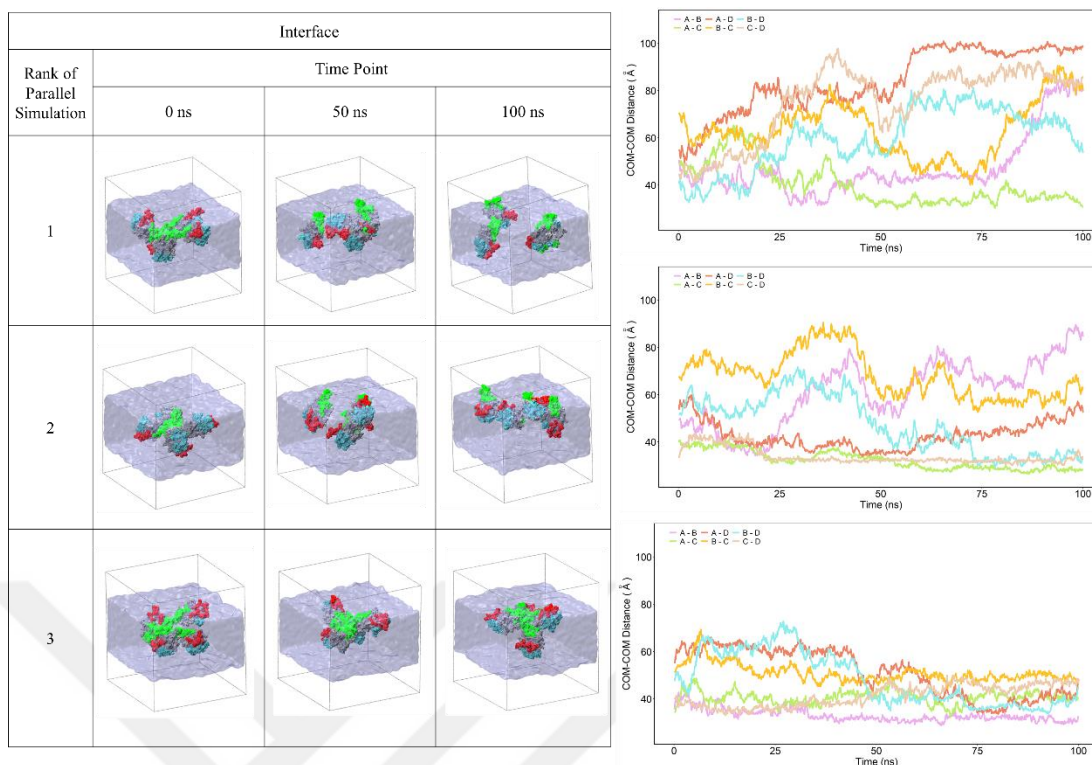


Figure 35. Snapshots of conformations at three different time points and COM–COM distances between chains of Orientation 2 interface simulations. Snapshots were taken at the beginning, middle, and end of the trajectory. L1, L2, and L3 loops and non-loop regions were denoted in cyan, green, red, and gray surface representations, respectively. The COM–COM distances between chains A & B, A & C, A & D, B & C, B & D, and C & D were shown in purple, green, orange, yellow, cyan, and beige lines, respectively.

As in Orientation 1 simulations, N-termini and L2 loops of chains were in interaction with other chains with the dominance of N-termini in solution (Figure 36). The highest number of atomic contacts and larger interacting regions were observed in L2 loops in solution clusters followed by N-termini at the interface. L2-L2 interactions had the highest probabilities in solution simulations and were found to be conserved in clusters of two monomers in at least two solution runs unlike interface (Figure 36). The number and variety of the conserved intermolecular interaction regions (N-termini, L1, L1-L2, L2, L2-L3, and L3) were higher, and the duration of the interaction was also longer at the air/water interface (Appendix 8). Although transient interactions were observed in various regions throughout the trajectories, conserved atomic contacts were concentrated in small regions at the interface clusters (Figure 37).

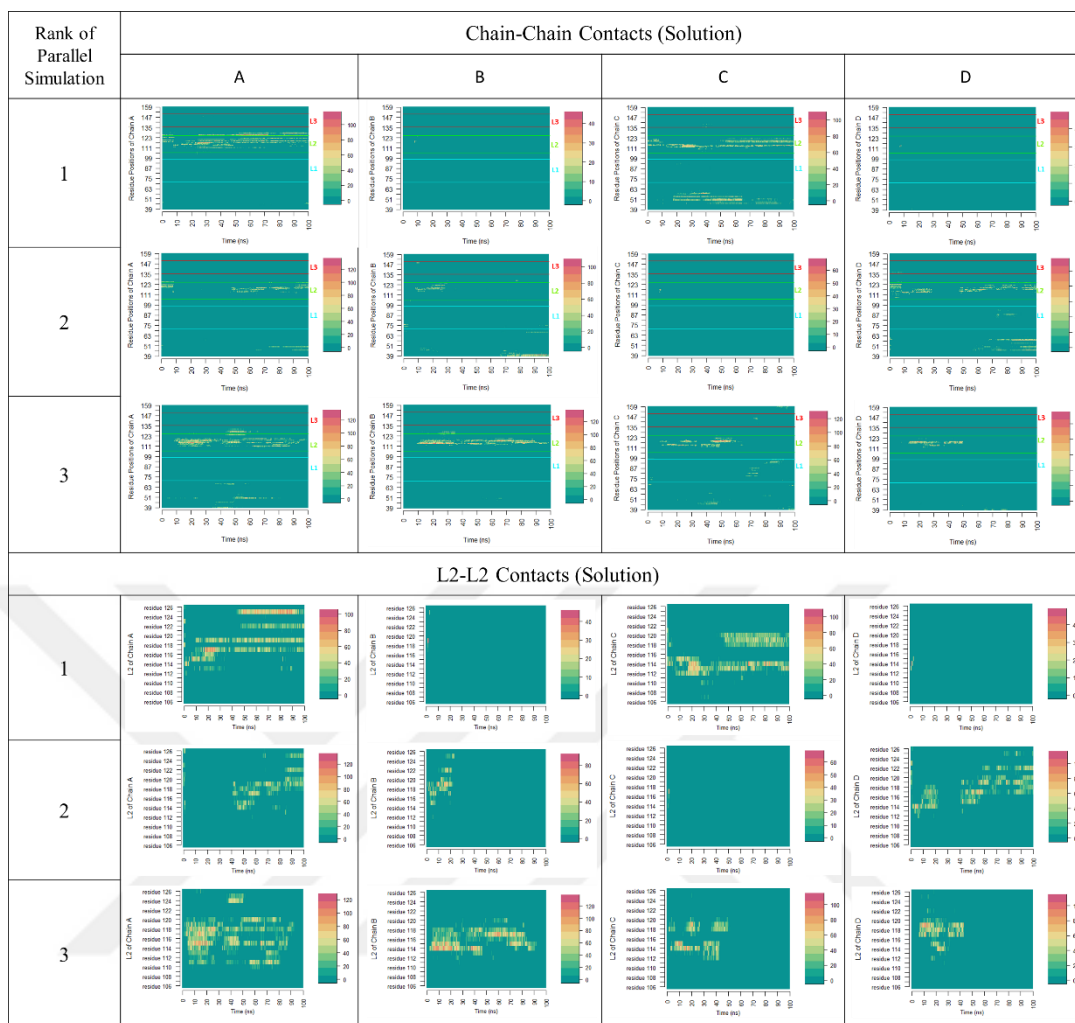


Figure 36. Contact maps of Orientation 2 solution simulations. Each map shows the interaction of the chain column with the remaining chains of the simulation system. L1, L2, and L3 loops were highlighted on the maps. L2-L2 interactions were measured by selecting only L2 loops. Light yellow background on the maps indicates no atomic interactions measured. The number of atomic contacts was assigned to a color gradient from dark green to red. The distance cutoff was set to 4 Å.

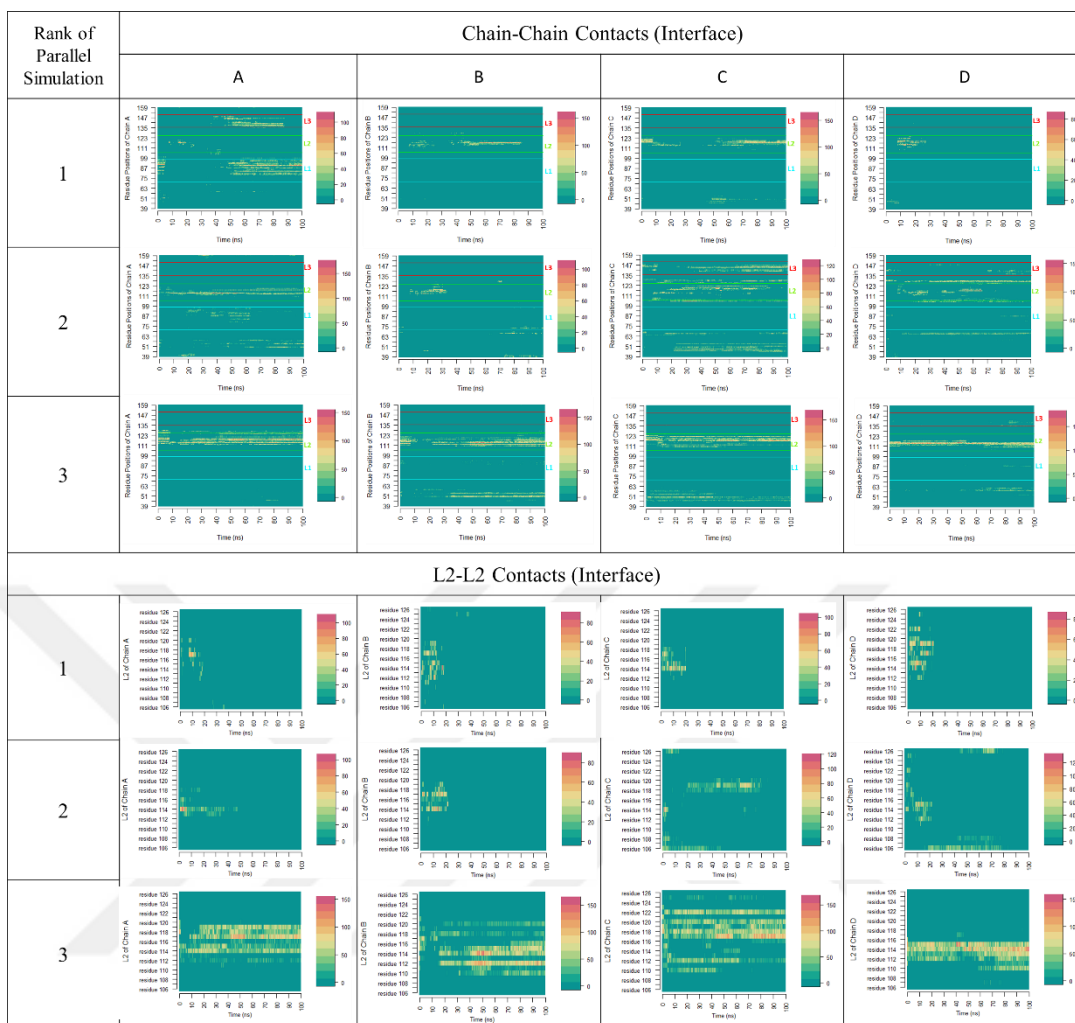


Figure 37. Contact maps of Orientation 2 interface simulations. Each map shows the interaction of the chain column with the remaining chains of the simulation system. L1, L2, and L3 loops were highlighted on the maps. L2-L2 interactions were measured by selecting only L2 loops. Light yellow background on the maps indicates no atomic interactions measured. The number of atomic contacts was assigned to a color gradient from dark green to red. The distance cutoff was set to 4 Å.

Hydrophobic interactions in Orientation 2 clusters were mostly established through residues in L2 loops of in both solution and interface simulations. The number of interaction regions was larger at the interface with the longer durations (Figure 38).

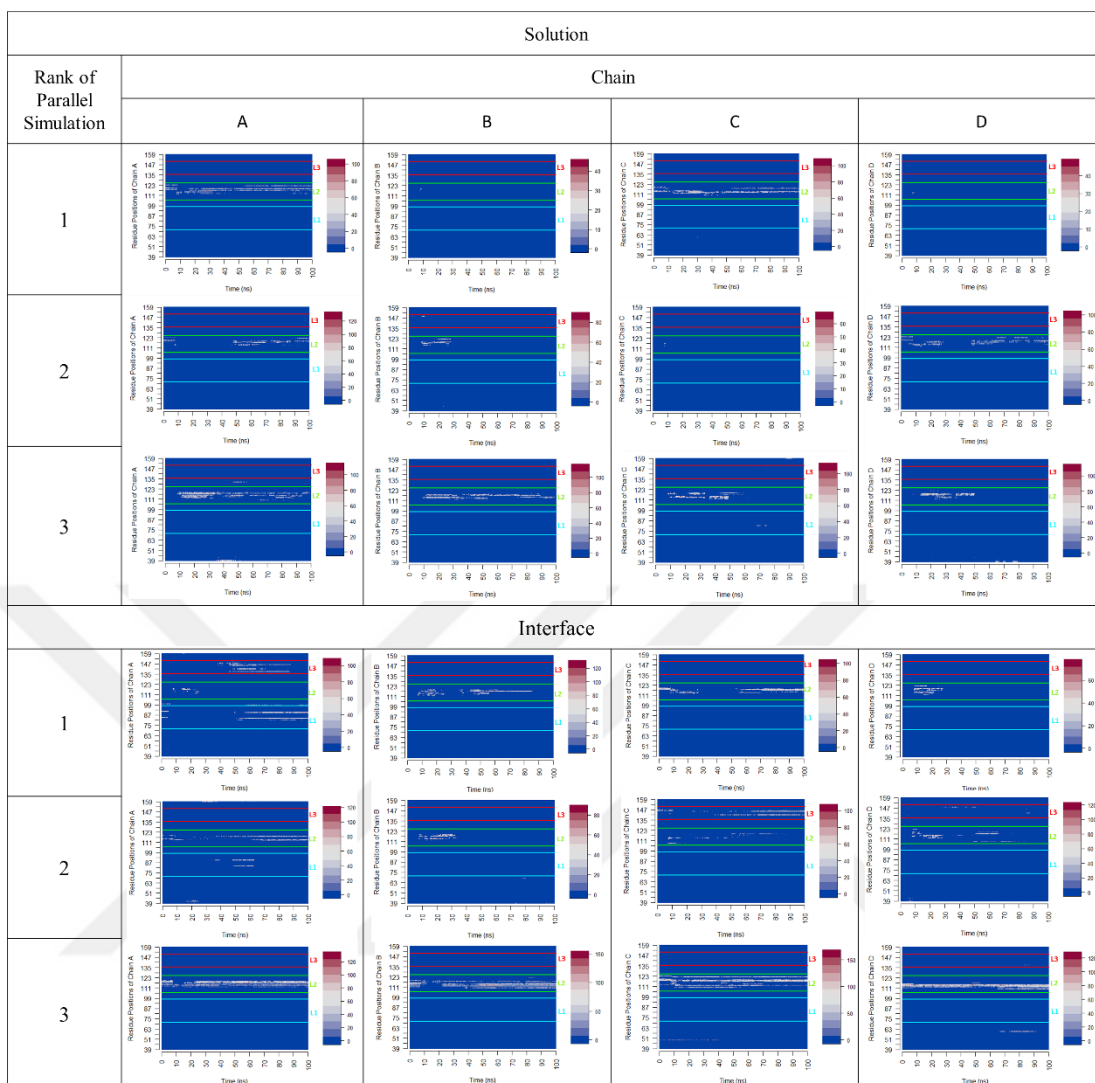


Figure 38. Hydrophobic contact maps of both Orientation 2 solution and interface simulations. Each map shows the interaction of the chain column with the remaining chains of the simulation system. L1, L2, and L3 loops were highlighted on the maps. The number of atomic contacts were assigned to a color gradient from dark blue to red. The distance cutoff was set to 4 Å.

In both simulation environments, the average number of hydrogen bonds was below five. Hydrogen bonds formed between chain pairs were not stable in solution. An increasing trend was observed in the number of bonds in the last 40 ns in the cluster only in the second solution run. At the interface simulations, hydrogen bonds between the cluster of chains A and C in the first run formed after 50 ns and conserved. In the second parallel run, the contribution of the hydrogen bonds to the cluster stability increased after 60 ns. It was observed that hydrogen bonds were formed in the early

simulation time steps of last two parallel runs and were conserved in the rest of the interface simulations (Figure 39).

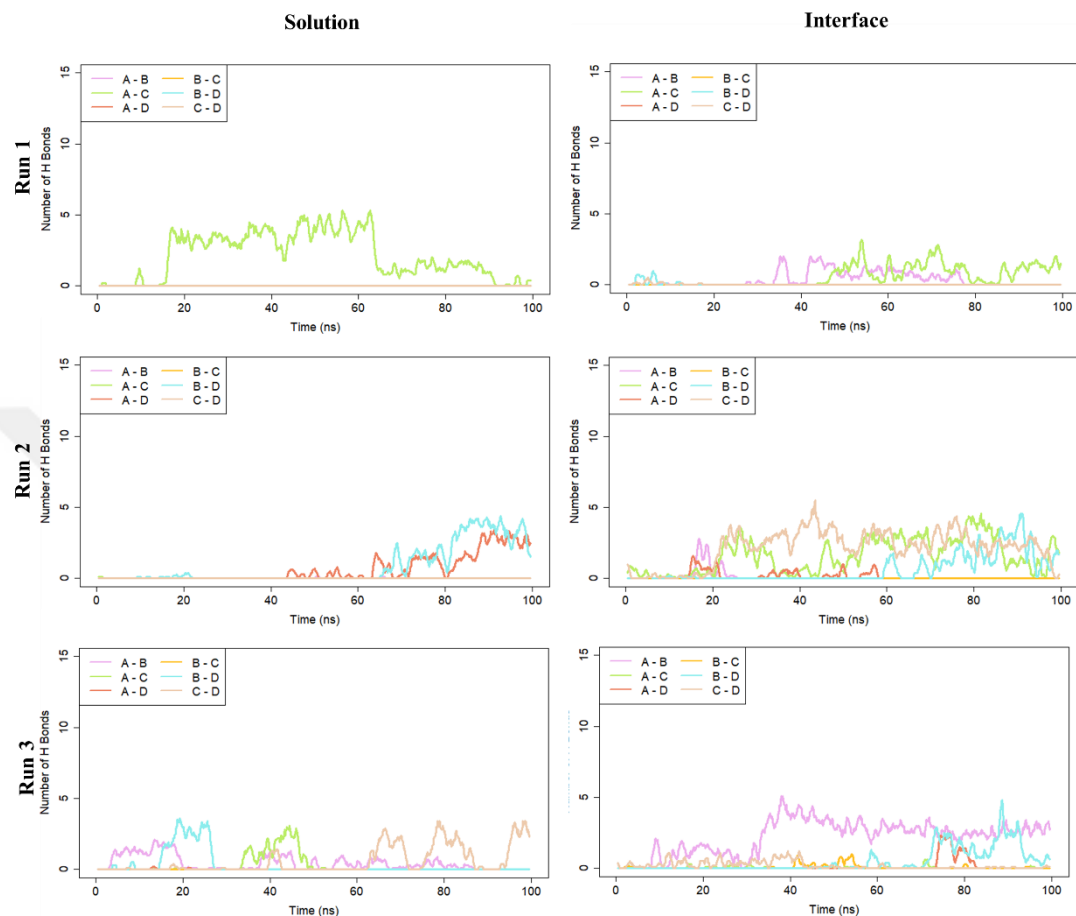


Figure 39. Measurements of the number of hydrogen bonds between chain pairs in Orientation 2 simulation systems in both environments. The chain pairs A & B, A & C, A & D, B & C, B & D, and C & D were shown as purple, green, orange, yellow, cyan, and beige lines in the graphs, respectively. Donor-acceptor distance and angle cutoff were set at 3.5 Å and 30° respectively.

Conserved salt bridges were not common; a total of three were stable for a significant time in one parallel run in both environments and there were no loop-loop interactions in the conserved salt bridges. Both ionic interactions were stabilizing the same chain pair in solution (Figure 40).

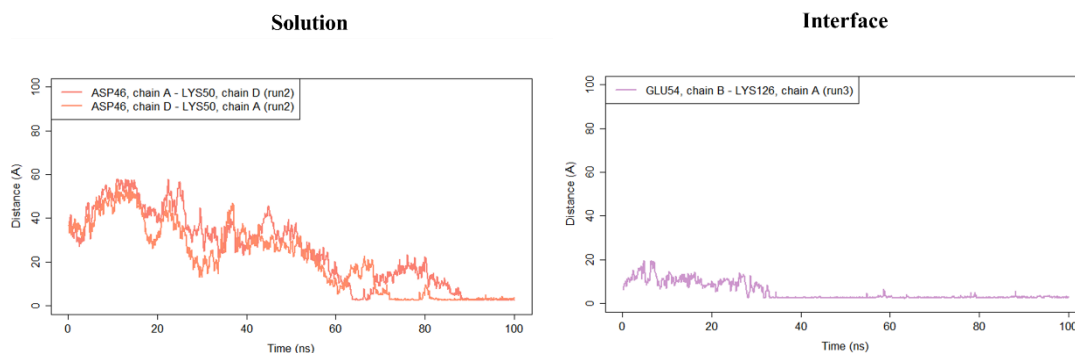


Figure 40. Measurements of conserved interchain salt bridges throughout the trajectories in Orientation 2 simulation systems for both environments. The distance between residues that form bridges was measured in the time course of the simulations. The chain pairs A & B and A & D were shown as purple and orange lines in the graphs, respectively. Oxygen-nitrogen distance cutoff was set to 3.2 Å. However, salt bridge distance up to 4 Å was acceptable as conserved interaction.

Most of monomers in interface simulations were more stable than the ones in solution simulations. The fluctuations in the RMSD measurements of interface monomers were resulted from the individual monomer behaviors; they were mainly (or after a time point) distant from the clusters. The cluster of four monomers in the last interface simulation was stable as the lower RMSD values were measured for the monomers in the cluster. L3 loops of monomers in solution were more mobile than the other loops both in solution and at the interface (Figure 41).

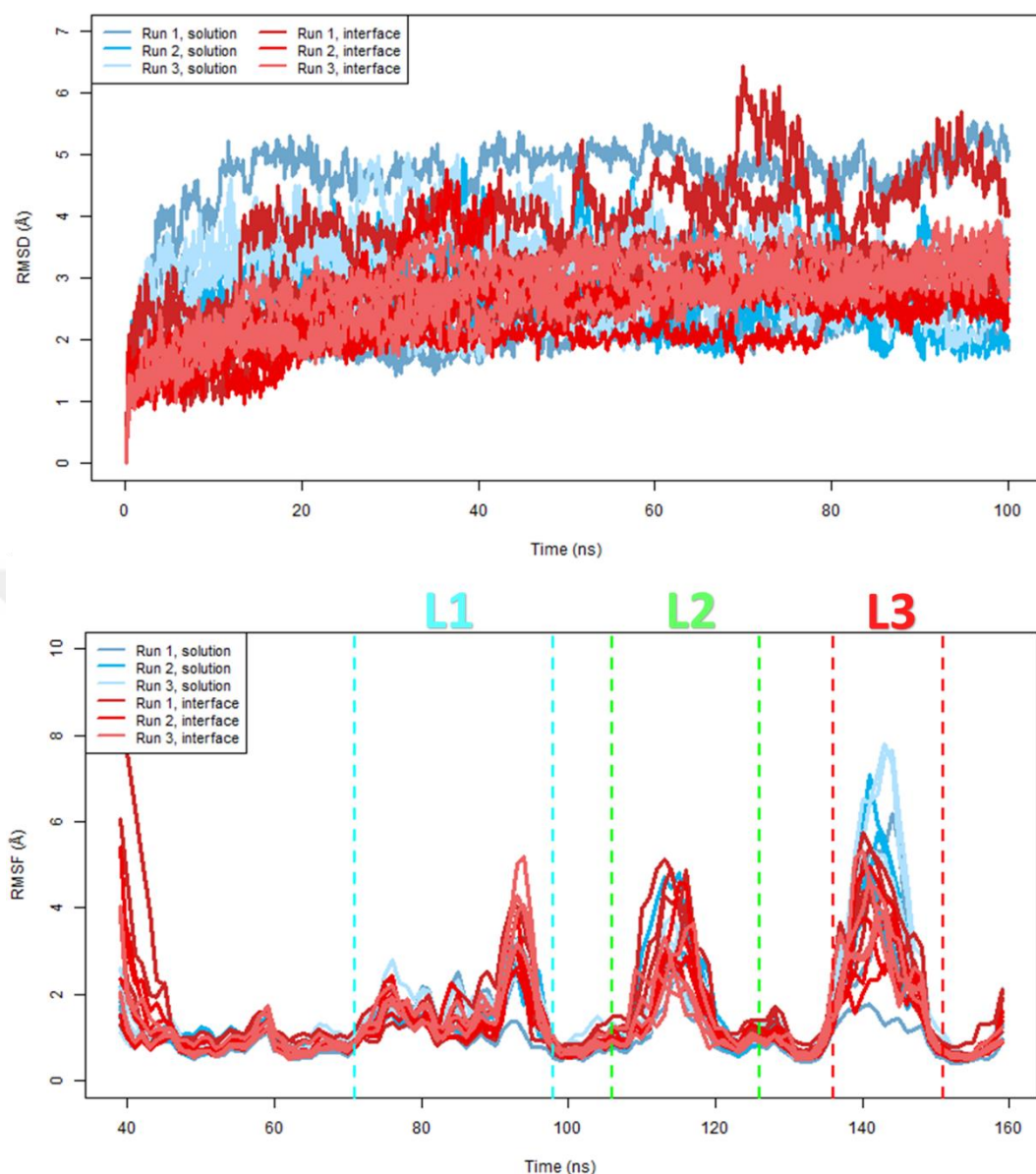


Figure 41. RMSD and RMSF measurements of individual chains in Orientation 2 simulation systems in both environments. Blue and red color gradients (from light to dark) were used to differentiate parallel solution and interface simulations, respectively. Regions between dashed cyan, green, and red lines show L1, L2, and L3 loops, respectively.

4.3.1.3 Orientation 3

Clusters of two or three monomers were observed in both simulation environments with the dominance of cluster of two at the interface. A sequential clustering behavior was observed in both simulation environments. There was no dominant cluster shape in both simulation environments. The clustering behavior was

observed in the earlier timesteps at the interface simulations (Figure 42). A conserved, planar-shaped cluster of chains A, B, and C formed after 25 ns and COM–COM distances between cluster monomers were below 40 Å in the second solution run. In the last run of the solution simulations, two clusters were observed: One from the beginning of the trajectory (chains B and D) and the other in the second half (chains A and B). However, neither of the clusters were stable even though the monomer distances were fluctuating below 30 Å (Figure 43).



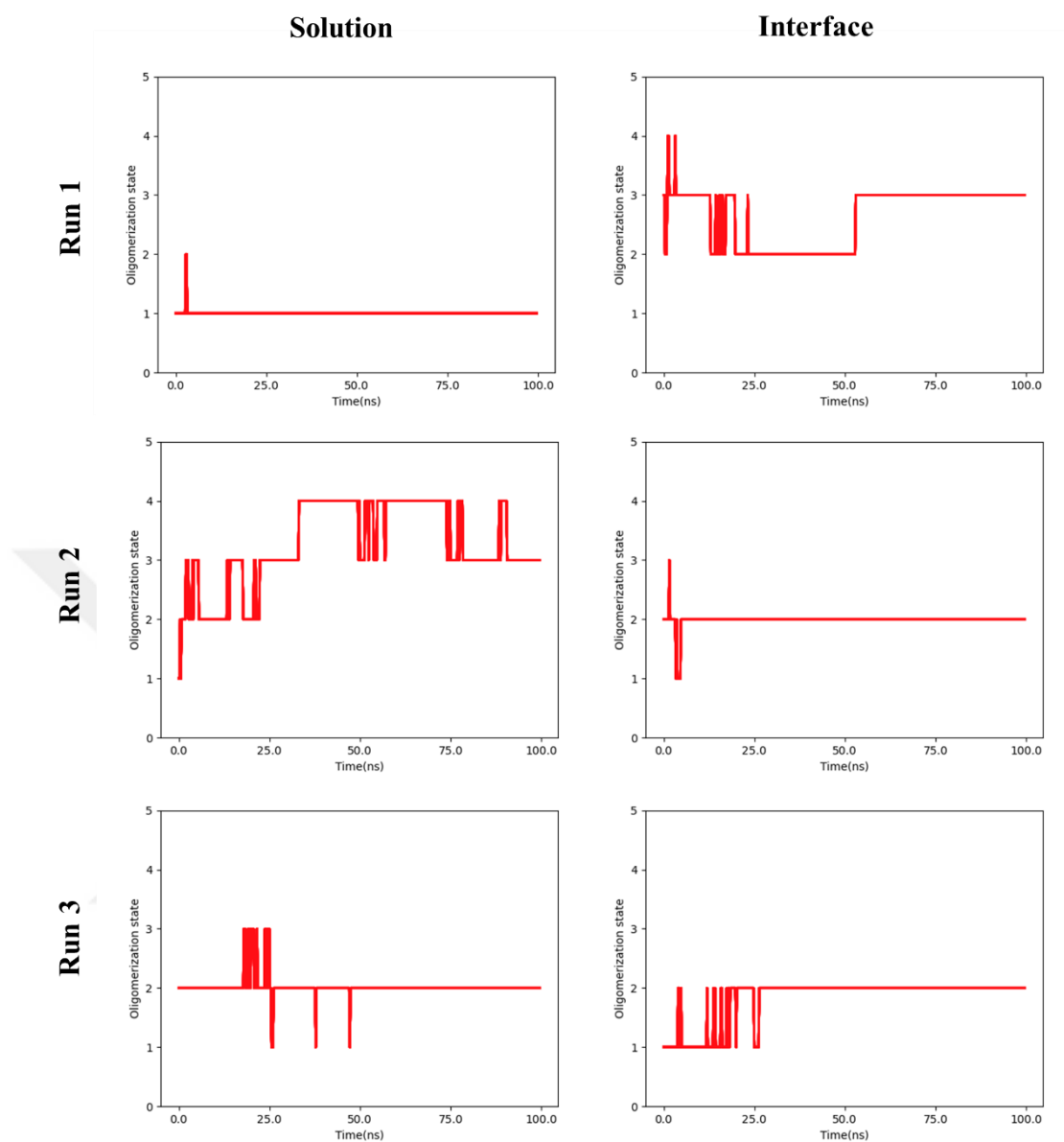


Figure 42. Cluster analysis of Orientation 3 simulations. Monomers in a 4 Å distance cutoff assumed as clustered in the simulation systems and the cluster lifetime was measured throughout the trajectories for both simulation environments.

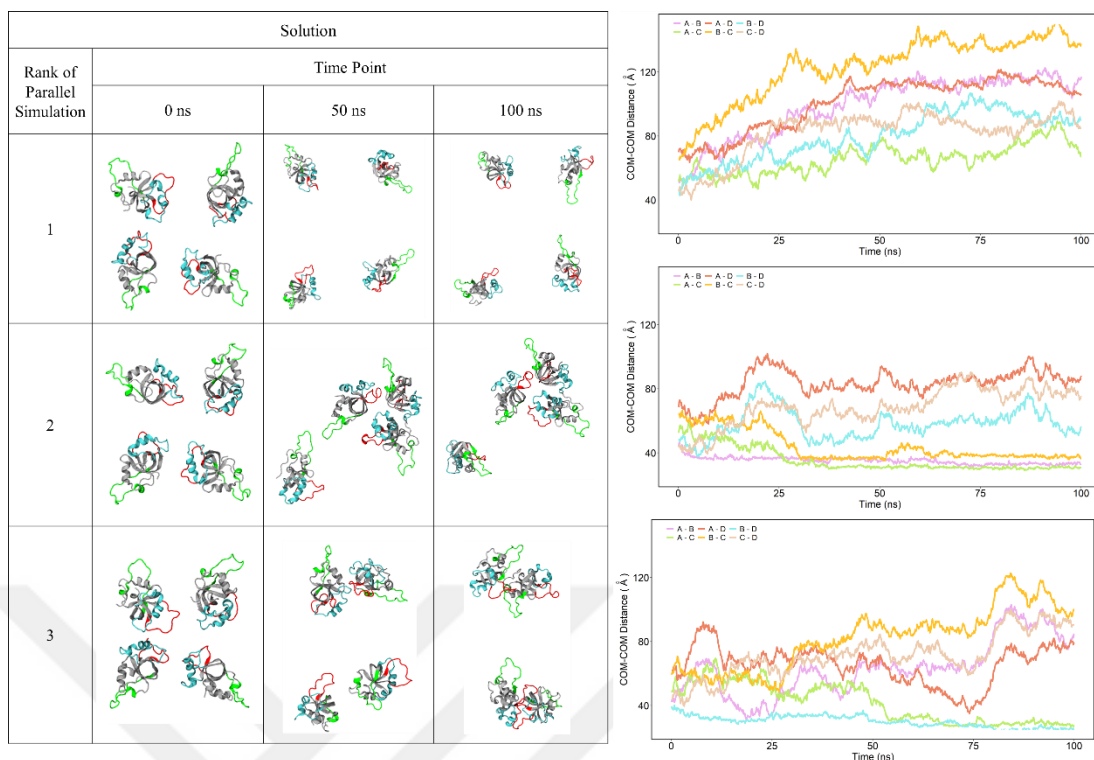


Figure 43. Snapshots of conformations at three different time points and COM–COM distances between chains of Orientation 3 solution simulations. Snapshots were taken at the beginning, middle, and end of the trajectory. L1, L2, and L3 loops and non-loop regions were denoted in cyan, green, red, and gray cartoon representations, respectively. The COM–COM distances between chains A & B, A & C, A & D, B & C, B & D, and C & D were shown as purple, green, orange, yellow, cyan, and beige lines, respectively.

Chains B, C, and D clustered linearly after 50 ns in the first interface simulation (Figure 42). Although the distance between chains C and D was stabilized around 30 Å in the last 20 ns, B, and D fluctuated relative to each other in the trimeric structure. A side-to-end cluster of chains C and D was conserved throughout the trajectory of the second parallel run. In the last parallel run, a stable cluster of chains A and B formed by interacting through their loop and non-loop regions after the first 20 ns. Surface adsorptions were mostly facilitated by L2 loops in all parallel runs (Figure 44).

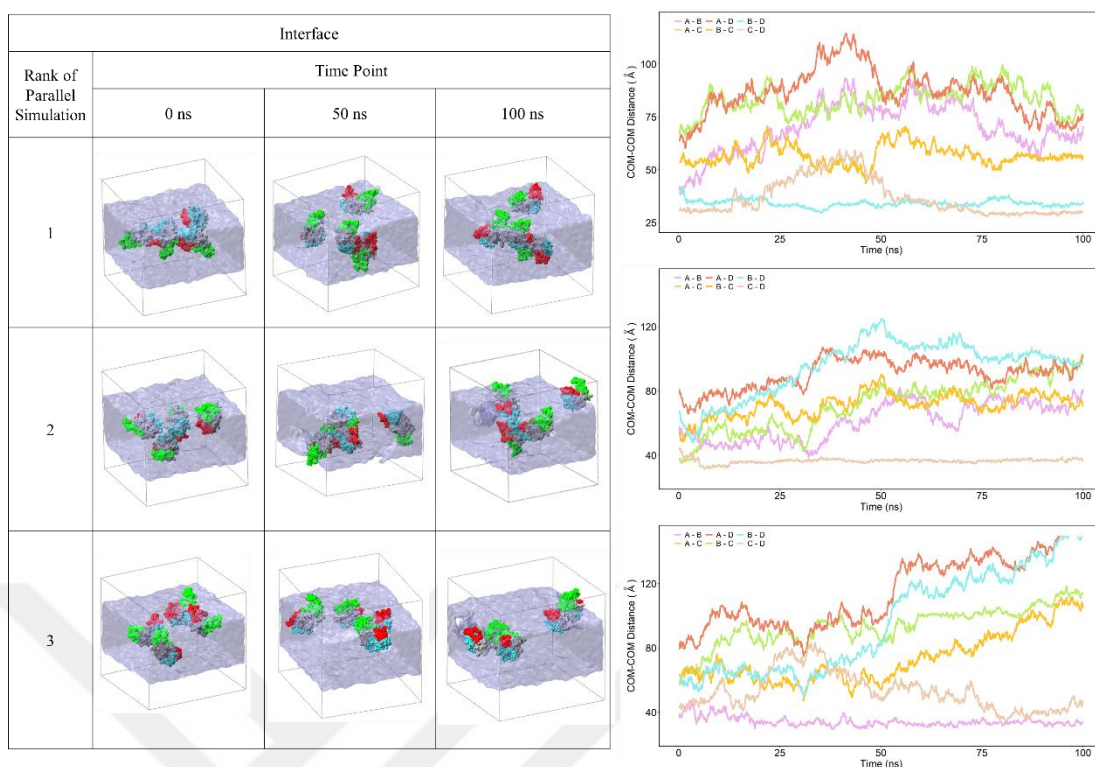


Figure 44. Snapshots of conformations at three different time points and COM–COM distances between chains of Orientation 3 interface simulations. Snapshots were taken at the beginning, middle, and end of the trajectory. L1, L2, and L3 loops and non-loop regions were denoted in cyan, green, red, and gray surface representations, respectively. The COM–COM distances between chains A & B, A & C, A & D, B & C, B & D, and C & D were shown as purple, green, orange, yellow, cyan, and beige lines, respectively.

The dominant chain–chain interaction regions were N-termini and L3 loop in solution and L2 loop in interface clusters in Orientation 3 simulations. Elevated number of atomic contacts also concentrated on these regions in solution clusters (Figure 45). However, L2-L3 regions and L3 loops was making high number of atomic contacts at the interface. The interaction regions in the solution monomers were distributed over the monomer surface and more transient interactions were observed compared to the monomers in the interface simulation systems (Figure 46). On the other hand, interactions concentrating on the distinct regions at the interface clusters and higher intermolecular interaction probabilities were measured in both homogenous and heterogenous loop-loop interactions (Appendix 9). Conserved L2-L2 interactions were only observed in the clusters of the third interface run (Figure 46).

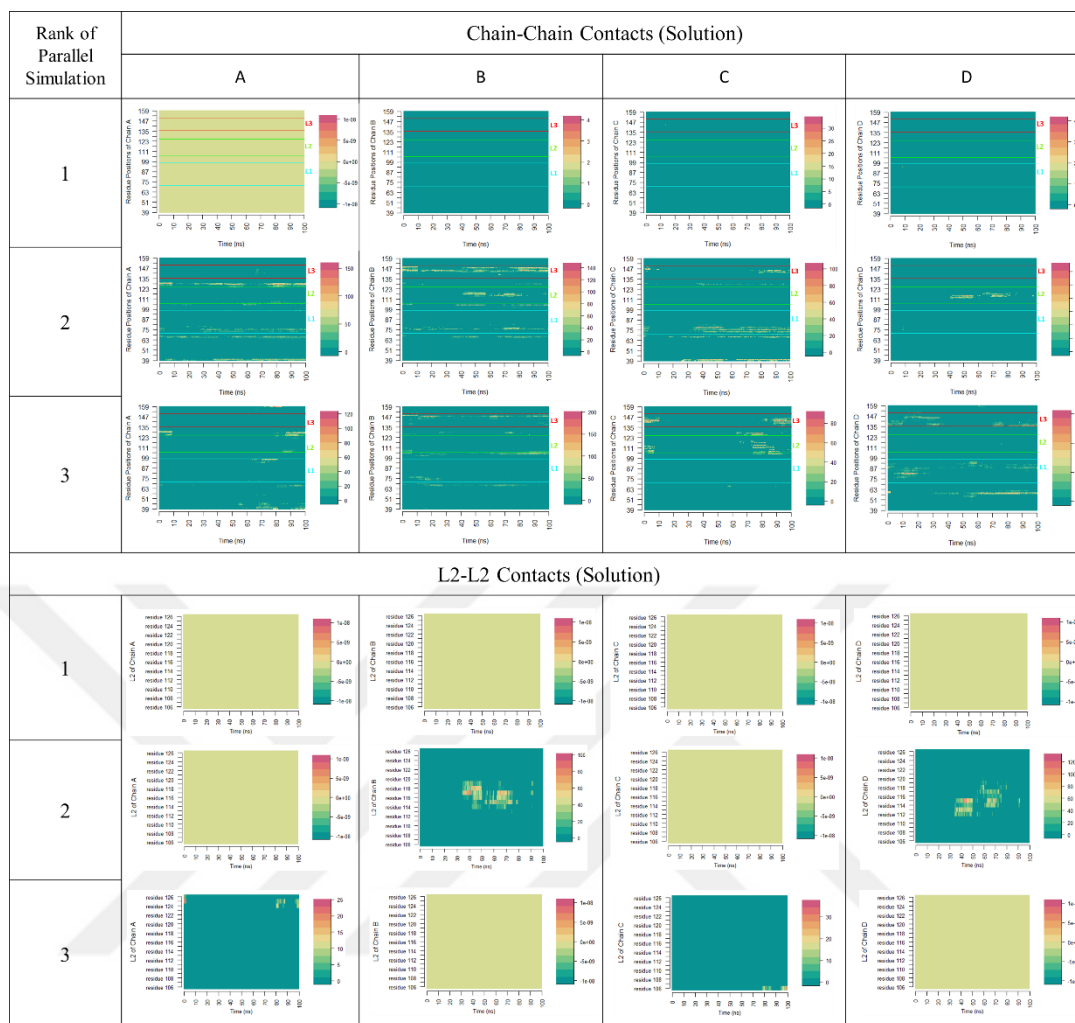


Figure 45. Contact maps of Orientation 3 solution simulations. Each map shows the interaction of the chain column with the remaining chains of the simulation system. L1, L2, and L3 loops were highlighted on the maps. L2-L2 interactions were measured by selecting only L2 loops. Light yellow background on the maps indicates no atomic interactions measured. The number of atomic contacts was assigned to a color gradient from dark green to red. The distance cutoff was set to 4 Å.

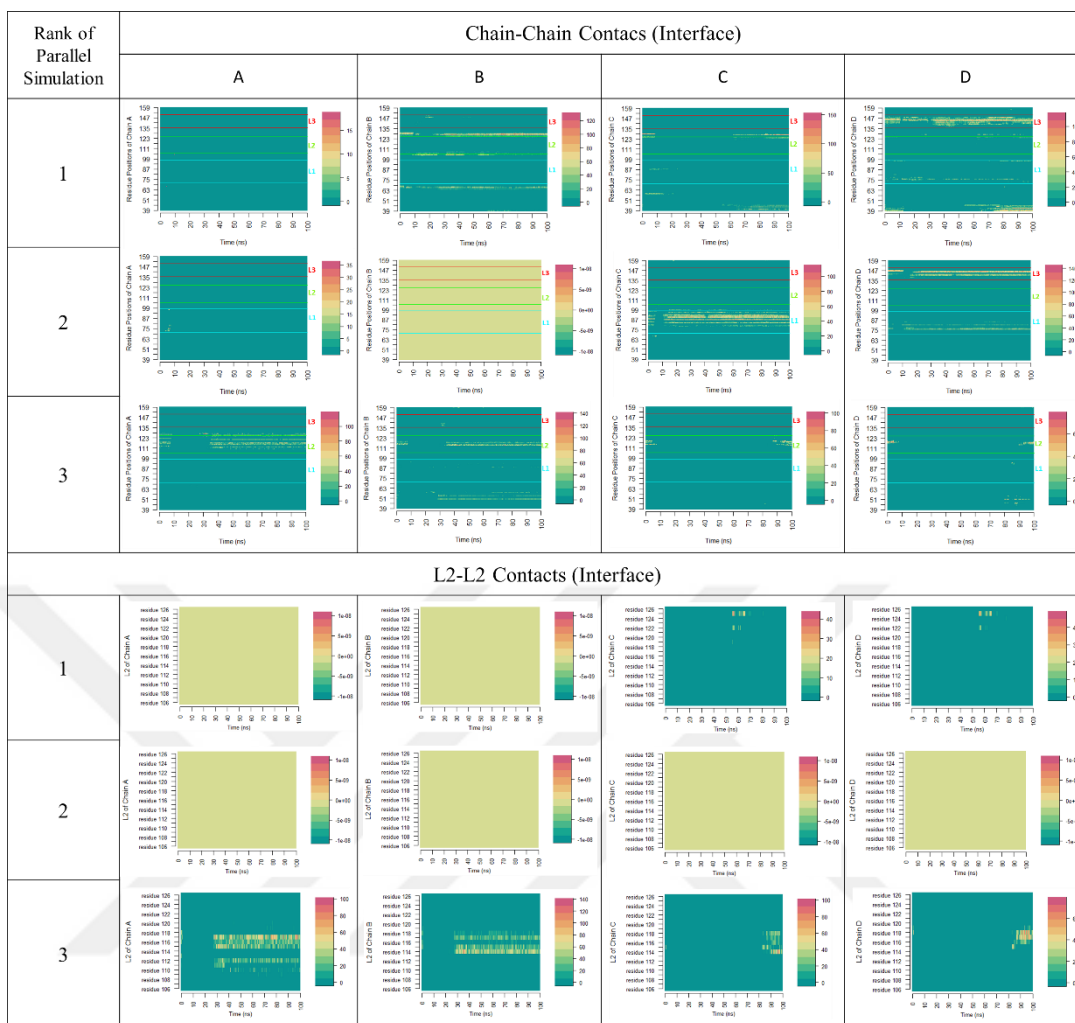


Figure 46. Contact maps of Orientation 3 interface simulations. Each map shows the interaction of the chain column with the remaining chains of the simulation system. L1, L2, and L3 loops were highlighted on the maps. L2-L2 interactions were measured by selecting only L2 loops. Light yellow background on the maps indicates no atomic interactions measured. The number of atomic contacts was assigned to a color gradient from dark green to red. The distance cutoff was set to 4 Å.

Hydrophobic interactions were observed only between L1 (chain D) and L3 (chain B) loops in the third parallel run of solution simulations. L2 and L3 in the second and each L2 loop of monomers in the last interface run had pronounced hydrophobic contacts (Figure 47).

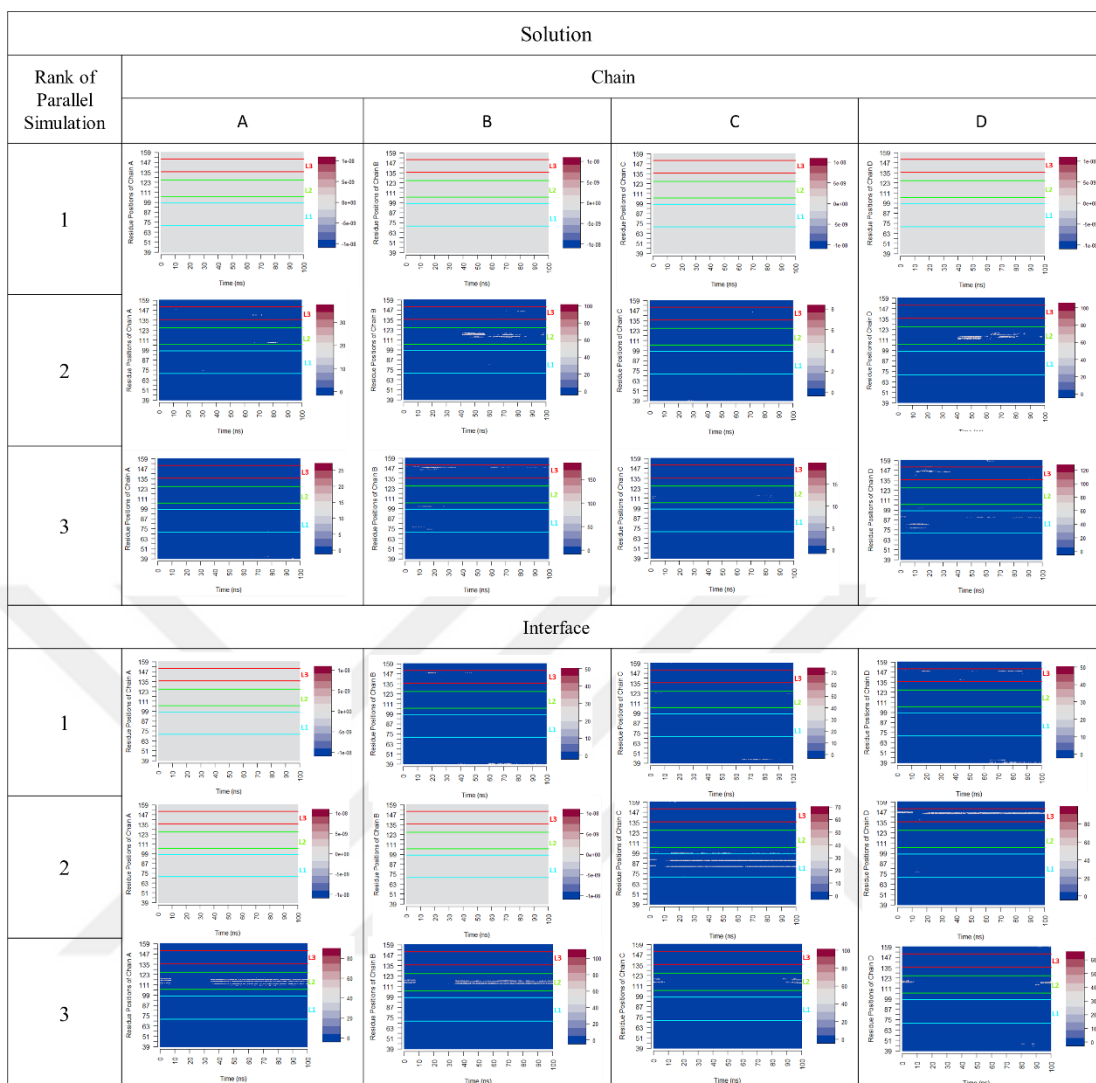


Figure 47. Hydrophobic contact maps of both Orientation 3 solution and interface simulations. Each map shows the interaction of the chain column with the remaining chains of the simulation system. L1, L2, and L3 loops were highlighted on the maps. The number of atomic contacts was assigned to a color gradient from dark blue to red. The distance cutoff was set to 4 Å.

A number of hydrogen bonds were formed and conserved mostly in last 60 ns in the clusters of the last two solution simulations. At the interface simulations, a high number of hydrogen bonds were observed both in clusters of two and three monomers; the number of hydrogen bonds were conserved with relatively small fluctuations (Figure 48).

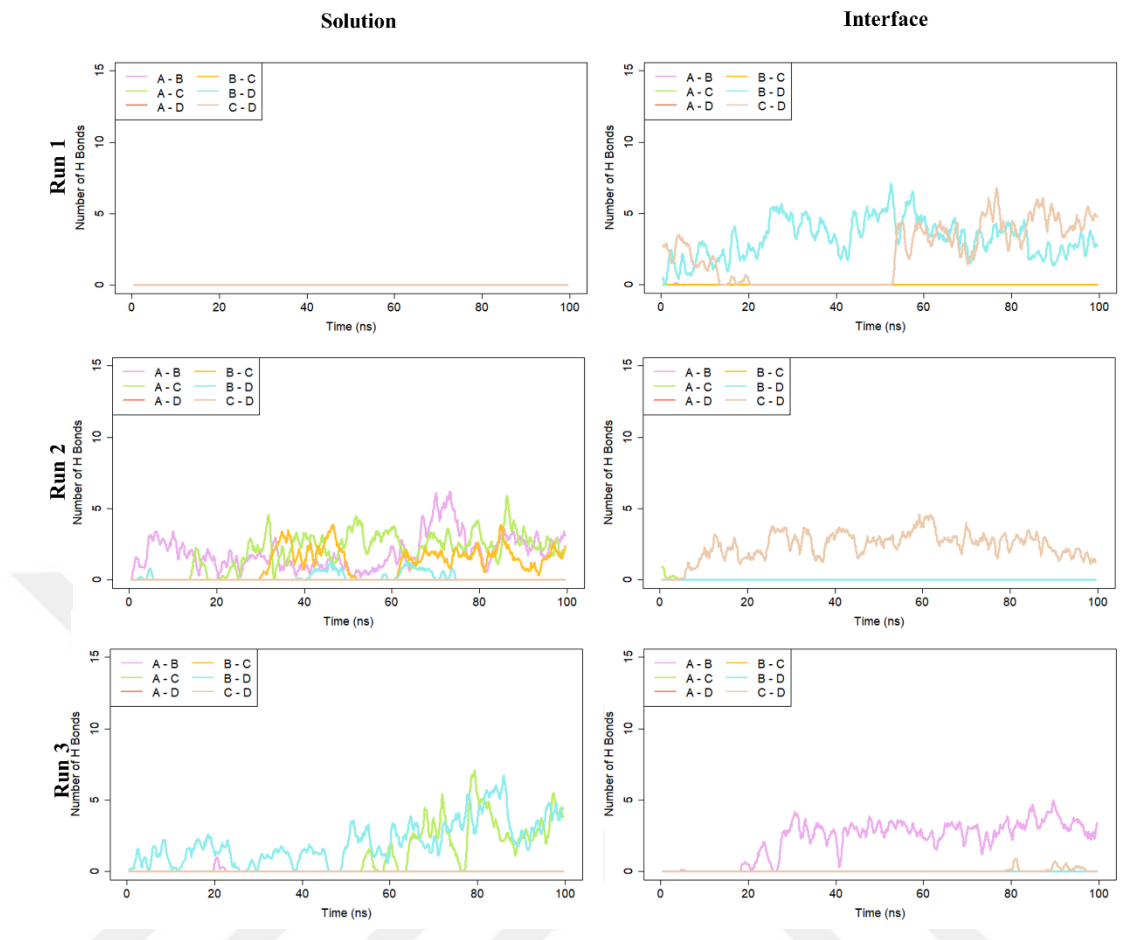


Figure 48. Measurements of the number of hydrogen bonds between chain pairs in Orientation 3 simulation systems in both environments. The chain pairs A & B, A & C, A & D, B & C, B & D, and C & D were shown as purple, green, orange, yellow, cyan, and beige lines in the graphs, respectively. Donor-acceptor distance and angle cutoff were set at 3.5 Å and 30° respectively.

The highest number of salt bridges was observed in the interface simulations of Orientation 3; at least one salt bridge was conserved in a run. The residues involved in the interaction were in L1 loop and a non-loop region in the second solution run. At the interface, the L1-L1 interaction formed with the help of a salt bridge in the second run. Other bridges were formed between non-loop regions or non-loop-loop regions (Figure 49).

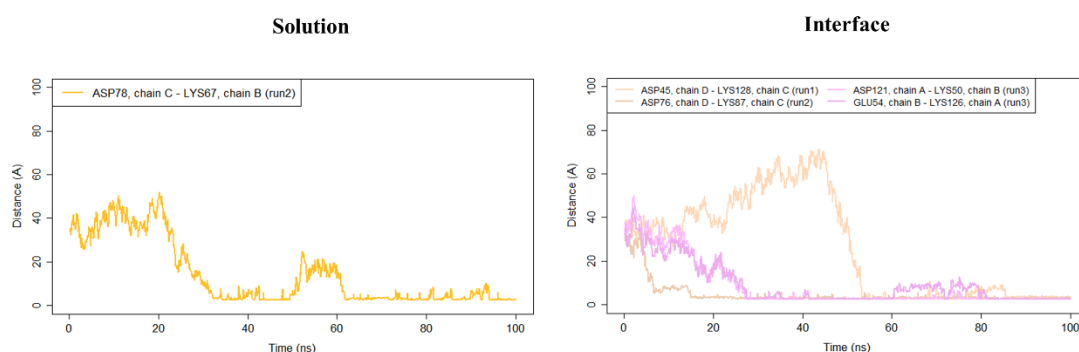


Figure 49. Measurements of conserved interchain salt bridges throughout the trajectories in Orientation 3 simulation systems for both environments. The distance between residues that form bridges was measured in the time course of the simulations. The chain pairs A & B and A & D were shown as purple and orange lines in the graphs, respectively. Oxygen-nitrogen distance cutoff was set to 3.2 Å. However, salt bridge distance up to 4 Å was acceptable as conserved interaction.

The structures of most monomers in solution simulations were more mobile than at the interface monomers. Some monomers at the interface were also experiencing relatively high mobility due to their individuality, especially in the first two simulations. L3 loops of interface monomers were mostly more stable than the monomers in the solution (Figure 50).

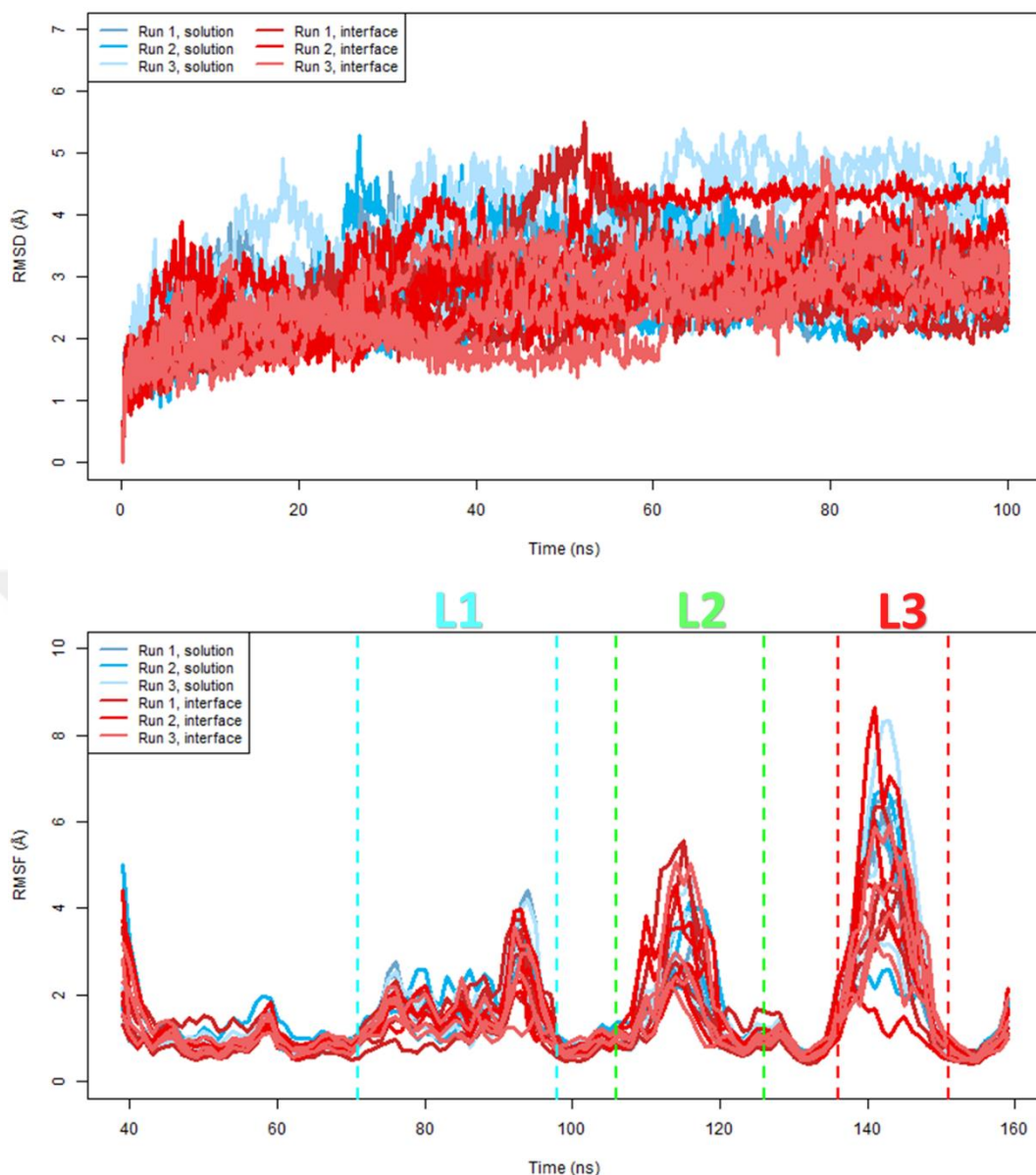


Figure 50. RMSD and RMSF measurements of individual chains in Orientation 3 simulation systems in both environments. Blue and red color gradients (from light to dark) were used to differentiate parallel solution and interface simulations, respectively. Regions between dashed cyan, green, and red lines show L1, L2, and L3 loops, respectively.

4.3.1.4 Orientation 4

In the last simulation systems, clusters of two monomers were observed in both simulation environments. In one of the parallel runs in two different environments, there was no conserved RodA assembly observed (Figure 51). In the first cluster, chains A and C were interacting through the loops with fluctuating COM–COM

distance at 45 Å, and an elongated shape was detectable for 70 ns. In the last 35 ns, the remaining chains were also clustered in a linear form. In the second run, the loops of chains C and D were in contact in the last 30 ns with a progressive decrease in the chain–chain distance. The cluster of chains A and B was not stable and dissociated at several time points throughout the trajectory in the last solution run (Figure 52).

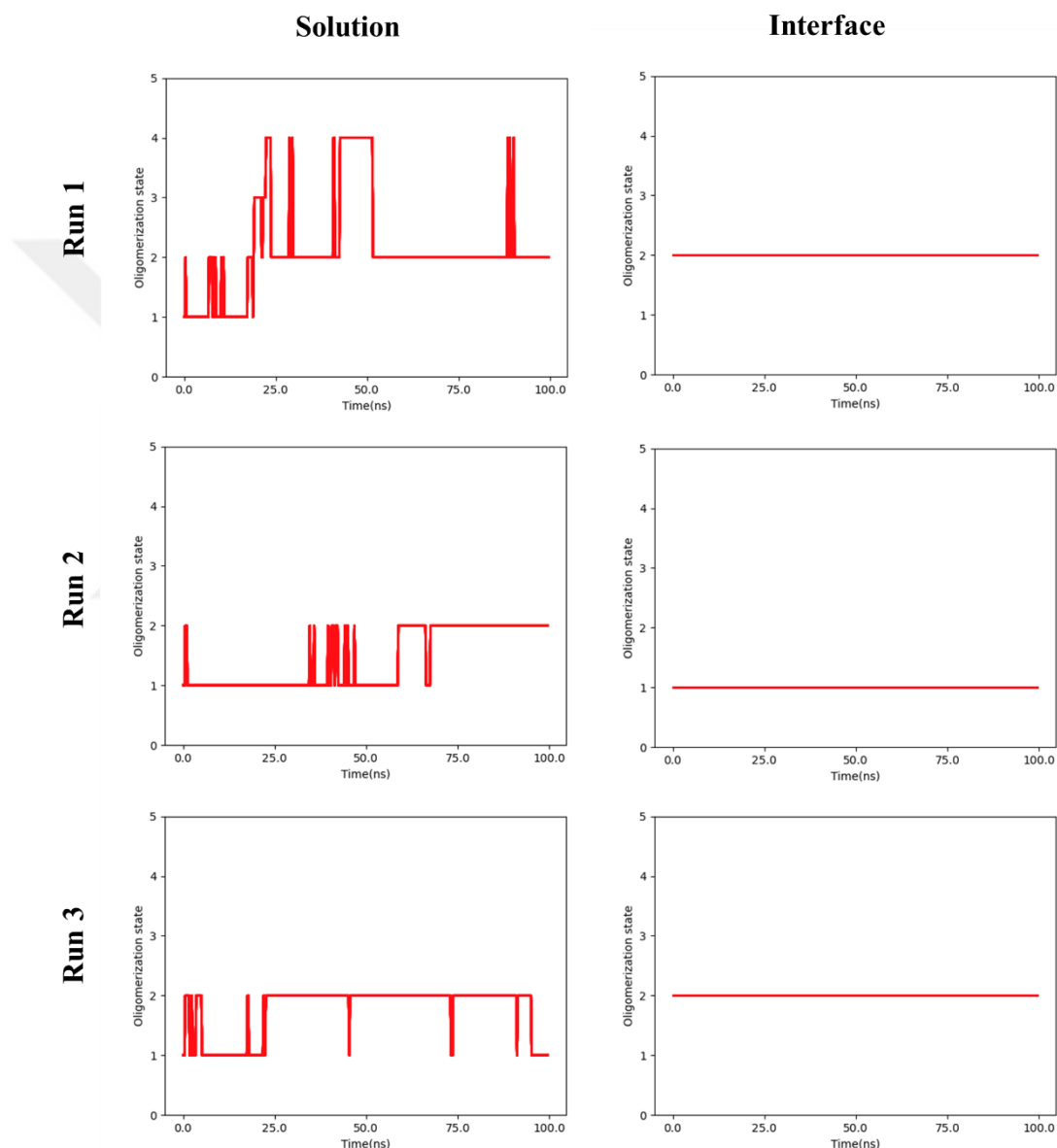


Figure 51. Cluster analysis of Orientation 4 simulations. Monomers in a 4 Å distance cutoff assumed as clustered in the simulation systems and the cluster lifetime was measured throughout the trajectories for both simulation environments.

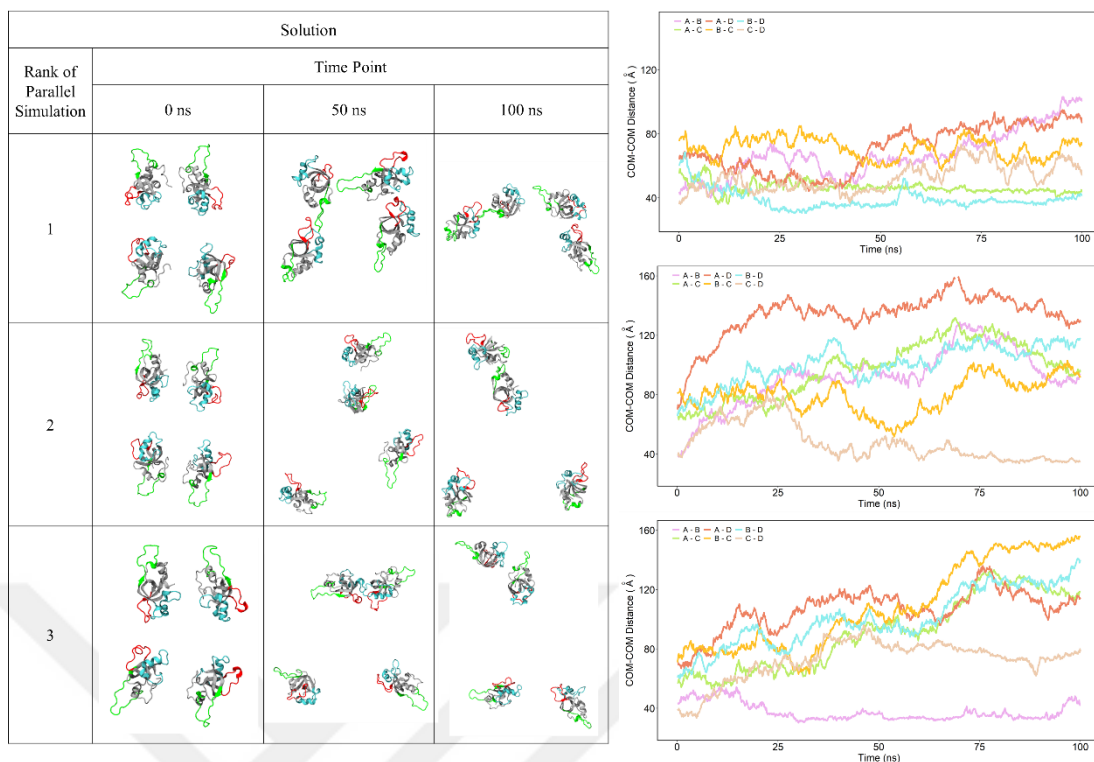


Figure 52. Snapshots of conformations at three different time points and COM–COM distances between chains of Orientation 4 solution simulations. Snapshots were taken at the beginning, middle, and end of the trajectory. L1, L2, and L3 loops and non-loop regions were denoted in cyan, green, red, and gray cartoon representations, respectively. The COM–COM distances between chains A and B, A and C, A and D, B and C, B and D, and C and D were shown as purple, green, orange, yellow, cyan, and beige lines, respectively.

At the air/water interface, early chain–chain interactions of conserved clusters were observed (Figure 51). The relative positions of aligned β -barrels of chains A and B were stable compared to chains C and D in the second cluster. Although the chain distance was strictly stable around 30 Å, while chain A in rotation, chain B at the upper interface was contacting with the air via its L2 and L3 loops. Only L2 loops of chains C and D were at the lower interface in this run. A similar clustering behavior was observed in the last parallel run. The central β -barrels of chains A and B displayed a stacking behavior and L2 loops were pointing towards the same direction after the 30 ns. In the second cluster, β -barrels were in proximity of each other. However, loops were rotating by conserving the intermolecular distance. The air interaction of the chains was highly similar to the previous observations (Figure 53).

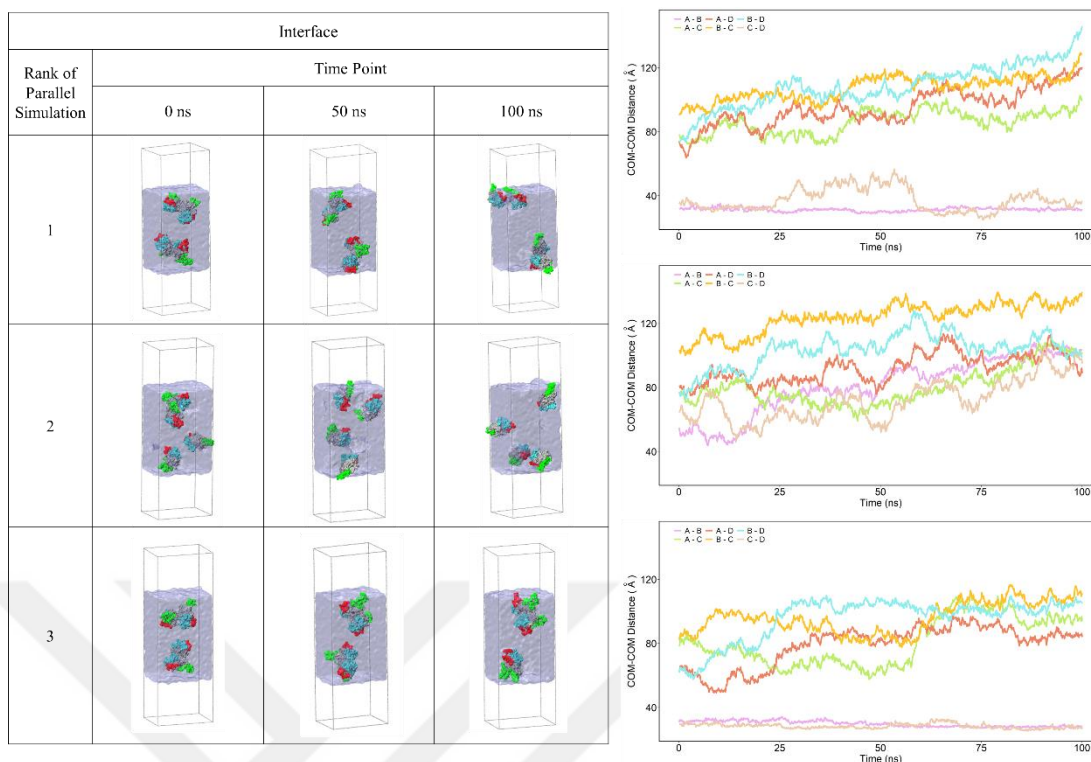


Figure 53. Snapshots of conformations at three different time points and COM–COM distances between chains of Orientation 4 interface simulations. Snapshots were taken at the beginning, middle, and end of the trajectory. L1, L2, and L3 loops and non-loop regions were denoted in cyan, green, red, and gray surface representations, respectively. The COM–COM distances between chains A & B, A & C, A & D, B & C, B & D and C & D were shown as purple, green, orange, yellow, cyan, and beige lines, respectively.

The dominant chain–chain interaction regions varied between simulation environments. L2 and L3 loops had slightly higher rates of intermolecular interactions with a number of atomic contacts compared to other regions in solution runs (Figure 54). On the other hand, N-termini were highly involved in the interactions with a higher number of atomic contacts followed by L1 loop in the interface clusters. The variety of the interaction regions was far more than the interface in solution clusters (Figure 55). Intermolecular interactions via loop-loop or loop-N terminal region were highly probable and established mostly through L2 loops in solution and only between N-termini at the interface monomer interactions (Appendix 10). The interaction regions were narrow but lasted longer in interface clusters. Conserved L2-L2 interactions were detected only in the last interface simulation run (Figure 55).

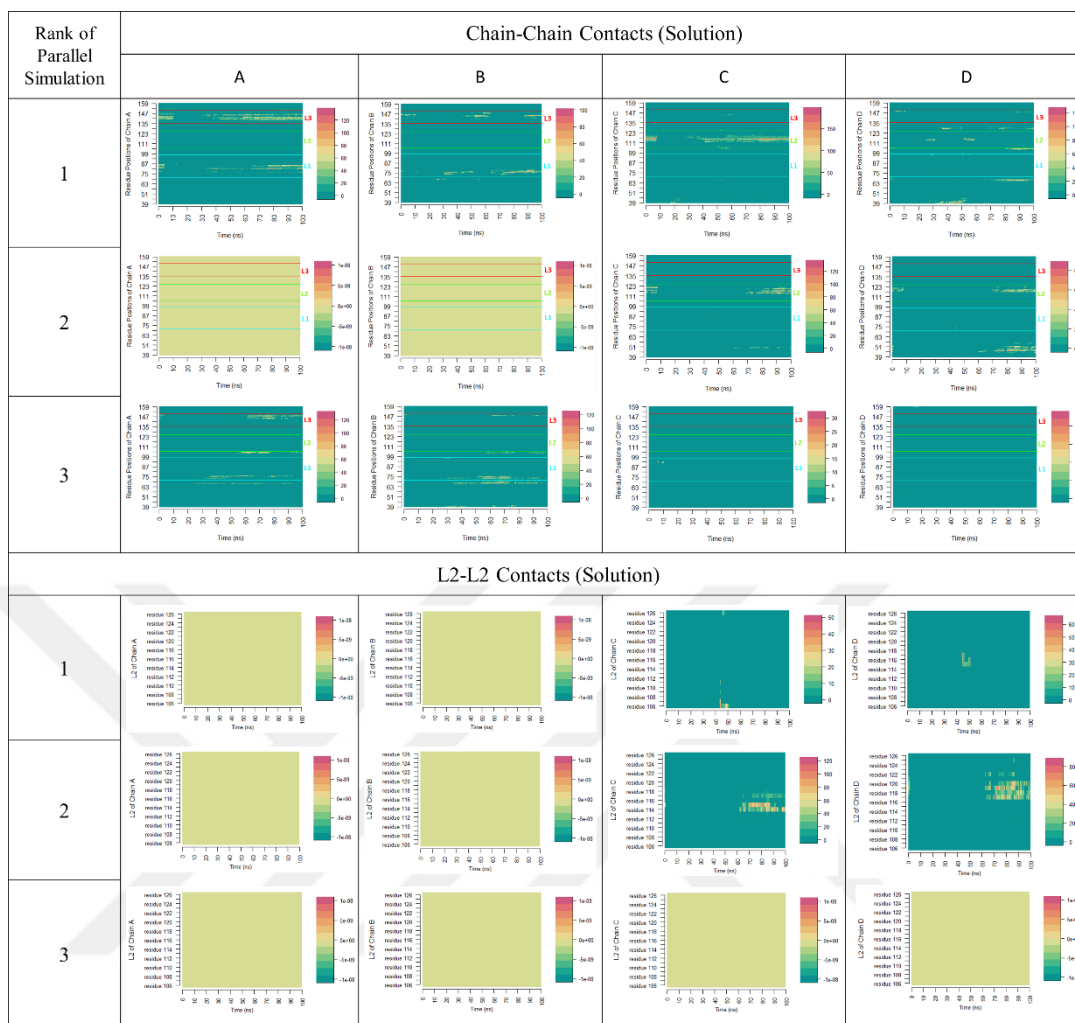


Figure 54. Contact maps of Orientation 4 solution simulations. Each map shows the interaction of the chain column with the remaining chains of the simulation system. L1, L2, and L3 loops were highlighted on the maps. L2-L2 interactions were measured by selecting only L2 loops. Light yellow background on the maps indicates no atomic interactions measured. The number of atomic contacts was assigned to a color gradient from dark green to red. The distance cutoff was set to 4 Å.

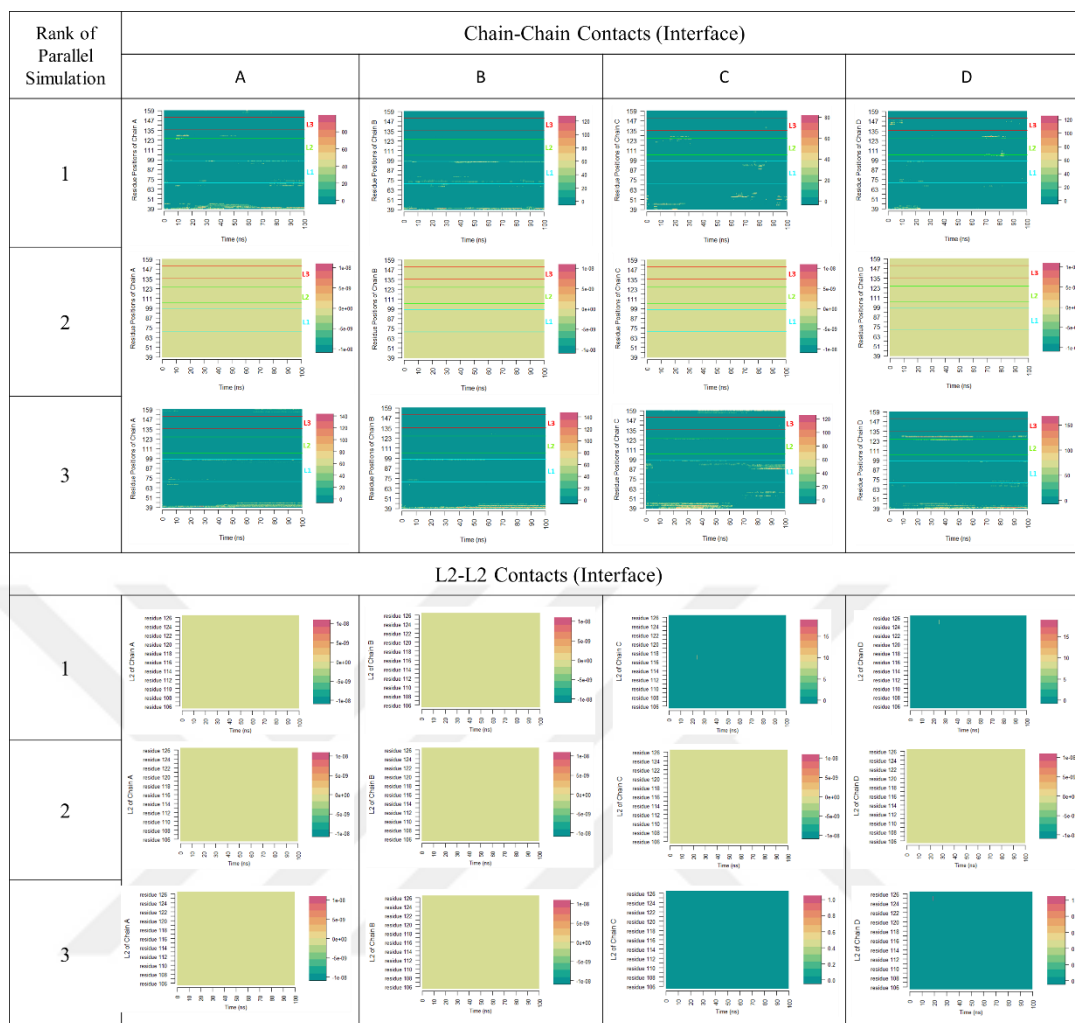


Figure 55. Contact maps of Orientation 4 interface simulations. Each map shows the interaction of the chain column with the remaining chains of the simulation system. L1, L2, and L3 loops were highlighted on the maps. L2-L2 interactions were measured by selecting only L2 loops. Light yellow background on the maps indicates no atomic interactions measured. The number of atomic contacts was assigned to a color gradient from dark green to red. The distance cutoff was set to 4 Å.

Hydrophobic interactions were formed through the loop regions; L2 was the dominant loop in solution simulations. Both homogenous and heterogenous hydrophobic loop contacts were observed in solution runs: L1-L2 and L3-L2 in the first and L2-L2 in the last run. At interface simulations, hydrophobic interactions were observed between N- and C-termini in the last parallel run (Figure 56).

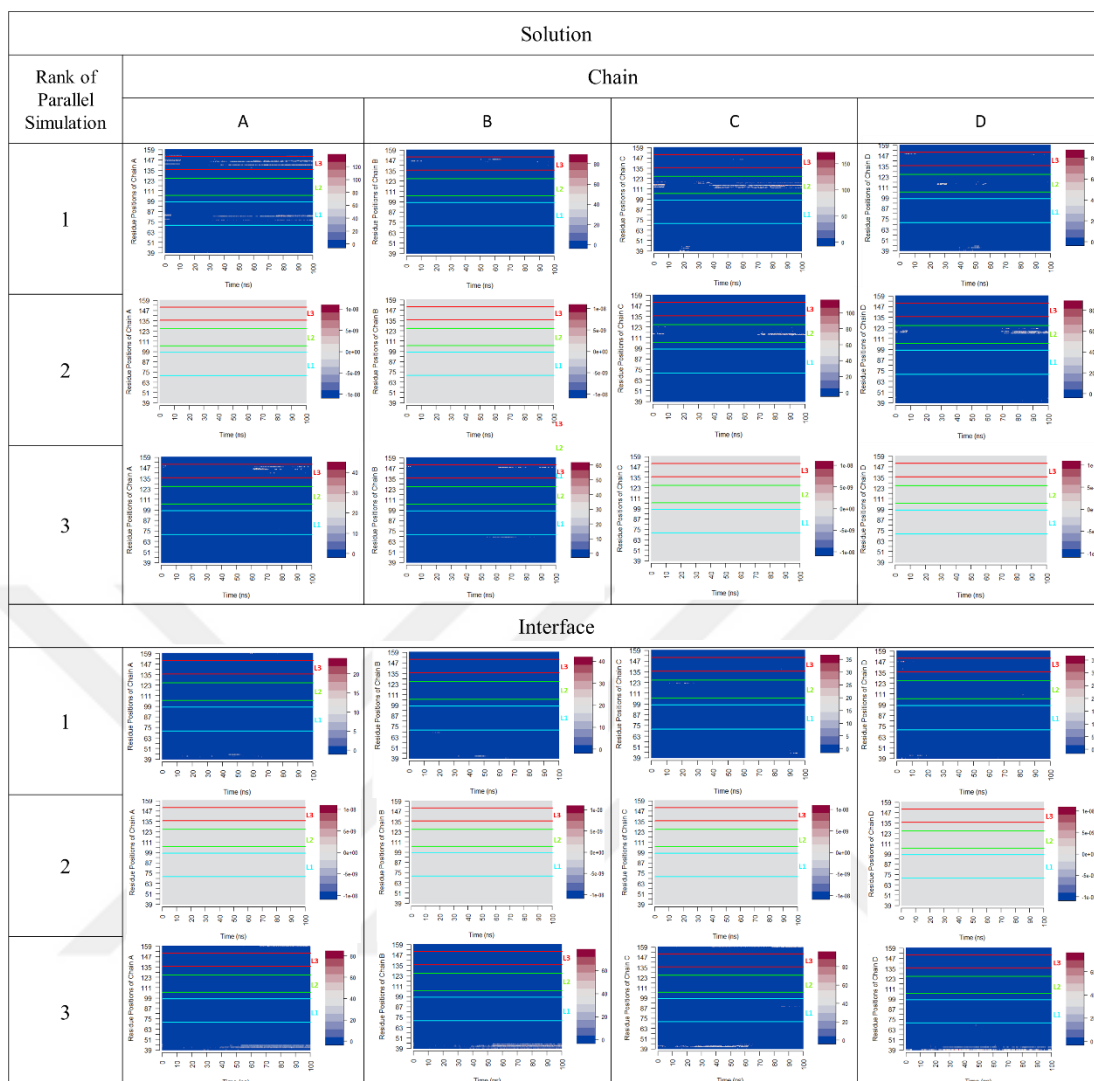


Figure 56. Hydrophobic contact maps of both Orientation 4 solution and interface simulations. Each map shows the interaction of the chain column with the remaining chains of the simulation system. L1, L2, and L3 loops were highlighted on the maps. The number of atomic contacts was assigned to a color gradient from dark blue to red. The distance cutoff was set to 4 Å.

Hydrogen bonds were not conserved in Orientation 4 solution simulations. In the first solution run, hydrogen bonds formed after 20 ns in both clusters but conserved only in one of the clusters. Clusters formed at interface had a slightly higher number of hydrogen bonds in total. The hydrogen bonds formed between the monomers of one of the clusters in the first (chain pair A-B) and last parallel run (chain pair C-D) were conserved in the last 50 ns (Figure 57).

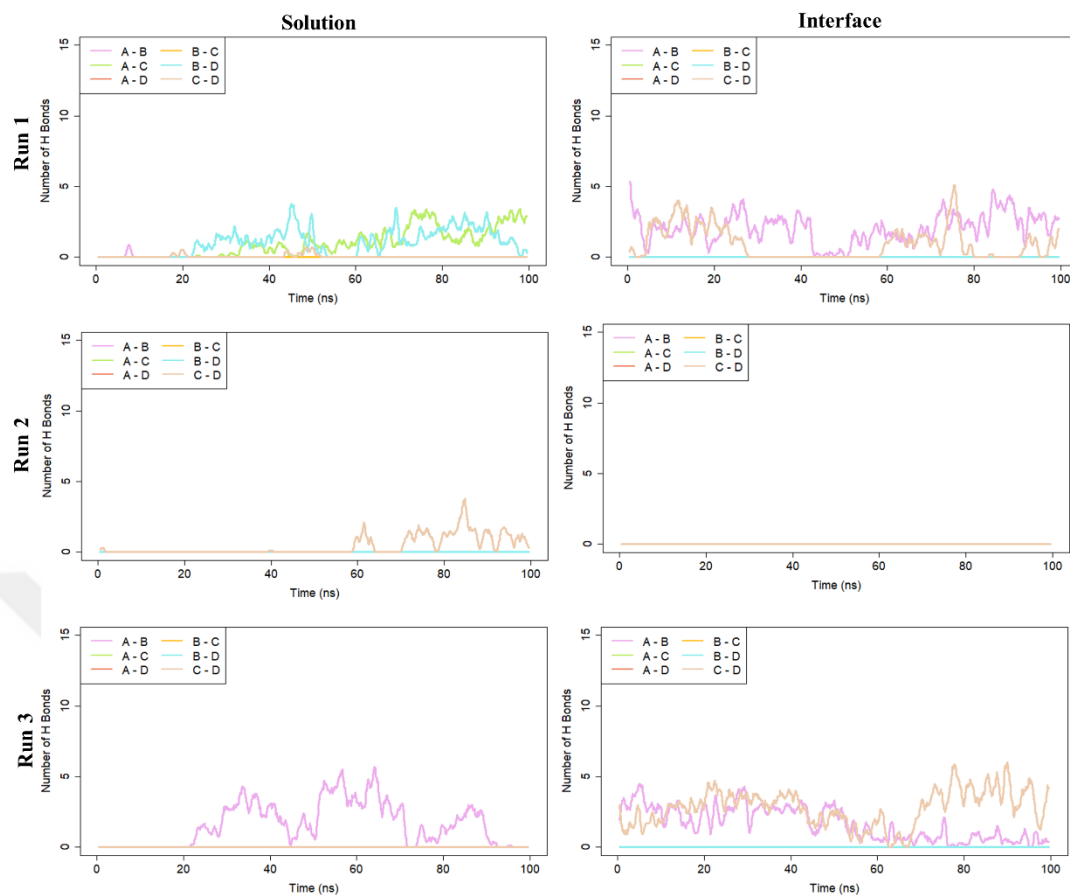


Figure 57. Measurements of the number of hydrogen bonds between chain pairs in Orientation 4 simulation systems in both environments. The chain pairs A & B, A & C, A & D, B & C, B & D, and C & D were shown as purple, green, orange, yellow, cyan, and beige lines in the graphs, respectively. Donor-acceptor distance and angle cutoff were set 3.5 Å and 30° respectively.

Potential salt bridges that might stabilize the clusters were also checked. However, there were no conserved interactions; all salt bridges formed between chains were transient. There was no difference observed between the stabilities of the monomers in solution and at interface simulations; the difference was observable in the level of one or two monomers. In solution, only a few monomers had slightly higher flexibility than monomers at the interface. Loops were mostly mobile at interface simulations due to intermolecular interactions through non-loop regions. Since there were loop interactions between solution monomers, RMSD values were lower for those monomers compared to interface (Figure 58).

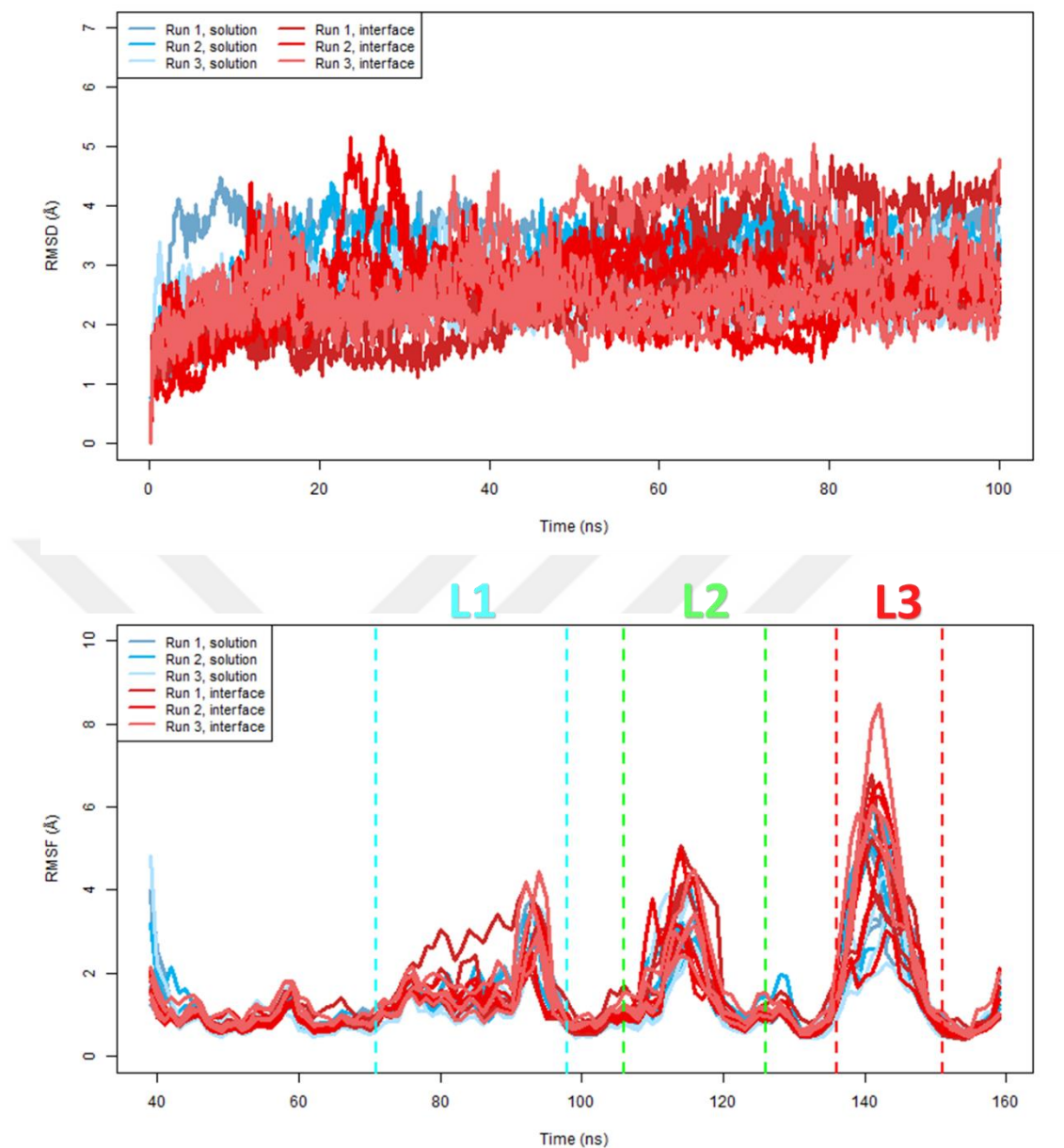


Figure 58. RMSD and RMSF measurements of individual chains in Orientation 4 simulation systems in both environments. Blue and red color gradients (from light to dark) were used to differentiate parallel solution and interface simulations, respectively. Regions between dashed cyan, green, and red lines show L1, L2, and L3 loops, respectively.

4.3.2 Simulation systems with positionally RWM

4.3.2.1 Orientation 1

Clusters of four monomers formed in most of the simulations in both solution and air/water interface (Figure 59). In the final compact cluster, chain A was the central monomer. After final rearrangement at 80 ns, chains formed a compact cluster in the first solution simulation (Figure 60). An L-shaped cluster of four monomers was observed in the last 20 ns, although the cluster formed in the earlier time steps of the second solution run. A-D and C-D distances were stable with small fluctuations over the trajectory and in the last 40 ns, respectively. In the last run, after 60 ns, monomers were aligned in the same way as in the second run by forming an L-shaped cluster. In this case, monomer pairs were interacting along their L2 loops. However, decreased fluctuations were observed, although the average distances were stable (Figure 60).

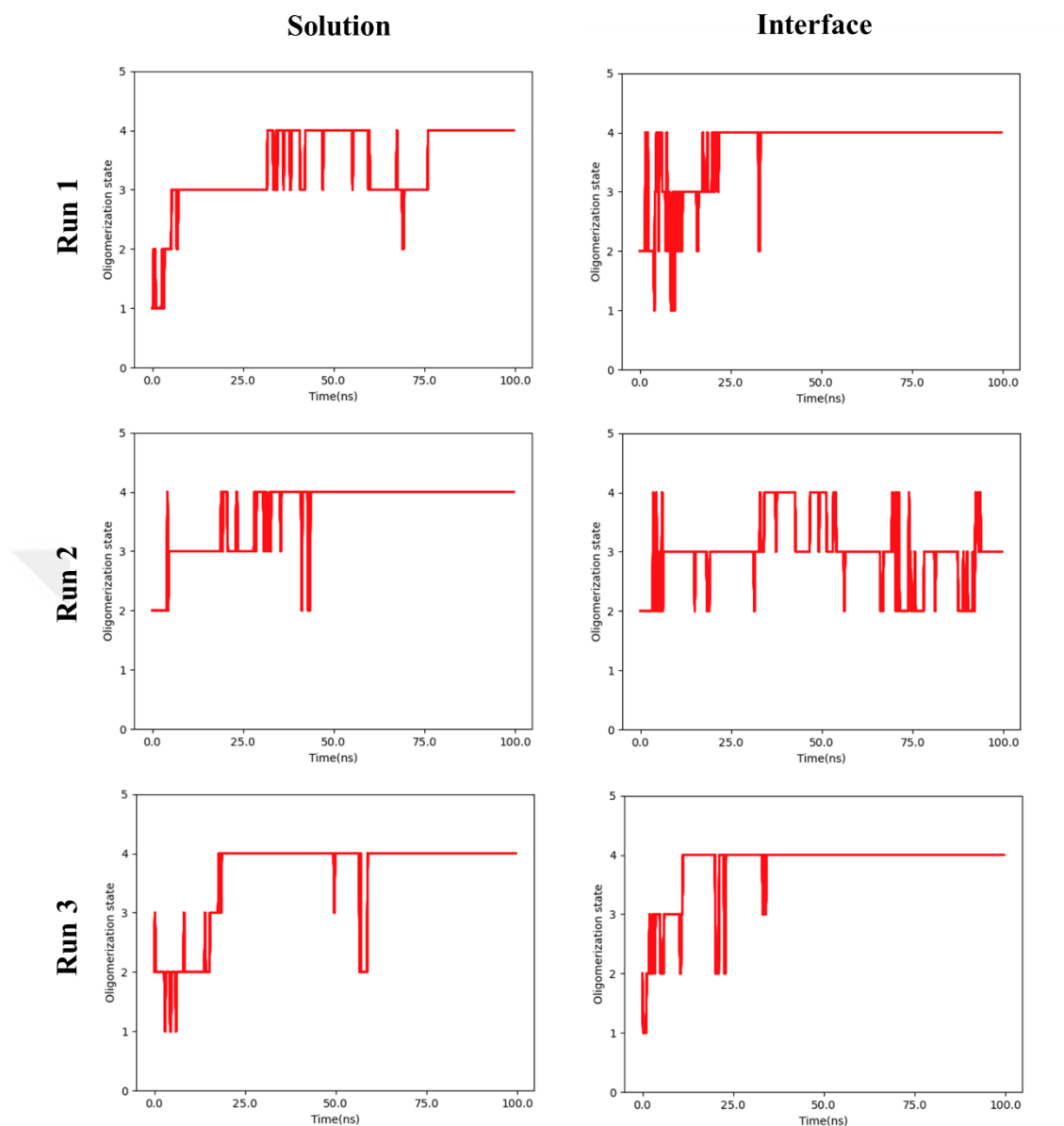


Figure 59. Cluster analysis of Orientation 1 simulation systems with positionally RWM. Monomers in a 4 Å distance cutoff assumed as clustered in the simulation systems and the cluster lifetime was measured throughout the trajectories for both simulation environments.

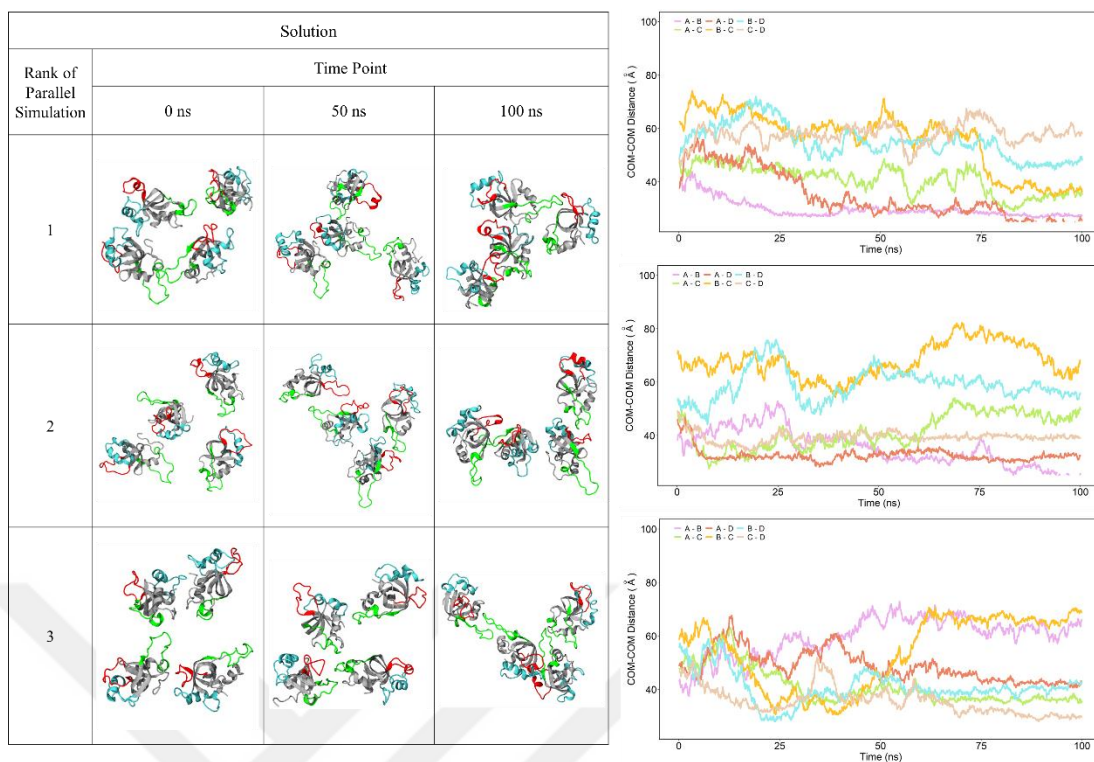


Figure 60. Snapshots of conformations at three different time points and COM-COM distances between chains of Orientation 1 solution simulation systems with positionally RWM. Snapshots were taken at the beginning, middle, and end of the trajectory. L1, L2, and L3 loops and non-loop regions were denoted in cyan, green, red, and gray cartoon representations, respectively. The COM-COM distances between chains A & B, A & C, A & D, B & C, B & D and C & D were shown as purple, green, orange, yellow, cyan, and beige lines, respectively.

In the first half of the first interface simulation, clusters of four monomers were formed (Figure 59). However, the final compact assembly formed in the last 15 ns. Both L2 and L3 loops of one monomer and only L2 loop of the second monomer interacted with the interface. In the second run, constant interactions were observed in only one chain pair (chains A and C). However, the intermolecular distance was not stabilized. Air interaction was observed in only one of the monomers through both its L2 and L3 loops; others clustered inside the solution. As in the second interface run, chain A was in contact with the air through its L3 loop; others formed a planar cluster submerged in the solution in the last interface run. The largest and most fluctuating distance was measured between chains A and D since they were not in interaction. At the interface, chains A and C were interacting through their L2 loops with longer interaction durations in each parallel run (Figure 61).

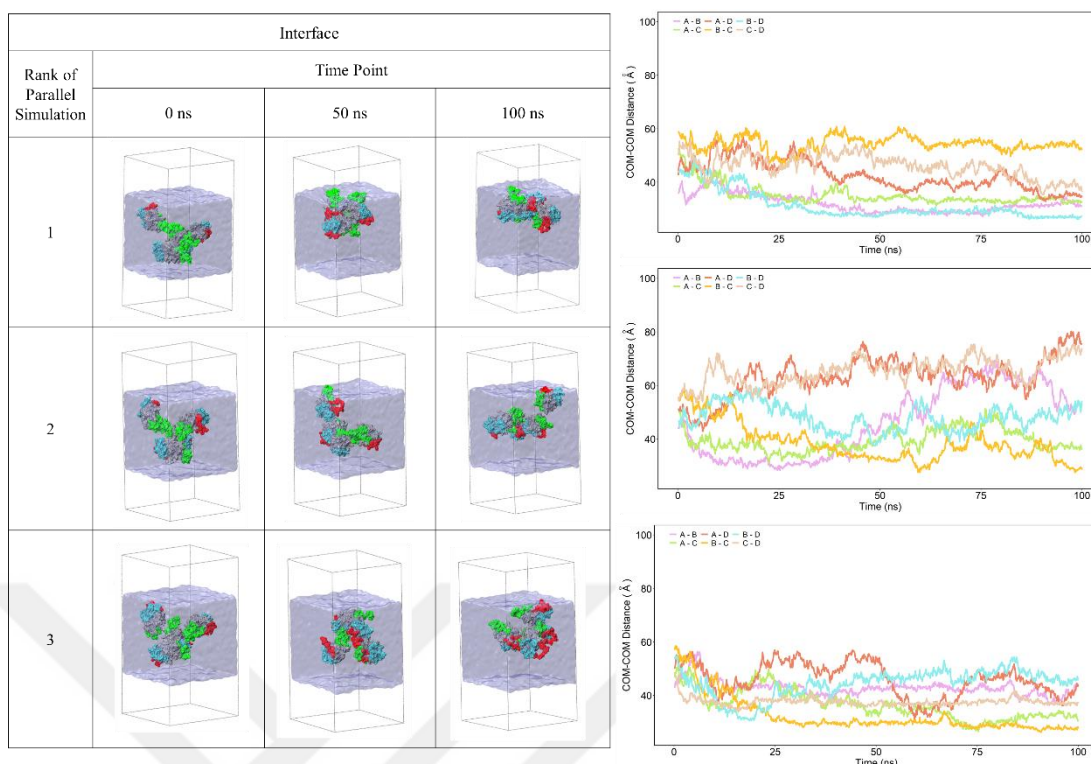


Figure 61. Snapshots of conformations at three different time points and COM–COM distances between chains of Orientation 1 interface simulation systems with positionally RWM. Snapshots were taken at the beginning, middle, and end of the trajectory. L1, L2, and L3 loops and non-loop regions were denoted in cyan, green, red, and gray surface representations, respectively. The COM–COM distances between chains A & B, A & C, A & D, B & C, B & D and C & D were shown as purple, green, orange, yellow, cyan, and beige lines, respectively.

Both N-termini and L2 loops were hotspots in the intermolecular interactions in both solution and interface simulations with the dominance of L2 in interface clusters (Appendix 11). Inter-loop region between L2 and L3 was highly involved in the conserved interactions and contributed to the cluster stability both in solution and at interface simulations. L3 loop was the following preferred interaction site at the interface simulations unlike L1 loop in the solution. In solution runs, monomers were interacting transiently through larger regions (Figure 62). In the case of interface, interactions regions were more stable and formed high number of atomic contacts. Although conserved L2-L2 contacts were observed in both environments, they were more pronounced in the interface clusters since they were observed in all runs (Figure 63).

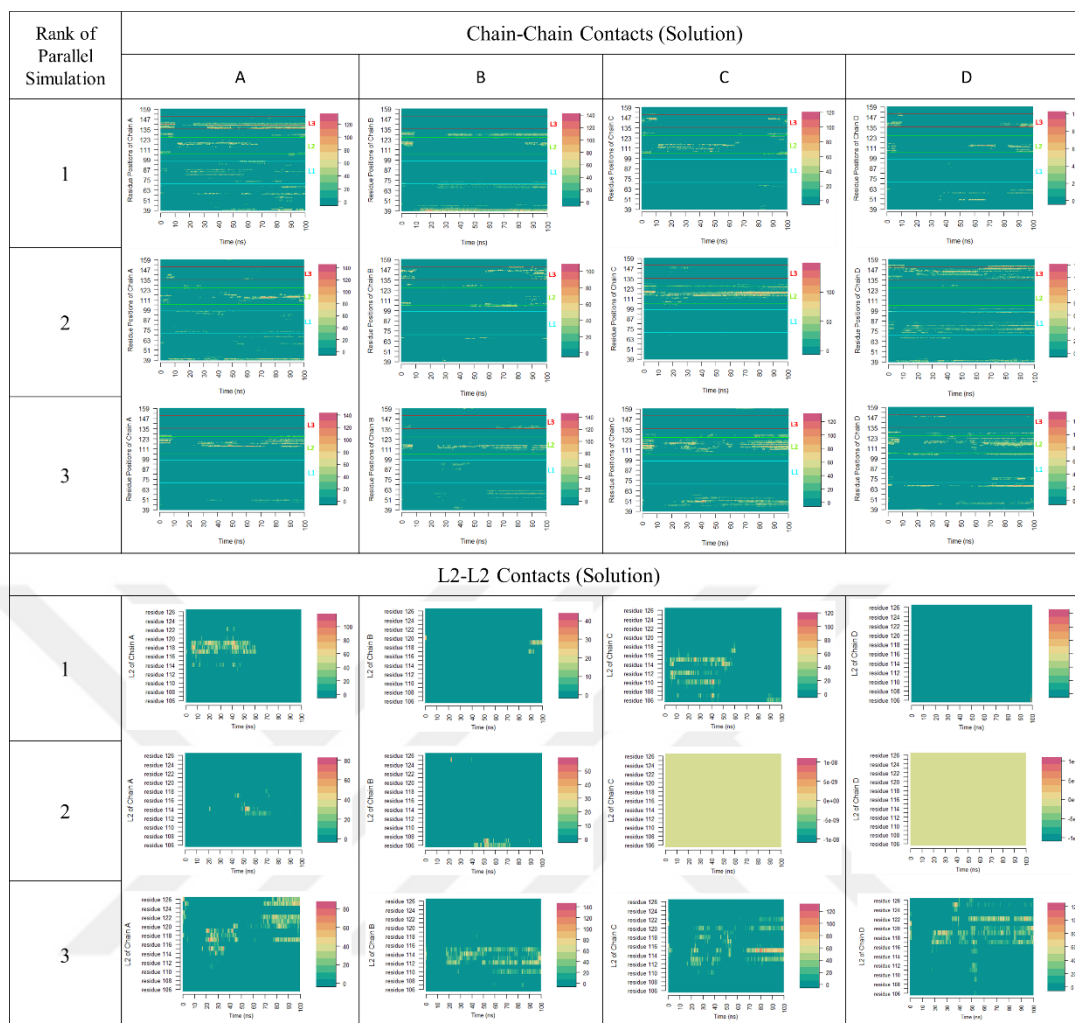


Figure 62. Contact maps of Orientation 1 solution simulation systems with positionally RWM. Each map shows the interaction of the chain column with the remaining chains of the simulation system. L1, L2 and L3 loops were highlighted on the maps. L2-L2 interactions were measured by selecting only L2 loops. Light yellow background on the maps indicates no atomic interactions measured. The number of atomic contacts was assigned to a color gradient from dark green to red. The distance cutoff was set to 4 Å.

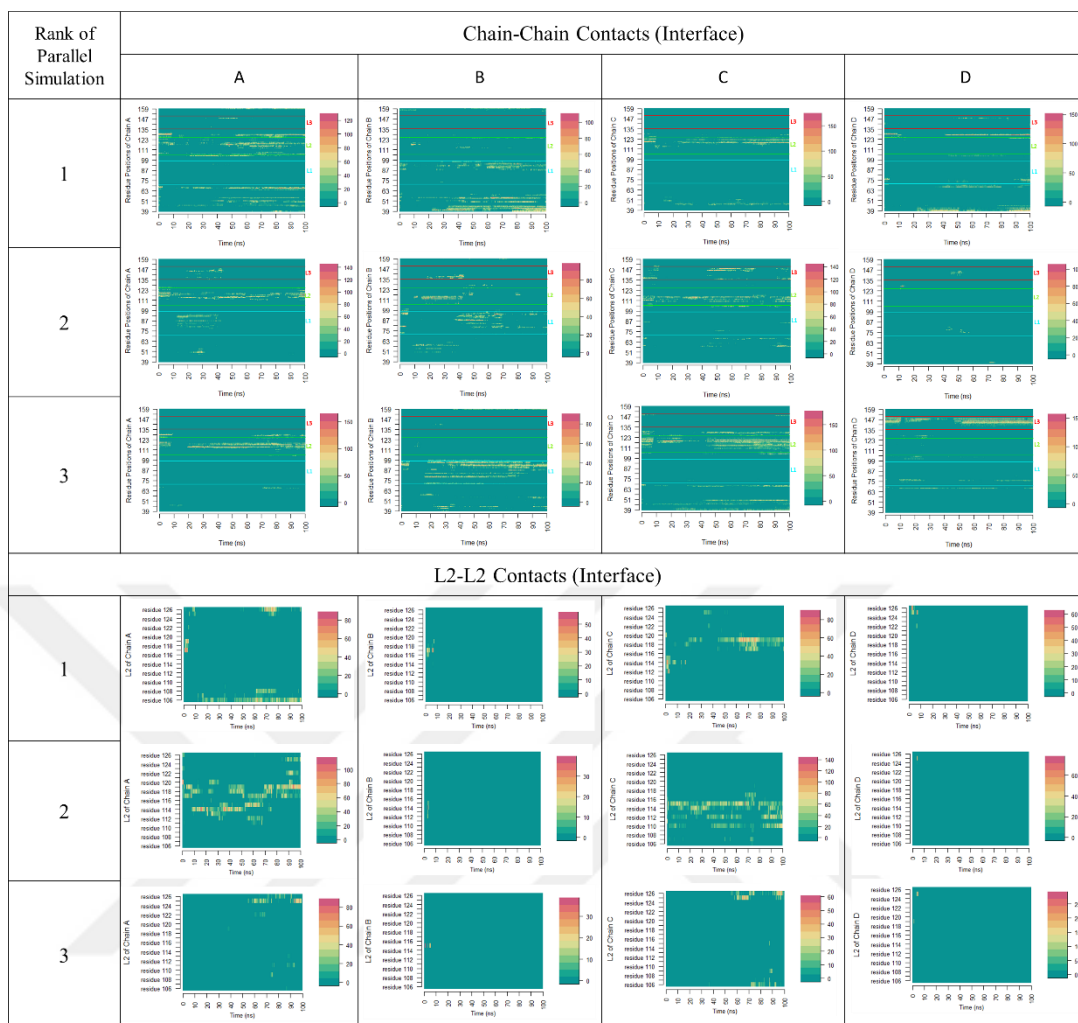


Figure 63. Contact maps of Orientation 1 interface simulation systems with positionally RWM. Each map shows the interaction of the chain column with the remaining chains of the simulation system. L1, L2 and L3 loops were highlighted on the maps. L2-L2 interactions were measured by selecting only L2 loops. Light yellow background on the maps indicates no atomic interactions measured. The number of atomic contacts was assigned to a color gradient from dark green to red. The distance cutoff was set to 4 Å.

Conserved hydrophobic contacts concentrated in L2 loops of monomers in both simulation environments. N-termini also had a small contribution to these interactions at air/water interface clusters (Figure 64). Cluster stabilities were supported with salt bridges and hydrogen bonds. Charged residues in loop regions formed heterogeneous loop-loop (residues in L2 and L3) and N termini-loop interactions. At the interface, homogeneous N-termini salt bridges were observed. Most of the bridges were formed in the late time steps of the simulations (Figure 65). Hydrogen bonds were observed between chains in the early time points in both simulation environments and conserved

in most of the runs. The number of bonds did not exceed five in most chain pairs and mostly increased in the second half of the simulations, especially in the last 15-20 ns (Figure 66).

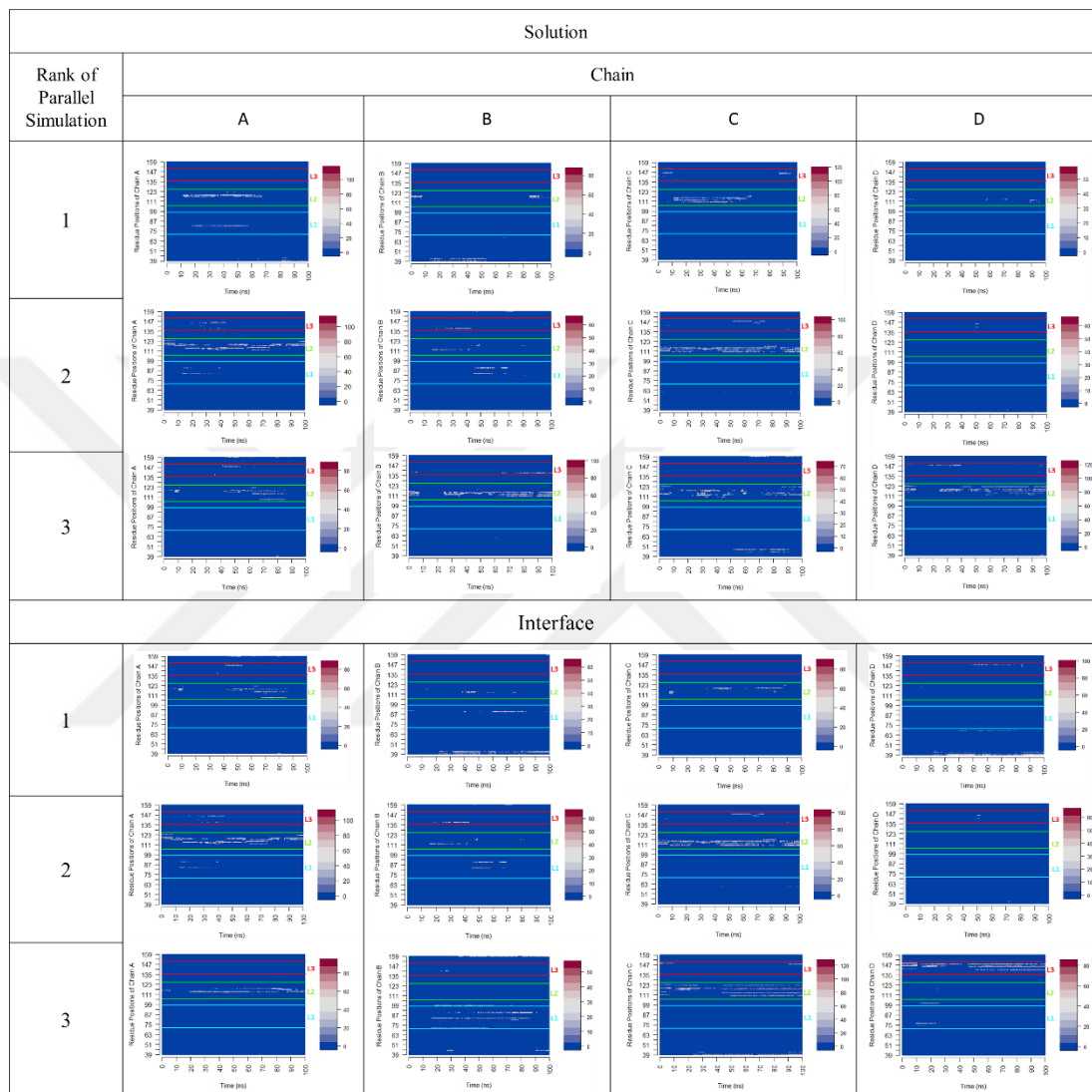


Figure 64. Hydrophobic contact maps of both Orientation 1 solution and interface simulation systems with positionally RWM. Each map shows the interaction of the chain column with the remaining chains of the simulation system. L1, L2, and L3 loops were highlighted on the maps. The number of atomic contacts was assigned to a color gradient from dark blue to red. The distance cutoff was set to 4 Å.

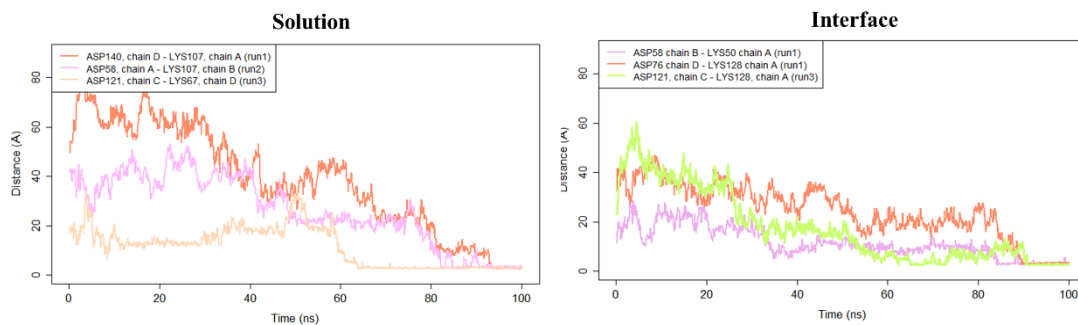


Figure 65. Measurements of conserved interchain salt bridges throughout the trajectories in Orientation 1 simulation systems with positionally RWM for both environments. The distance between residues that form bridges was measured in the time course of the simulations. The chain pairs A & B, A & C, A & D, and C & D were shown as purple, green, orange, and beige lines in the graphs, respectively. Oxygen-nitrogen distance cutoff was set to 3.2 Å. However, salt bridge distance up to 4 Å was acceptable as conserved interaction.

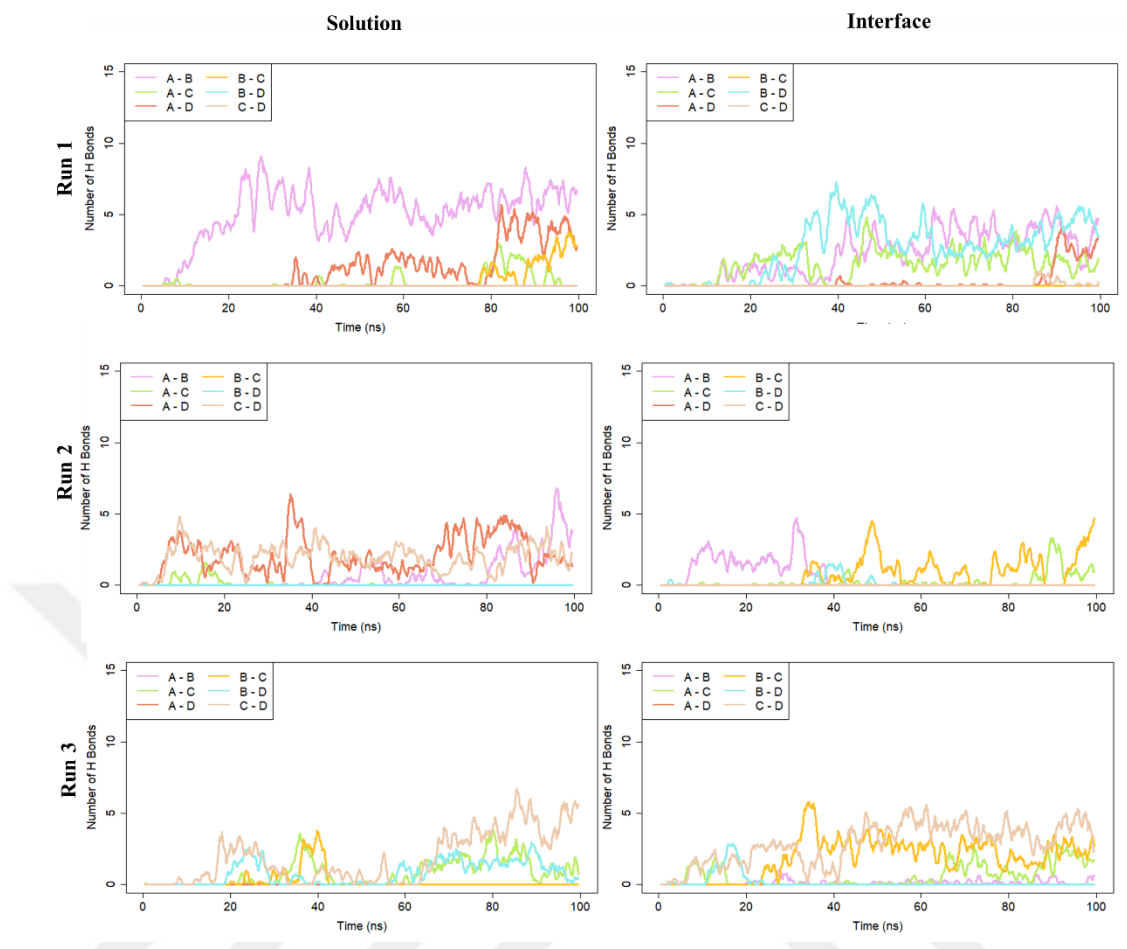


Figure 66. Measurements of the number of hydrogen bonds between chain pairs in Orientation 1 simulation systems with positionally RWM in both environments. The chain pairs A & B, A & C, A & D, B & C, B & D, and C & D were shown in purple, green, orange, yellow, cyan, and beige lines in the graphs, respectively. Donor-acceptor distance and angle cutoff were set at 3.5 Å and 30° respectively.

Most of the monomers in the interface simulations were stable compared to the solution. Elevated RMSD values were measured for relatively free monomers at the interface. N-termini, L1 and L3 loops of solution monomers had higher fluctuations (Figure 67).

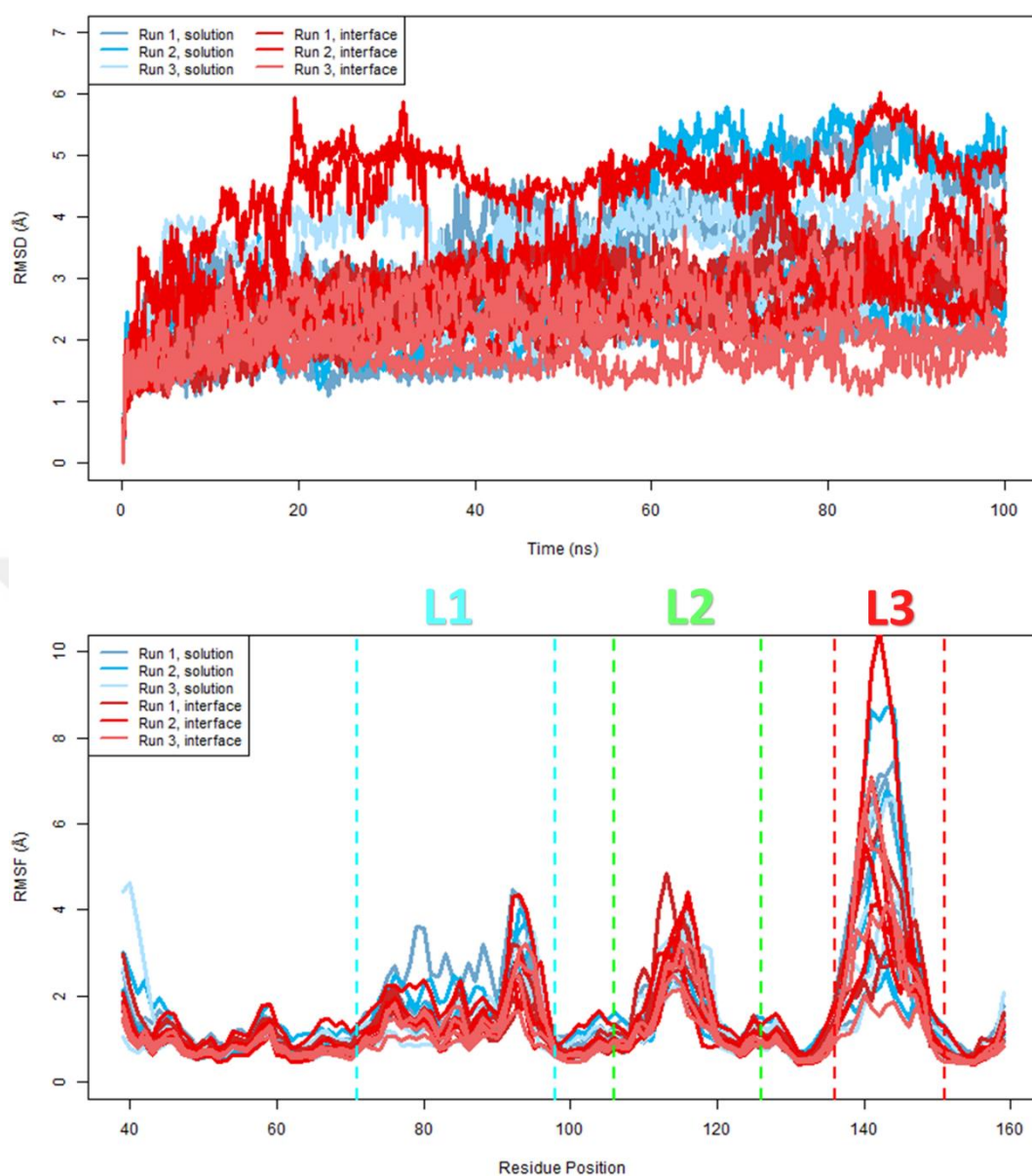


Figure 67. RMSD and RMSF measurements of chains in Orientation 1 simulations with positionally RWM in both environments. Blue and red color gradients (from light to dark) were used to differentiate parallel solution and interface simulations, respectively. Regions between dashed cyan, green, and red lines show L1, L2, and L3 loops, respectively.

4.3.2.2 Orientation 2

In the systems in which L2 loops were centered, interface simulations formed clusters that contain at least 3 monomers. Varied cluster sizes were observed in solution simulations (clusters of two, three and four monomers) (Figure 68). In the

first run, a cluster of two monomers was conserved in the last 10 ns. Other monomers were transiently interacting with each other. In the second parallel run, a cluster of four monomers formed and was conserved for 75 ns; three interacting chains were more stable than the cluster of four. The intermolecular distances experienced smaller fluctuations in the cluster compared to the first run. An L-shaped cluster of three was formed in the early steps of the last solution simulation trajectory. The distance between chain A and C was stable throughout the simulation and had the lowest COM–COM distance due to the closer central beta barrels of the monomers (Figure 69).



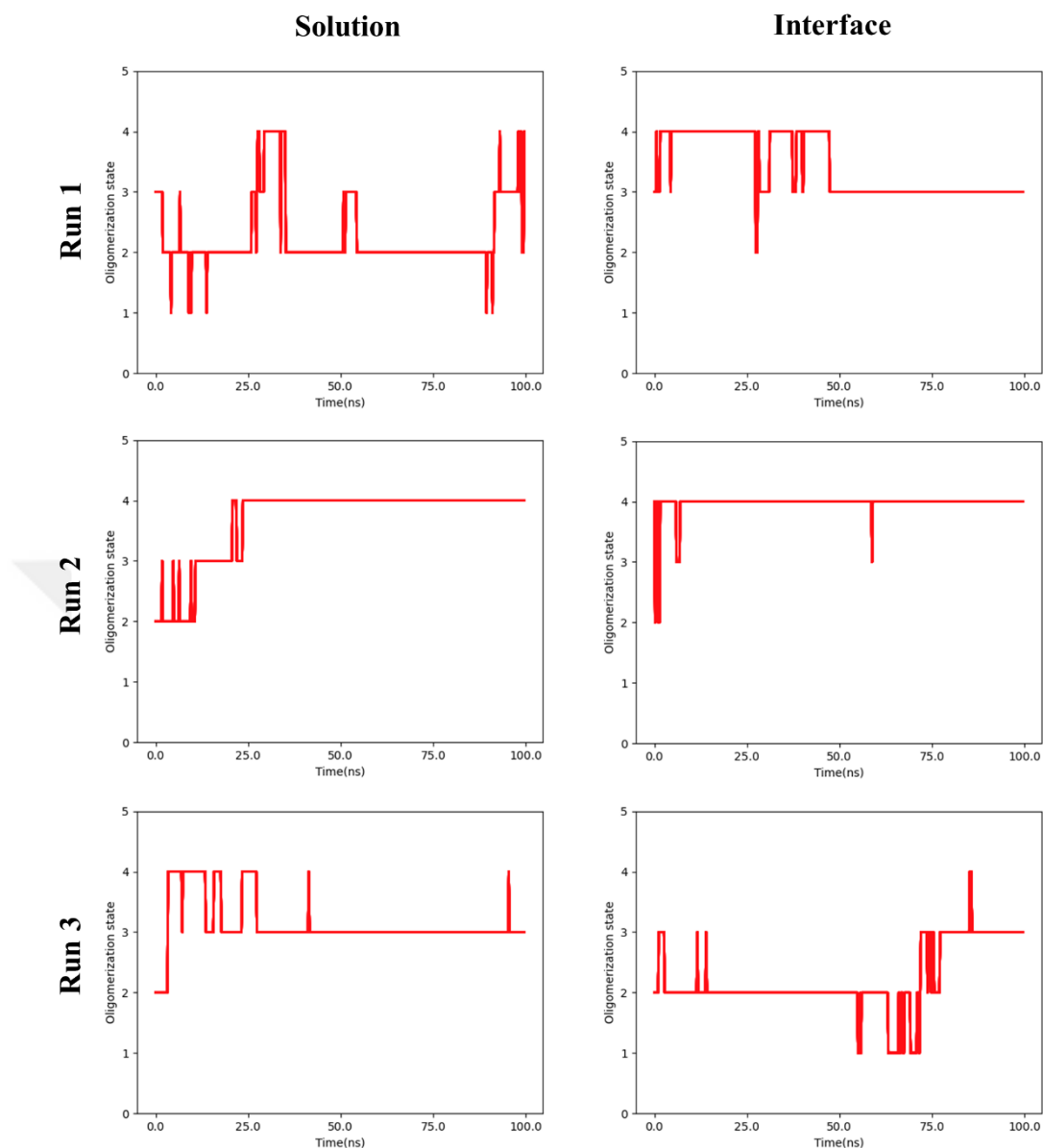


Figure 68. Cluster analysis of Orientation 2 simulation systems with positionally RWM. Monomers in a 4 Å distance cutoff assumed as clustered in the simulation systems and the cluster lifetime was measured throughout the trajectories for both simulation environments.

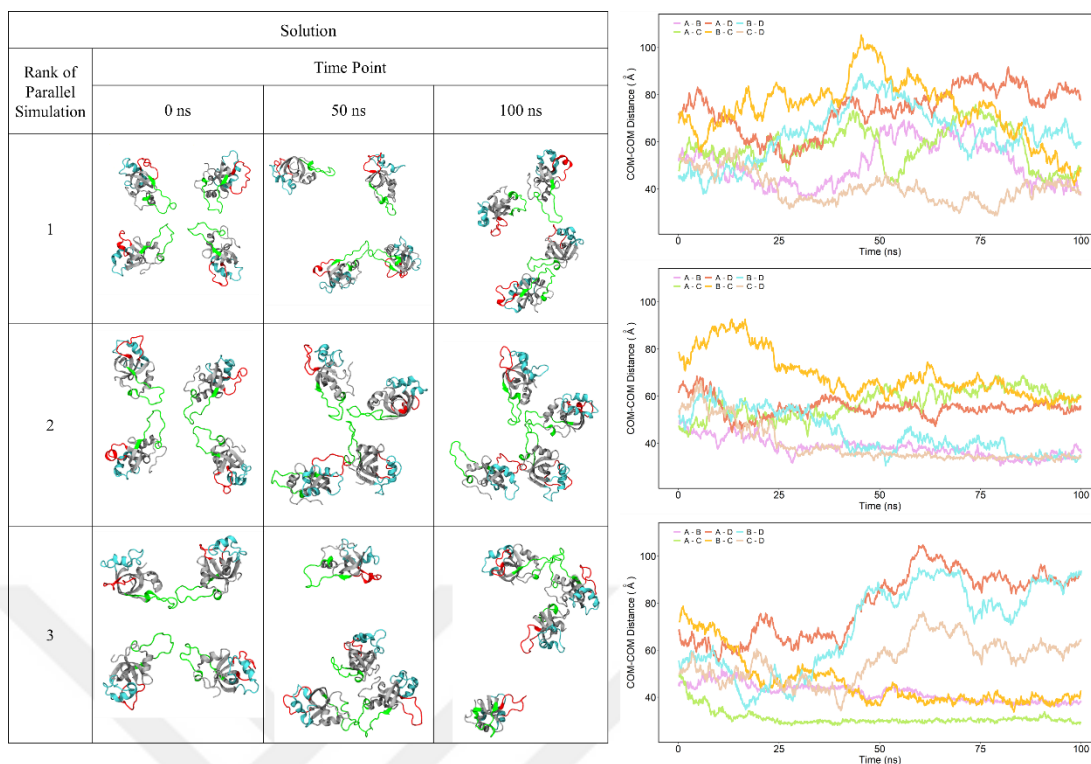


Figure 69. Snapshots of conformations at three different time points and COM–COM distances between chains of Orientation 2 solution simulation systems with positionally RWM. Snapshots were taken at the beginning, middle, and end of the trajectory. L1, L2, and L3 loops and non-loop regions were denoted in cyan, green, red, and gray cartoon representations, respectively. The COM–COM distances between chains A & B, A & C, A & D, B & C, B, & D and C & D were shown as purple, green, orange, yellow, cyan, and beige lines, respectively.

At the interface, monomers formed a planar cluster by facing centered L2 loops to the interface in the first run. Chains A and C formed a compact pair by stacking of both L2 and L3 loops; chain D was slightly distant from this pair next to chain A. Monomer distances were more stable in the last 40 ns. In the second run, three monomers were interacting from the tips of L2 loops which prevented the formation of compact cluster. B-D and A-C chain distances were shorter in this cluster. In the last run, a cluster of three monomers was observed in the later time steps. The intermolecular distances were stable in the last 20 ns due to the rotation of the monomers to face L2 loops interface. All monomers were at the upper air/water interface (Figure 70).

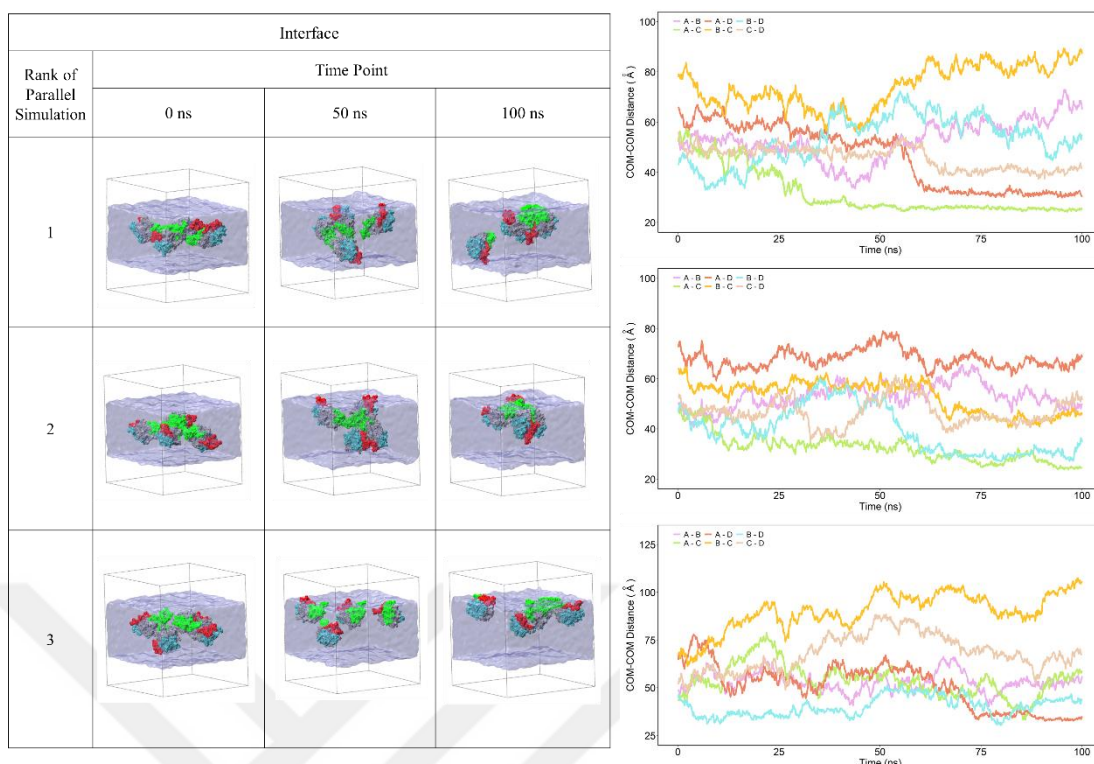


Figure 70. Snapshots of conformations at three different time points and COM–COM distances between chains of Orientation 2 interface simulation systems with positionally RWM. Snapshots were taken at the beginning, middle, and end of the trajectory. L1, L2, and L3 loops and non-loop regions were denoted in cyan, green, red, and gray surface representations, respectively. The COM–COM distances between chains A & B, A & C, A & D, B & C, B & D and C & D were shown as purple, green, orange, yellow, cyan, and beige lines, respectively.

Although L2 region was a key interaction site, the N-termini of the monomers also had a significant impact on intermolecular interactions in the clusters formed in both simulation environments. The number of atomic contacts was mostly higher in L2 loops. In the inter-chain regions (L1-L2, L2-L3), there were constantly interacting residues contributing to the stabilization, especially of the interface clusters (Figure 72). Although L2-L2 contacts were conserved for long time steps in both environments and each parallel run (Appendix 12), the number of monomers involved in L2-L2 interactions was higher at the interface with larger interaction regions (Figure 72). Varied sizes of sequence regions were observed in clusters in both simulation environments. However, larger chain–chain interfaces were observed in solution clusters (Figure 71). The total number of regions involved in the intermolecular

contacts was higher at the interface systems with an elevated number of atomic contacts (Figure 72).

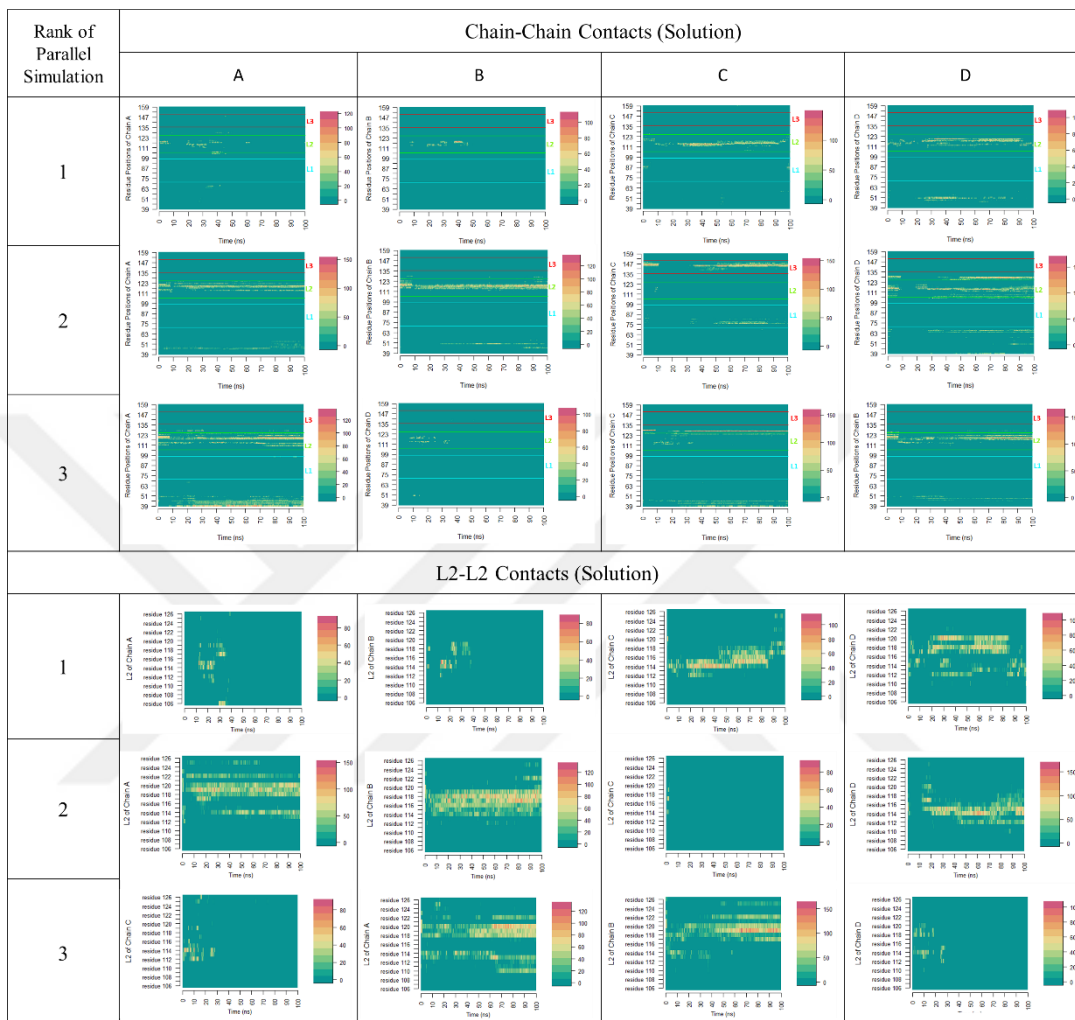


Figure 71. Contact maps of Orientation 2 solution simulation systems with positionally RWM. Each map shows the interaction of the chain column with the remaining chains of the simulation system. L1, L2, and L3 loops were highlighted on the maps. L2-L2 interactions were measured by selecting only L2 loops. Light yellow background on the maps indicates no atomic interactions measured. The number of atomic contacts was assigned to a color gradient from dark green to red. The distance cutoff set to 4 Å.

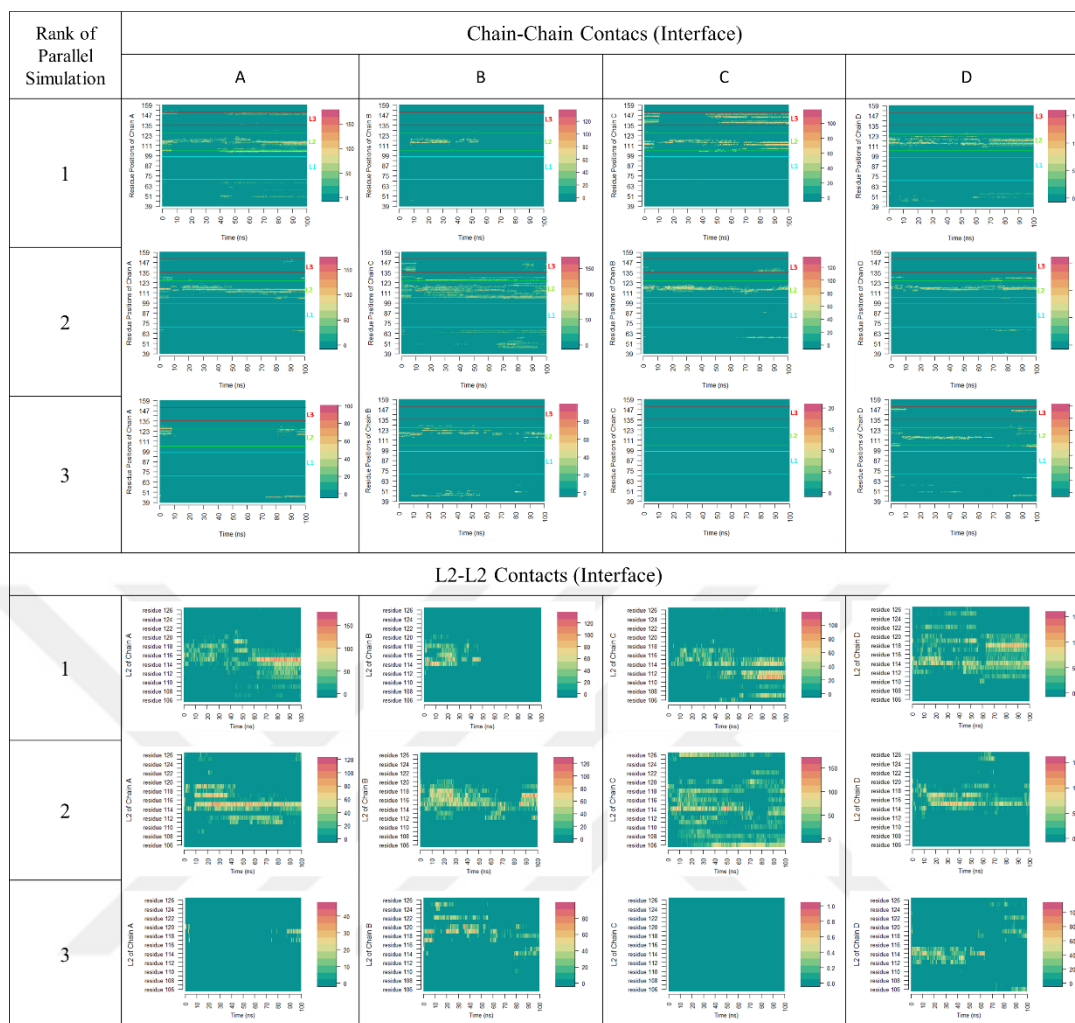


Figure 72. Contact maps of Orientation 2 interface simulation systems with positionally RWM. Each map shows the interaction of the chain column with the remaining chains of the simulation system. L1, L2, and L3 loops were highlighted on the maps. L2-L2 interactions were measured by selecting only L2 loops. Light yellow background on the maps indicates no atomic interactions measured. The number of atomic contacts was assigned to a color gradient from dark green to red. The distance cutoff set to 4 Å.

Hydrophobic residues contributed to cluster stability; hydrophobic interactions were conserved in L2 loops in both solution and interface clusters. The second most dominant hydrophobic interactions were at the N-termini and in L3 loop in solution and interface monomers, respectively (Figure 73). Despite large clusters formed, salt bridges were not a key interaction that contributed to the cluster stability. The interacting residues were mostly at the N-termini and only a few numbers of interactions formed in total (Figure 74). A small number of hydrogen bonds (≤ 5) formed between monomers in the early time steps and were more stable in the clusters

of solution simulations. At the interface, the number of hydrogen bonds increased in the last 30-40 ns (Figure 75).

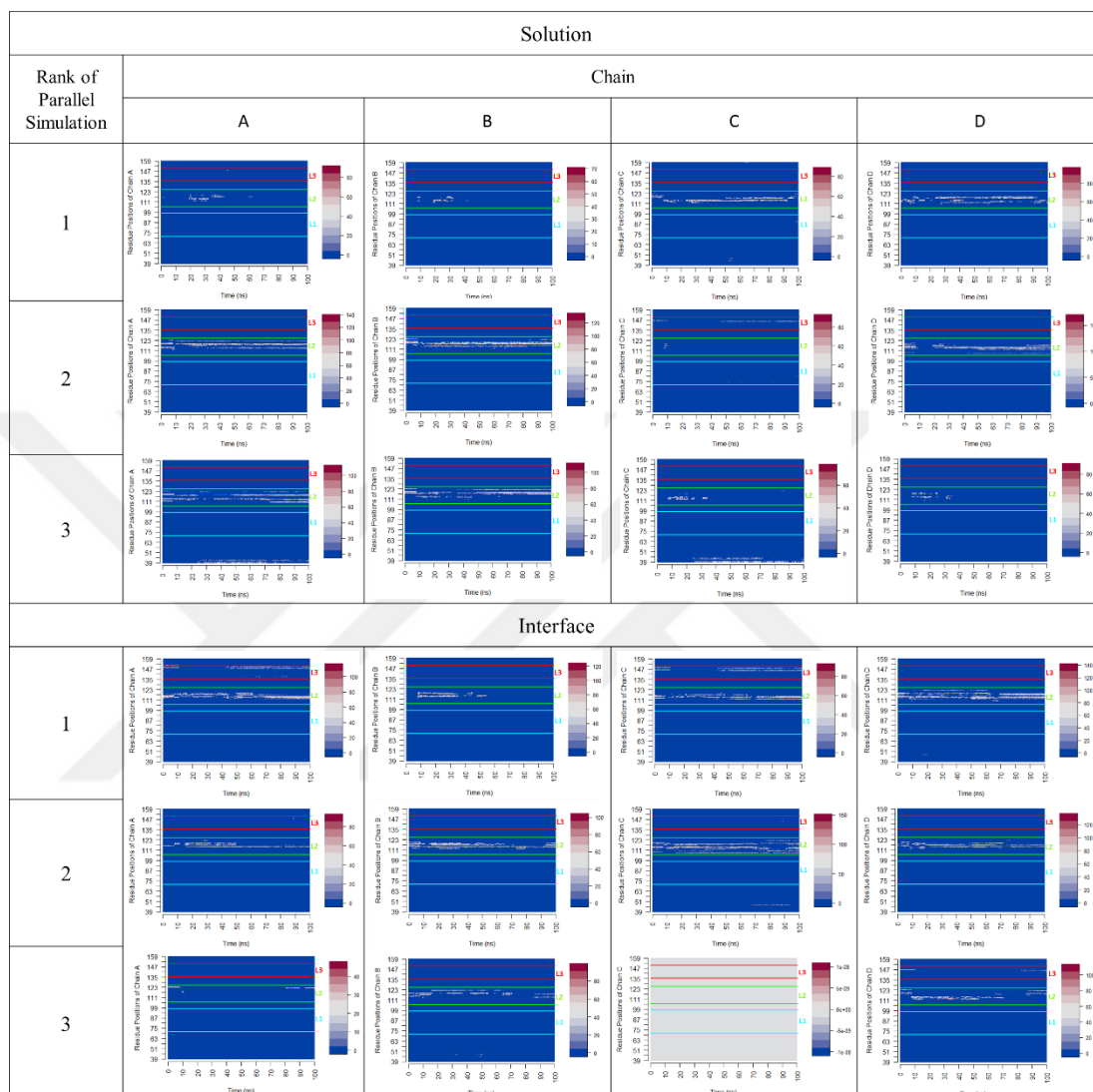


Figure 73. Hydrophobic contact maps of both Orientation 2 solution and interface simulation systems with positionally RWM. Each map shows the interaction of the chain column with the remaining chains of the simulation system. L1, L2, and L3 loops were highlighted on the maps. The number of atomic contacts was assigned to a color gradient from dark blue to red. The distance cutoff was set to 4 Å.

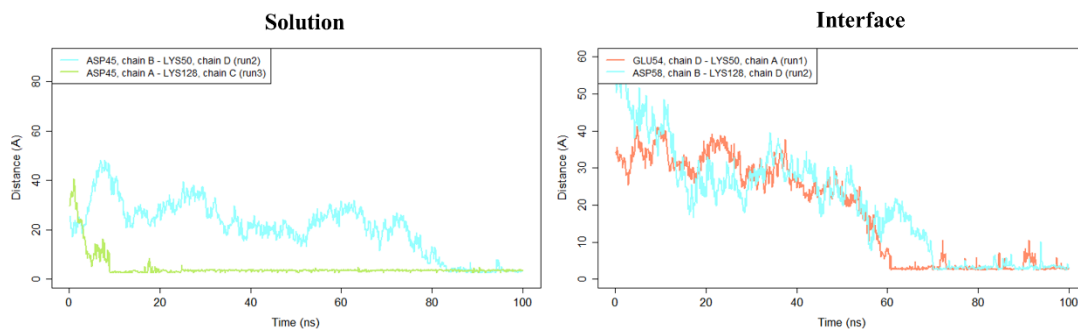


Figure 74. Measurements of conserved interchain salt bridges throughout the trajectories in Orientation 2 simulation systems with positionally RWM for both environments. The distance between residues that form bridges was measured in the time course of the simulations. The chain pairs A & C, A & D, and B & D were shown as green, orange, and cyan lines in the graphs, respectively. Oxygen-nitrogen distance cutoff was set to 3.2 Å. However, salt bridge distance up to 4 Å was acceptable as conserved interaction.

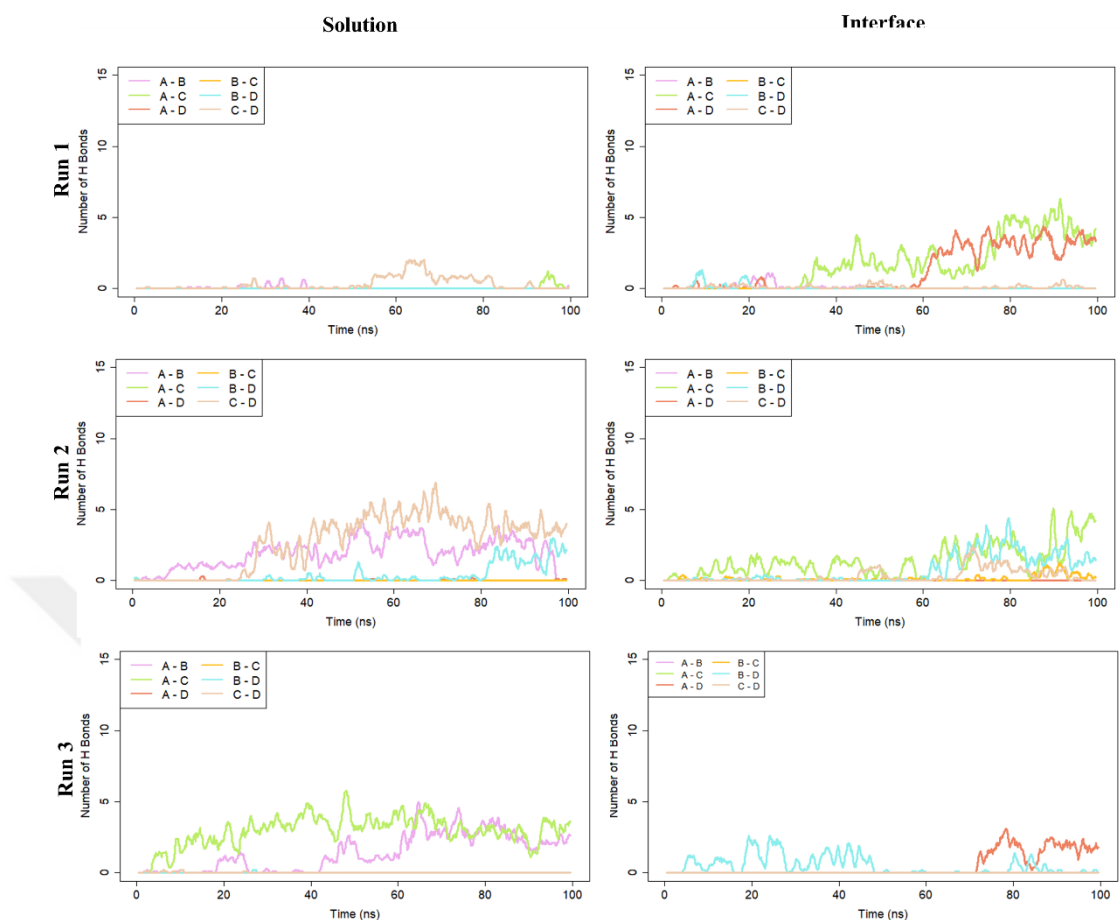


Figure 75. Measurements of the number of hydrogen bonds between chain pairs in Orientation 2 simulation systems with positionally RWM in both environments. The chain pairs A & B, A & C, A & D, B & C, B & D, and C & D were shown as purple, green, orange, yellow, cyan, and beige lines in the graphs, respectively. Donor-acceptor distance and angle cutoff were set at 3.5 Å and 30° respectively.

The monomer mobility was higher than the average (higher RMSD) only in two isolated interface monomers (run1 and run3). L3 of solution monomers were in high fluctuation compared to interface (Figure 76).

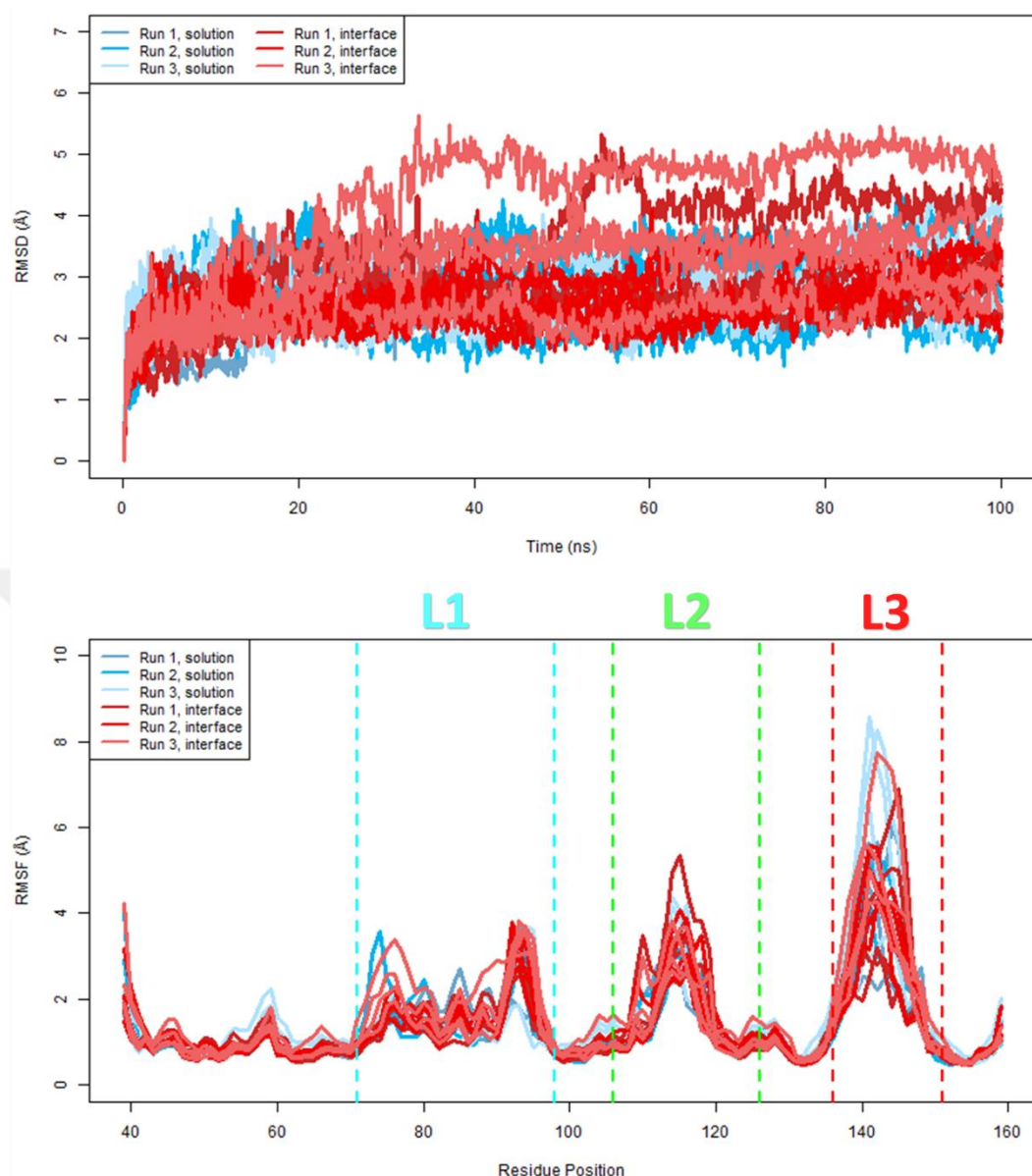


Figure 76. RMSD and RMSF measurements of chains in Orientation 2 simulations with positionally RWM in both environments. Blue and red color gradients (from light to dark) were used to differentiate parallel solution and interface simulations, respectively. Regions between dashed cyan, green, and red lines show L1, L2, and L3 loops, respectively.

4.3.2.3 Orientation 3

In the case of Orientation 3 simulations, monomers were placed in a position where L1 and L3 loops were pointed towards the center of mass of the four monomers. Clusters of at least two and three monomers were observed in solution and at interface simulations, respectively (Figure 77). In solution simulations, four monomers

sequentially interacted and formed an elongated-shaped assembly in the first solution run. COM–COM distances between chains A & B and C & D were the smallest due to the proximity of central β -barrels from the early steps of the trajectory. In the second run, two clusters were observed after 20 ns and conserved in the rest of the trajectory. The distances between chains A & B and C & D decreased after 60 ns and were stabilized around 30 Å. A stable interaction was observed only in a chain pair in the last solution run (Figure 78).



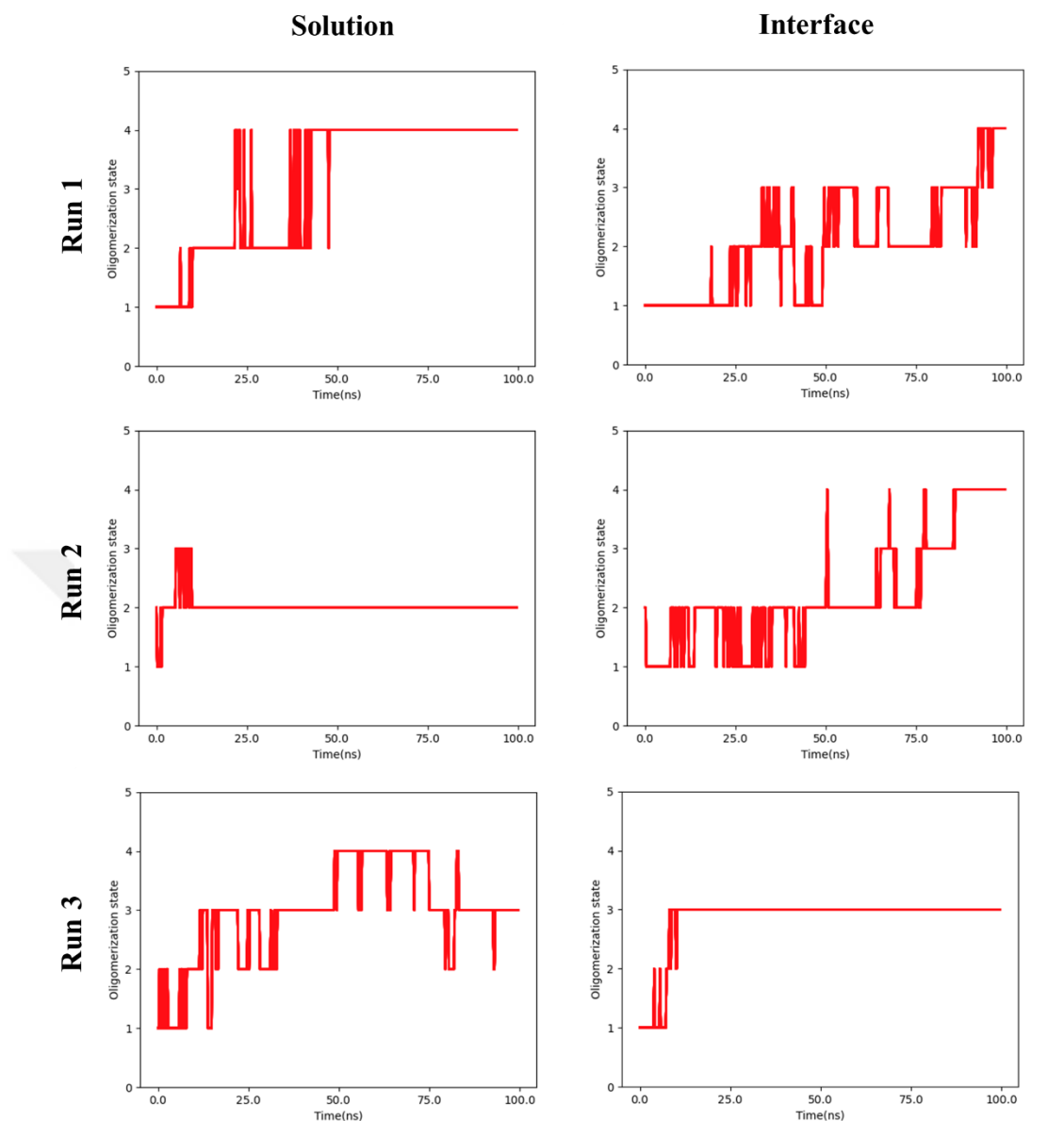


Figure 77. Cluster analysis of Orientation 3 simulation systems with positionally RWM. Monomers in a 4 Å distance cutoff assumed as clustered in the simulation systems and the cluster lifetime was measured throughout the trajectories for both simulation environments.

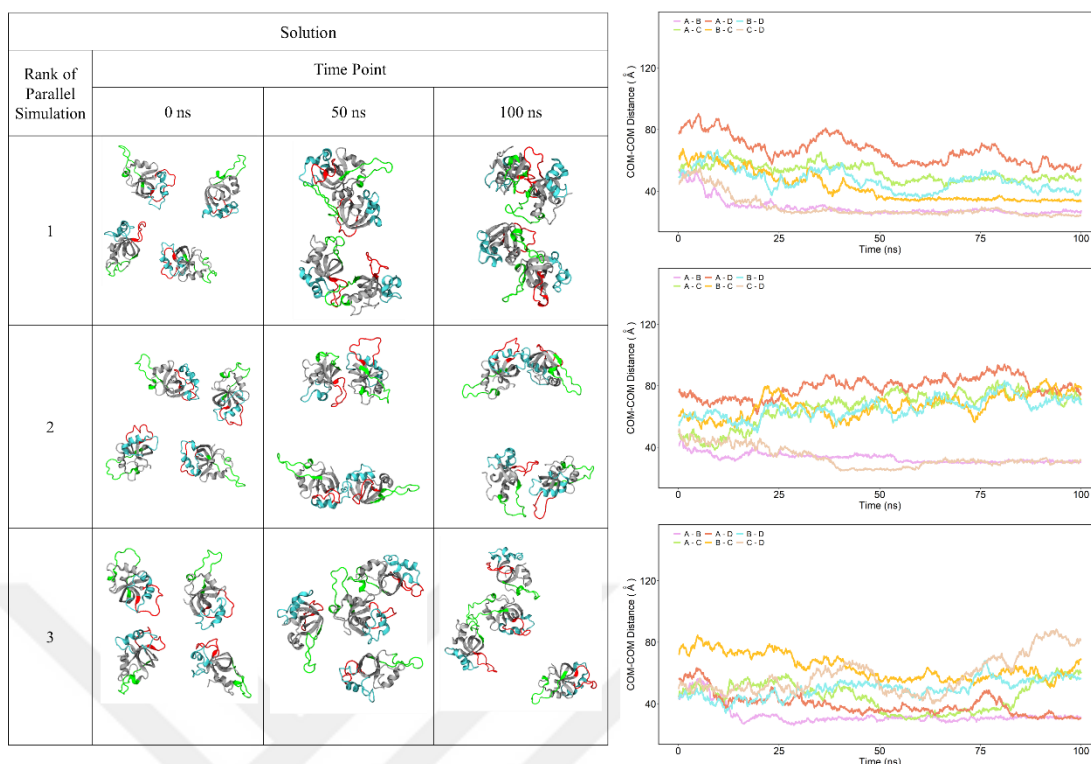


Figure 78. Snapshots of conformations at three different time points and COM–COM distances between chains of Orientation 2 solution simulation systems with positionally RWM. Snapshots were taken at the beginning, middle, and end of the trajectory. L1, L2, and L3 loops and non-loop regions were denoted in cyan, green, red, and gray cartoon representations, respectively. The COM–COM distances between chains A & B, A & C, A & D, B & C, B & D and C & D were shown as purple, green, orange, yellow, cyan, and beige lines, respectively.

At the air/water interface, clusters formed by three or four monomers mostly in the later time steps of the simulations (Figure 77). In the first run, two, three, and four monomers were sequentially clustered and formed an elongated assembly (L-shaped) and all four monomers were adsorbed to the interfaces by facing L2, L3 or both loops to the upper or lower interface. In these large cluster, monomer pairs were more stable compared to three or four monomers. Similar sequential clustering behavior was also observed in the second parallel run in the later time points. Monomers were randomly oriented at the upper (three monomers) and lower (only one monomer) interface in the cluster and each monomer was in interaction with the interface through L2 loops. B-D and B-C-D cluster was formed sequentially with decreasing monomer distances respectively in the last 30 ns; chain A joined the cluster in the last 15 ns. In the last interface run, a cluster of three formed in the early steps and was conserved. Each

monomer in the cluster was in interaction with the interface: Two with the upper and one with the lower interface. A and D were facing same air/water interface due to the Y-shaped cluster. The intermolecular distance was not stable (Figure 79).

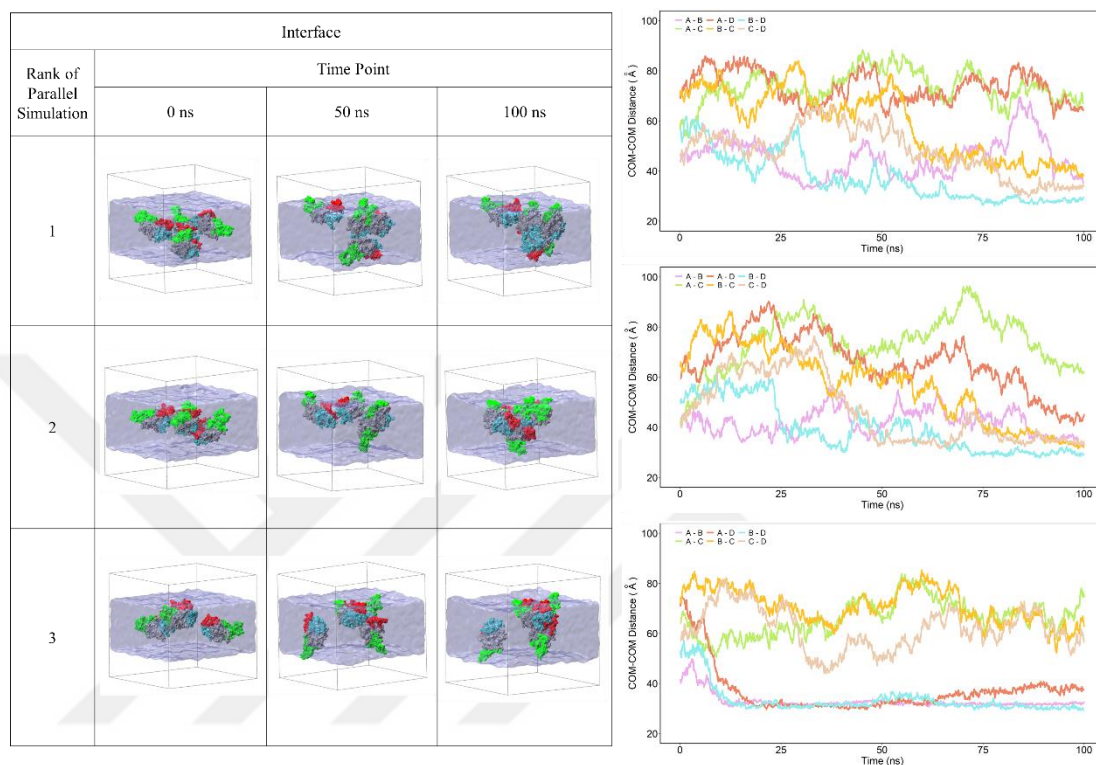


Figure 79. Snapshots of conformations at three different time points and COM–COM distances between chains of Orientation 3 interface simulation systems with positionally RWM. Snapshots were taken at the beginning, middle, and end of the trajectory. L1, L2, and L3 loops and non-loop regions were denoted in cyan, green, red, and gray surface representations, respectively. The COM–COM distances between chains A & B, A & C, A & D, B & C, B & D, and C & D were shown as purple, green, orange, yellow, cyan, and beige lines, respectively.

The intermolecular interaction sites of the monomers in a cluster varied in each run of both simulation environments. Although L1 and L3 loops were centered in the initial orientations, most conserved chain interactions were through N-termini and L2 loops were observed in addition to L3 loop in solution simulations. The inter-loop regions also contributed to the cluster stabilities in different parallel runs (Figure 80). N-termini were making contacts in the significant time steps of the trajectories with the loop regions (L1 and L3) (Appendix 13). Relatively larger regions of L2 loops were involved in interactions in solution simulations. At the air/water interface, N-

termini were the hotspot for the intermolecular interactions (Figure 81). Larger monomer regions were found to be involved in the intermolecular interactions in solution clusters. The high number of atomic contacts concentrated in the small regions in both simulation environments. Conserved L2-L2 contacts were high in number in chain pairs in solution clusters (Figure 80).

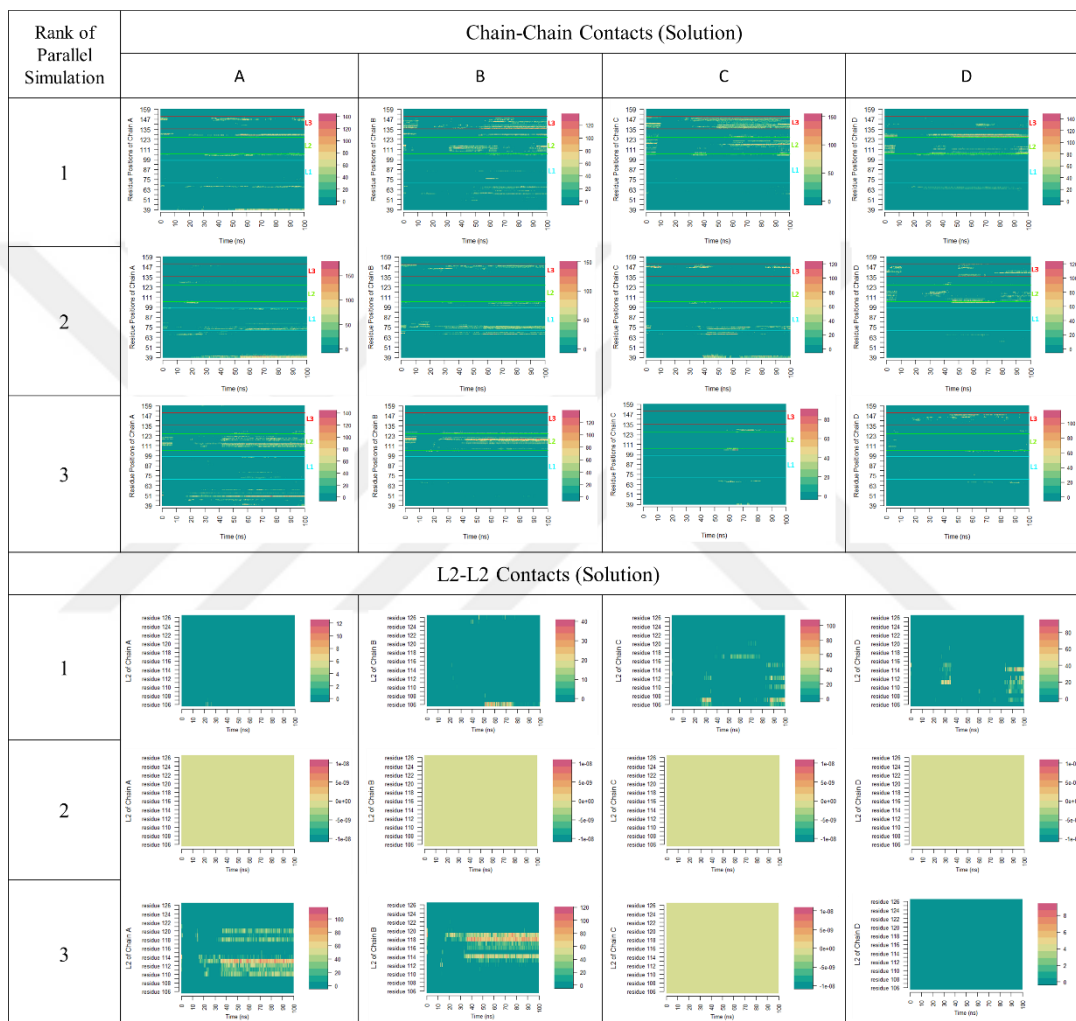


Figure 80. Contact maps of Orientation 3 solution simulation systems with positionally RWM. Each map shows the interaction of the chain column with the remaining chains of the simulation system. L1, L2, and L3 loops were highlighted on the maps. L2-L2 interactions were measured by selecting only L2 loops. Light yellow background on the maps indicates no atomic interactions measured. The number of atomic contacts was assigned to a color gradient from dark green to red. The distance cutoff was set to 4 Å.

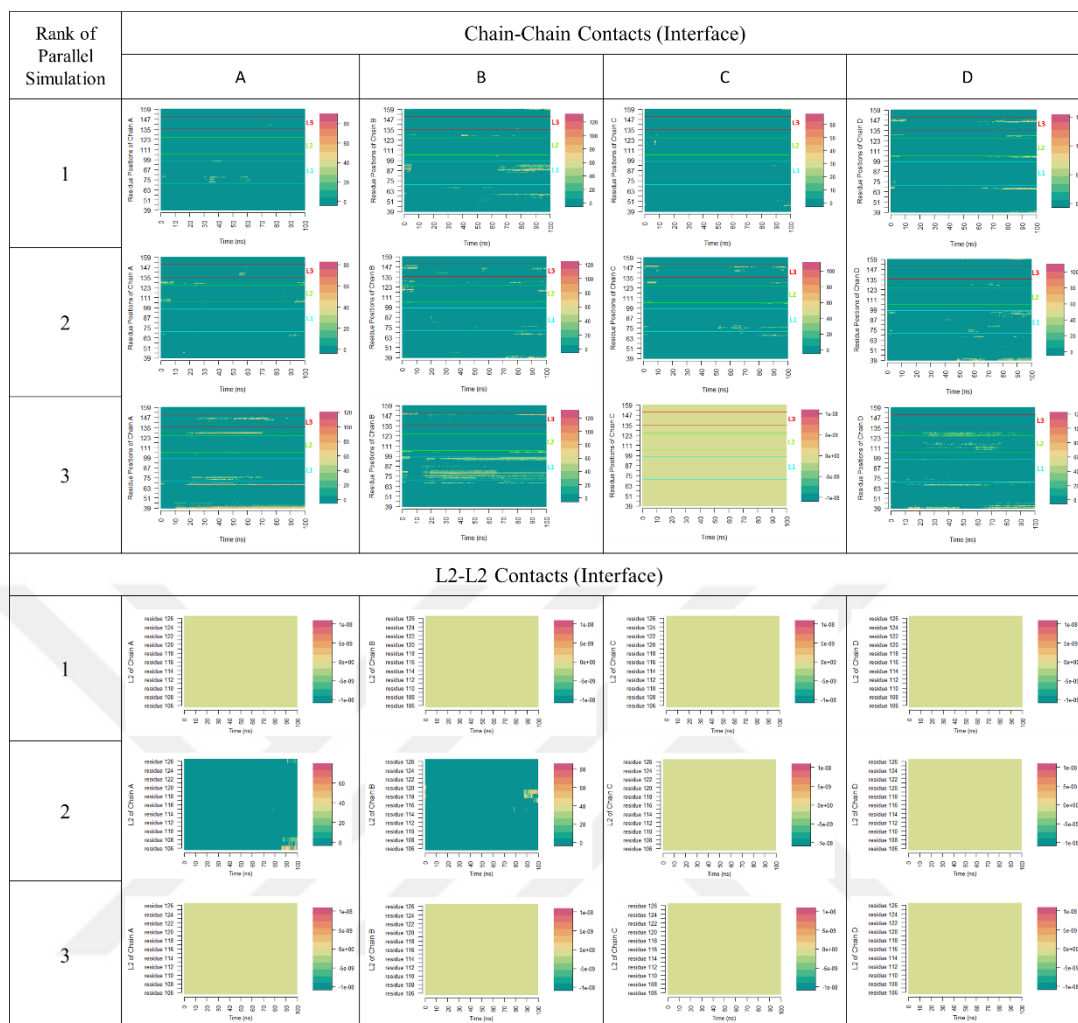


Figure 81. Contact maps of Orientation 3 interface simulation systems with positionally RWM. Each map shows the interaction of the chain column with the remaining chains of the simulation system. L1, L2, and L3 loops were highlighted on the maps. L2-L2 interactions were measured by selecting only L2 loops. Light yellow background on the maps indicates no atomic interactions measured. The number of atomic contacts was assigned to a color gradient from dark green to red. The distance cutoff was set to 4 Å.

Conserved hydrophobic contacts were mostly in the loop regions (L2 and L3) of the solution clusters. On the other hand, hydrophobic interactions were not dominant at the interface (Figure 82). The number of salt bridges was higher between monomers clustered freely in the solution and both loop-loop and N termini-loop salt bridges were observed. At the interface, ionic interactions formed mostly between N-termini and loops, and most of the bridges formed in the second half of the simulations of the last parallel run (Figure 83). After 50 ns, the number of hydrogen bonds and the chain pairs forming these bonds increased in the first two parallel solution simulations (≥ 5) and

only a chain pair conserved its hydrogen bonds in the last run; others were fluctuating. At the air/water interface, there was no hydrogen bond observed in the first half of the first two parallel runs; a small number of bonds were mostly formed between chain pairs in the later time steps. In the last run, the A-B-D cluster formed conserved hydrogen bond networks from the early time steps of the simulation (Figure 84).

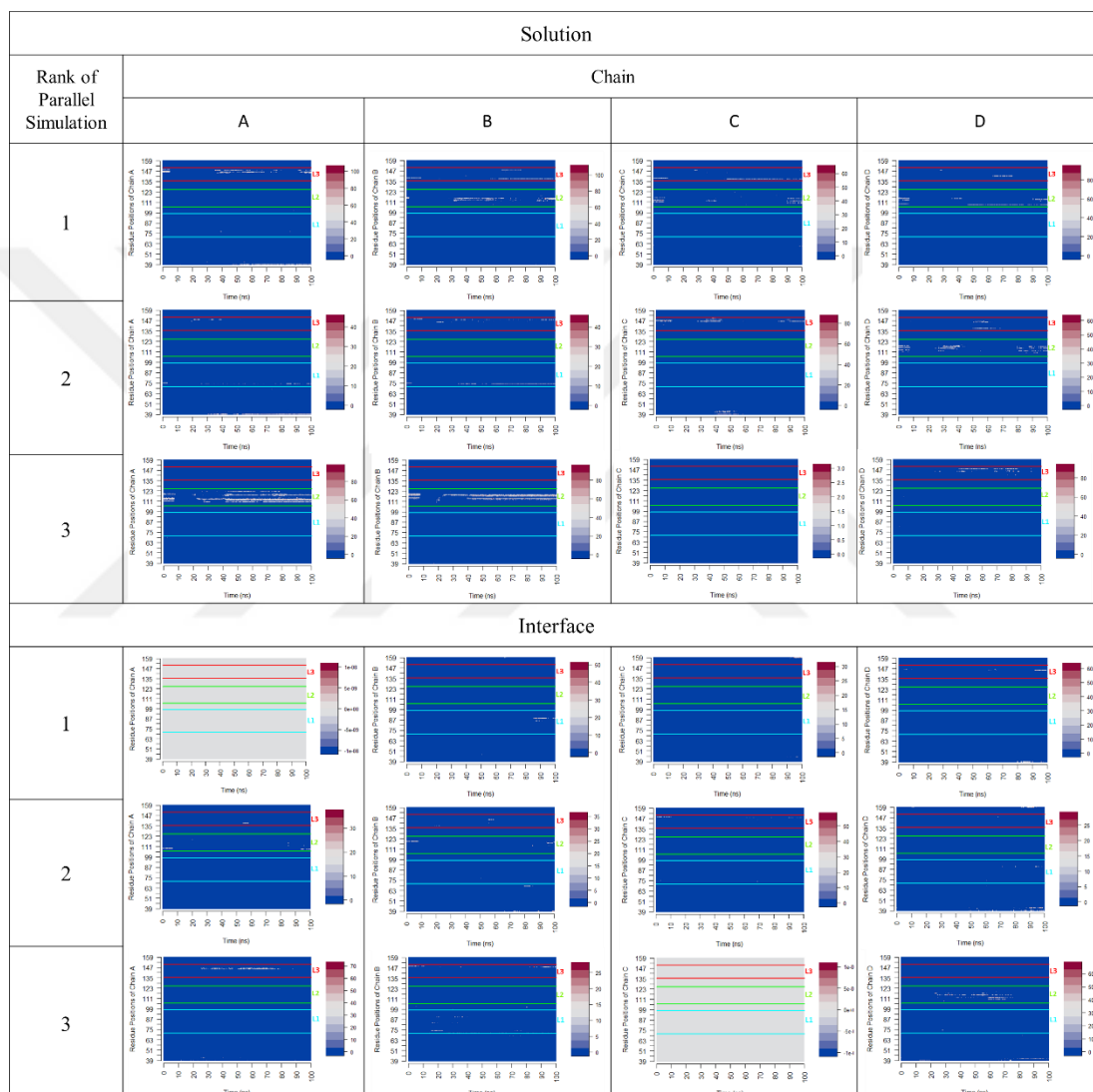


Figure 82. Hydrophobic contact maps of both Orientation 3 solution and interface simulation systems with positionally RWM. Each map shows the interaction of the chain column with the remaining chains of the simulation system. L1, L2, and L3 loops were highlighted on the maps. The number of atomic contacts was assigned to a color gradient from dark blue to red. The distance cutoff was set to 4 Å.

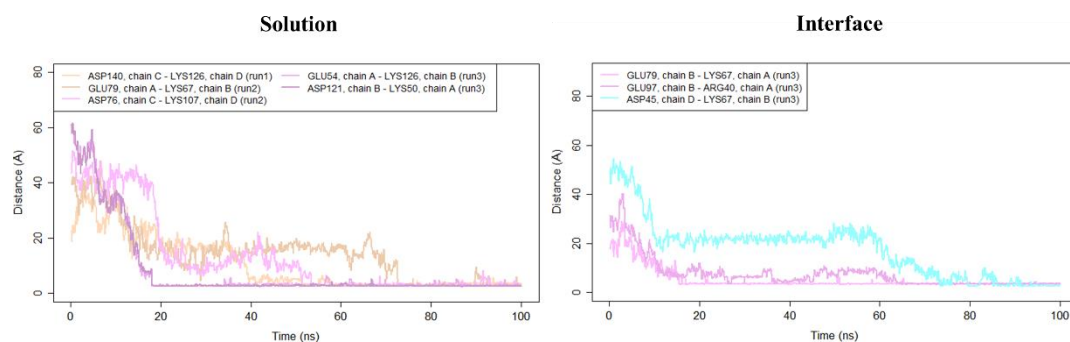


Figure 83. Measurements of conserved interchain salt bridges throughout the trajectories in Orientation 3 simulation systems with positionally RWM for both environments. The distance between residues that form bridges was measured in the time course of the simulations. The chain pairs A & B, B & D, and C & D were shown as purple, cyan and beige lines in the graphs, respectively. Oxygen-nitrogen distance cutoff was set to 3.2 Å. However, salt bridge distance up to 4 Å was acceptable as conserved interaction.

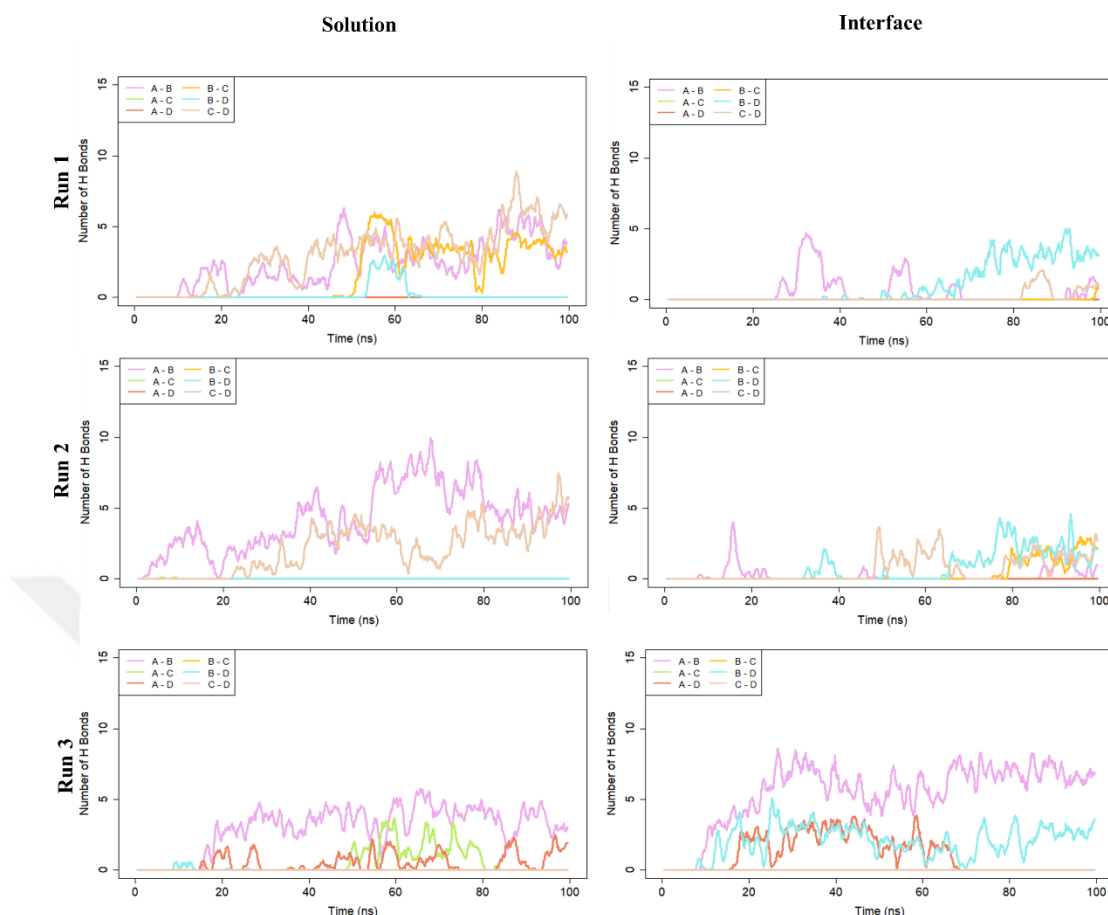


Figure 84. Measurements of the number of hydrogen bonds between chain pairs in Orientation 3 simulation systems with positionally RWM in both environments. The chain pairs A & B, A & C, A & D, B & C, B & D, and C & D were shown as purple, green, orange, yellow, cyan, and beige lines in the graphs, respectively. Donor-acceptor distance and angle cutoff were set at 3.5 Å and 30° respectively.

Solution monomers were more mobile in the Orientation 3 simulations and L3 loops were fluctuating on large scale. The mobility of L3 loops of monomers at interface simulations was restricted due to the air interaction at the air/water interface (Figure 85).

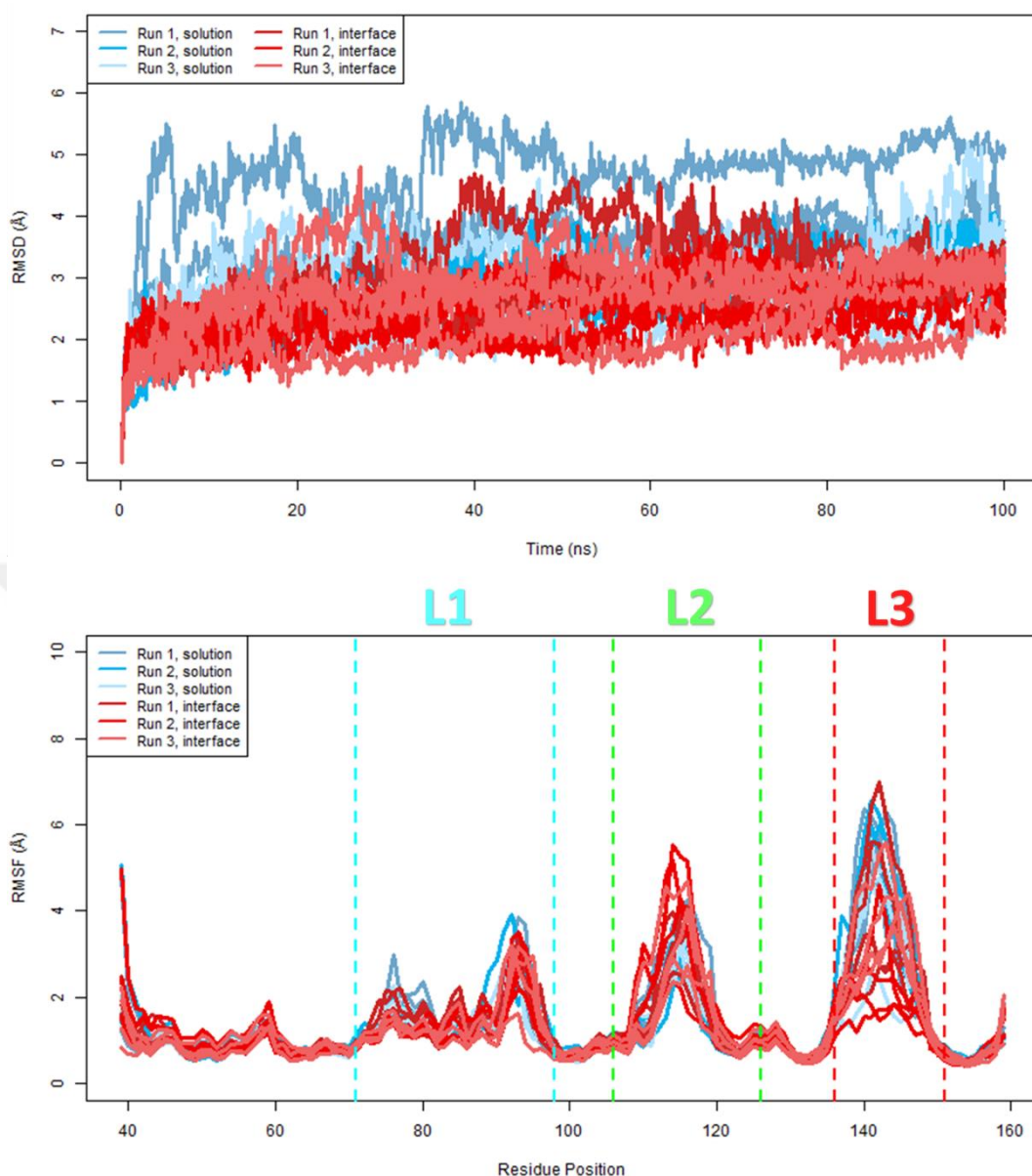


Figure 85. RMSD and RMSF measurements of chains in Orientation 3 simulations with positionally RWM in both environments. Blue and red color gradients (from light to dark) were used to differentiate parallel solution and interface simulations, respectively. Regions between dashed cyan, green, and red lines show L1, L2, and L3 loops, respectively.

4.3.2.4 Orientation 4

Clusters of two monomers were dominant in the solution and at interface simulations of Orientation 4. One of the runs in both environments was unstable and grouped into more crowded clusters (Figure 86). In the first run of solution simulations, four monomers assembled in a circular cluster by interacting through their

loops. The chains A & B and C & D COM–COM distances were shorter due to the proximities of the central β -barrels of the monomers. Two clusters of two monomers formed after the first 20 ns; one of the clusters was more stable and loops of similar type pointed to the same direction in the second parallel run. In the second cluster, the loop directions of the interacting monomers were not stable relative to each other. Loops were mostly free and fluctuating in both clusters. In the last parallel run of the solution simulations, a number of clusters of similar type was formed as in the second run. However, the directions of L2 and L3 loops were not stable. A-B and C-D chain distances reached an equilibrium in the last 20 ns, although the monomer pairs were formed in the earlier steps (Figure 87).



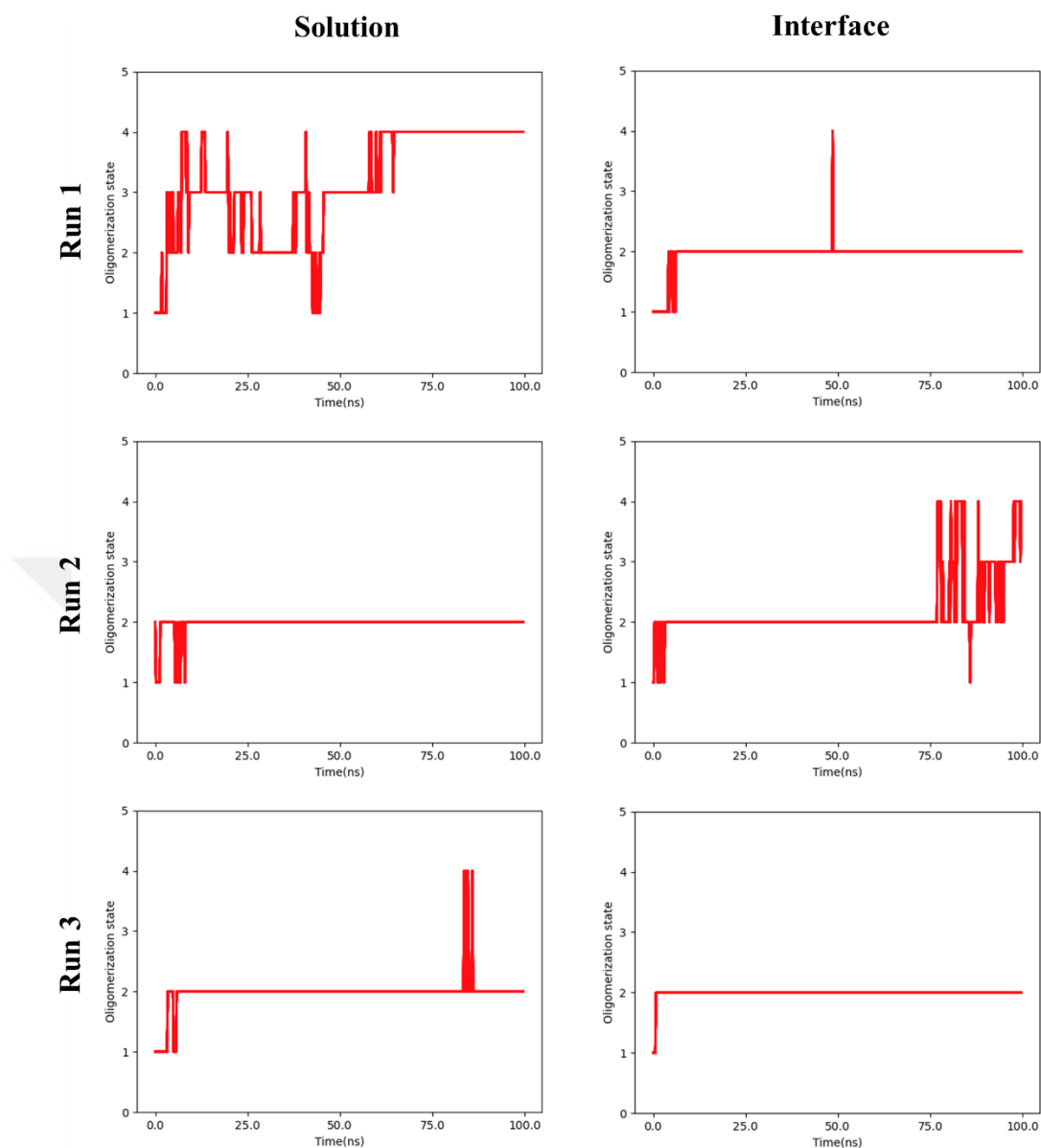


Figure 86. Cluster analysis of Orientation 4 simulation systems with positionally RWM. Monomers in a 4 Å distance cutoff assumed as clustered in the simulation systems and the cluster lifetime was measured throughout the trajectories for both simulation environments.

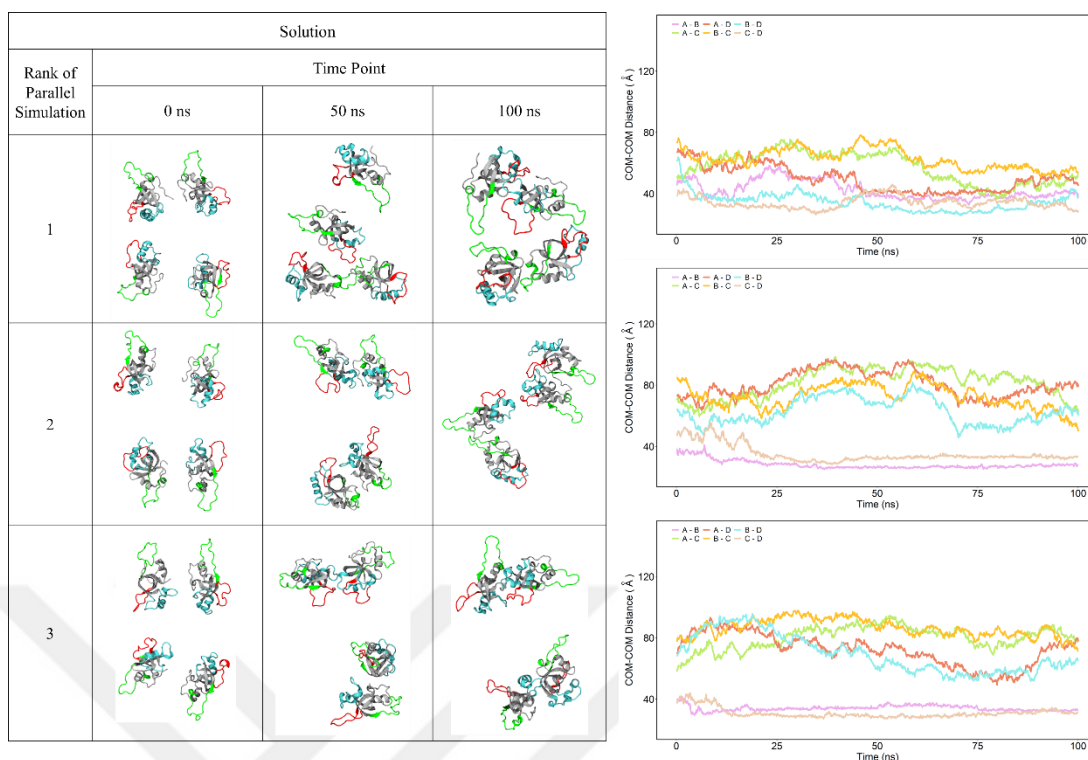


Figure 87. Snapshots of conformations at three different time points and COM-COM distances between chains of Orientation 4 solution simulation systems with positionally RWM. Snapshots were taken at the beginning, middle, and end of the trajectory. L1, L2, and L3 loops and non-loop regions were denoted in cyan, green, red, and gray cartoon representations, respectively. The COM-COM distances between chains A & B, A & C, A & D, B & C, B & D, and C & D were shown as purple, green, orange, yellow, cyan, and beige lines, respectively.

In the first interface run, two clusters formed both at the interface and in the bulk solution. The monomers in the cluster formed in bulk solution was rotating in a way to position the same types of loops in the same direction (for L2 and L3). Initially, two clusters were observed around 20 ns which lasted for ~55 ns in the second parallel run at the interface simulations; linear transient interactions of four monomers were observed in the later steps of the simulation. As in the first run, two clusters were observed from the early steps of the third parallel run at the upper and lower interface. In one of the clusters, L2 loop of only one monomer was in interaction with the interface; the loops of the other monomer were submerged due to interchain interaction through its loops. In the second cluster, monomers were in a relatively parallel orientation with a fluctuating intermolecular distance. Both L2 loops of the cluster were in interaction with the air at the interface (Figure 88).

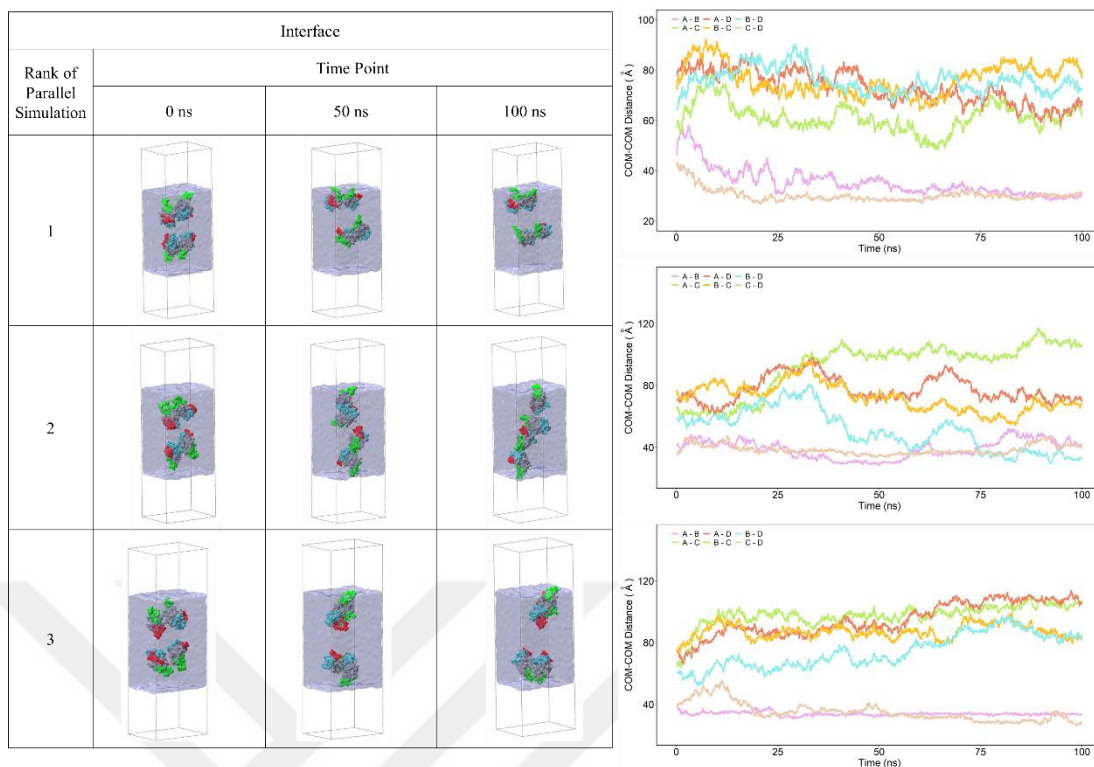


Figure 88. Snapshots of conformations at three different time points and COM–COM distances between chains of Orientation 4 interface simulation systems with positionally RWM. Snapshots were taken at the beginning, middle, and end of the trajectory. L1, L2, and L3 loops and non-loop regions were denoted in cyan, green, red, and gray surface representations, respectively. The COM–COM distances between chains A & B, A & C, A & D, B & C, B & D, and C & D were shown as purple, green, orange, yellow, cyan, and beige lines, respectively.

The intermolecular contacts were dominantly observed in the N-termini of the monomers followed by L2 loops in both simulation environments, especially at the interface clusters. Moreover, these regions were in interaction with each other with higher probability in total (Appendix 14). However, the highest number of atomic interactions was not necessarily observed in these dominant interaction regions. L2-L3 inter-loop region also contributed to the cluster stabilities and formed a high number of atomic contacts in both simulation environments. The total number of interaction regions was higher in solution clusters (Figure 89). Additionally, interactions involved small concentrated sequence sites in most of the cases. Conserved L2-L2 interactions were detected in the last 20 ns of one of the parallel runs in both simulation environments (Figure 89-90).

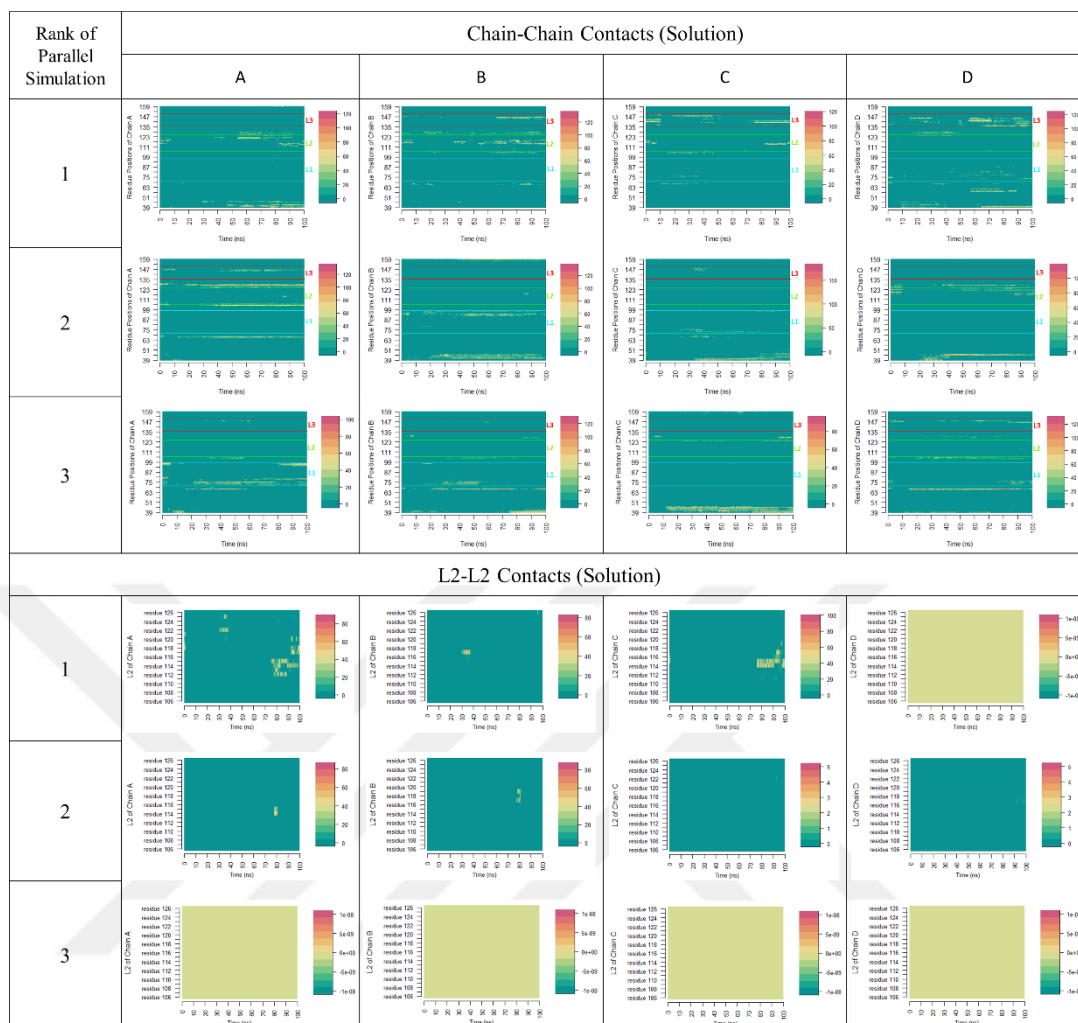


Figure 89. Contact maps of Orientation 4 solution simulation systems with positionally RWM. Each map shows the interaction of the chain column with the remaining chains of the simulation system. L1, L2, and L3 loops were highlighted on the maps. L2-L2 interactions were measured by selecting only L2 loops. Light yellow background on the maps indicates no atomic interactions measured. The number of atomic contacts was assigned to a color gradient from dark green to red. The distance cutoff was set to 4 Å.

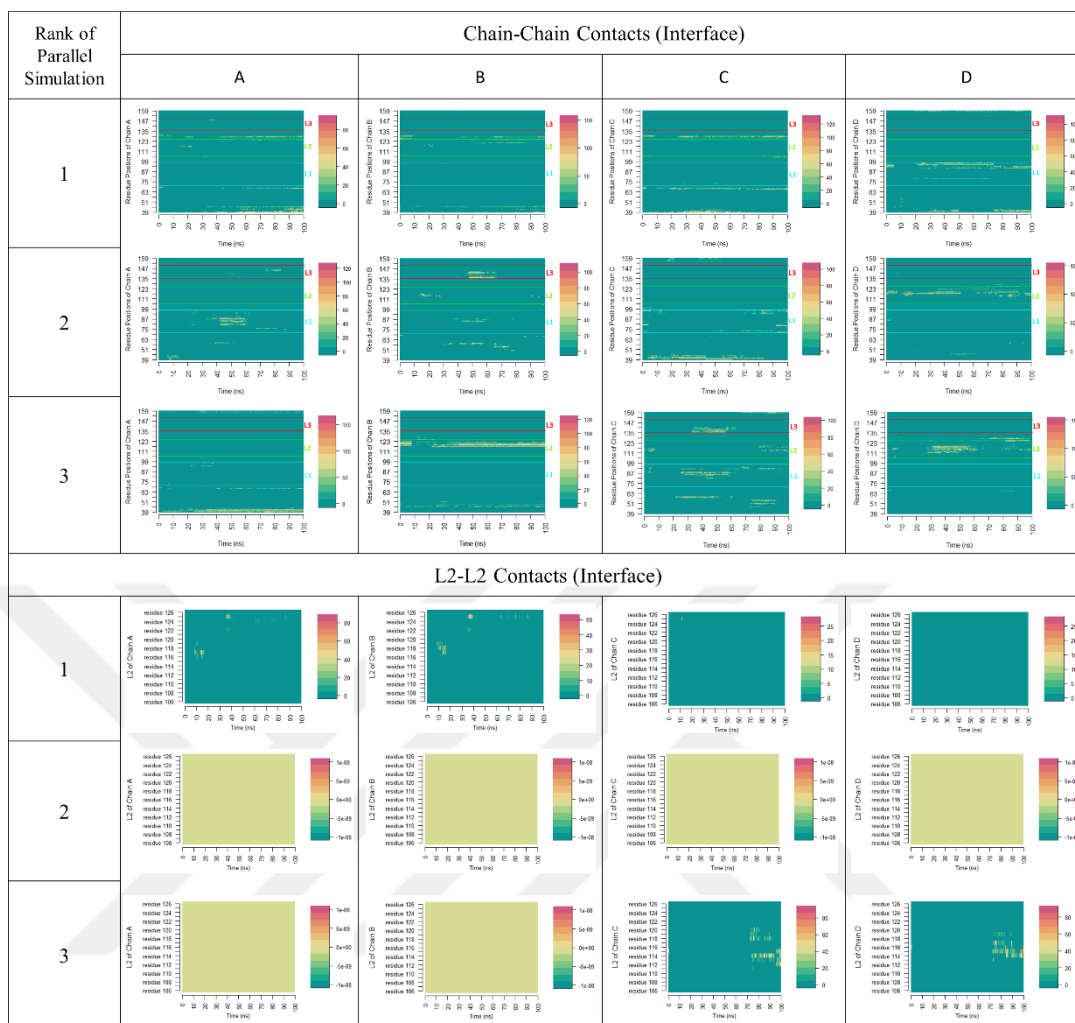


Figure 90. Contact maps of Orientation 4 interface simulation systems with positionally RWM. Each map shows the interaction of the chain column with the remaining chains of the simulation system. L1, L2, and L3 loops were highlighted on the maps. L2-L2 interactions were measured by selecting only L2 loops. Light yellow background on the maps indicates no atomic interactions measured. The number of atomic contacts was assigned to a color gradient from dark green to red. The distance cutoff was set to 4 Å.

Conserved hydrophobic interactions were only measured in the clusters of the solution simulations dominantly through the L2 loops in the first two runs. At the interface, no conserved interactions were observed in the first two runs; hydrophobic interactions were dominant in L2 loops in the last parallel interface simulation (Figure 91). The contribution of the ionic bonds to the cluster stability was not pronounced in the clusters of two monomers (Figure 92). There was no difference between the number of hydrogen bonds both in solution and interface clusters. However, hydrogen

bonds played a key role in the stability of the clusters from the early time steps of the simulations in both environments (Figure 93).

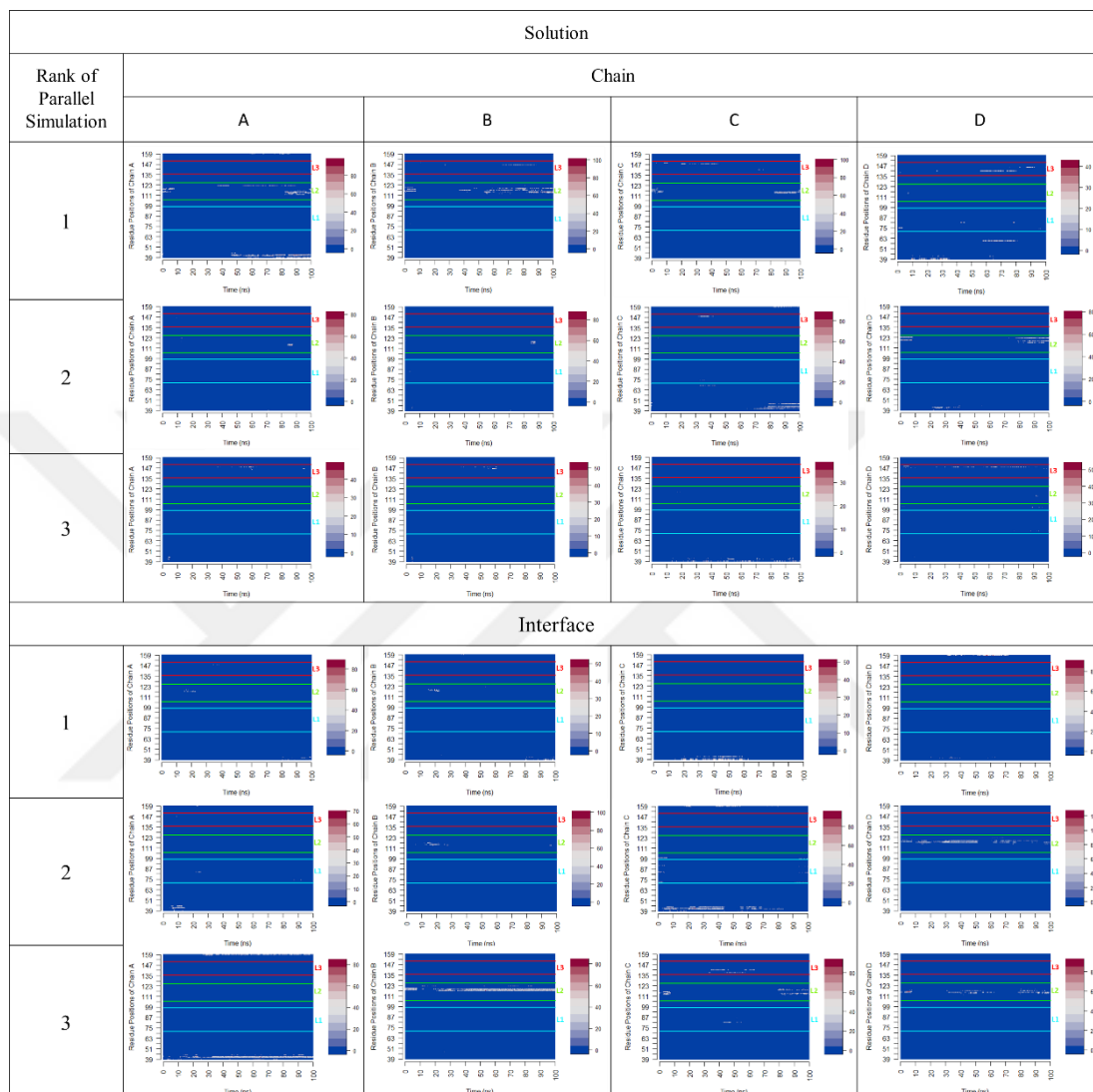


Figure 91. Hydrophobic contact maps of both Orientation 4 solution and interface simulation systems with positionally RWM. Each map shows the interaction of the chain column with the remaining chains of the simulation system. L1, L2, and L3 loops were highlighted on the maps. The number of atomic contacts was assigned to a color gradient from dark blue to red. The distance cutoff was set to 4 Å.

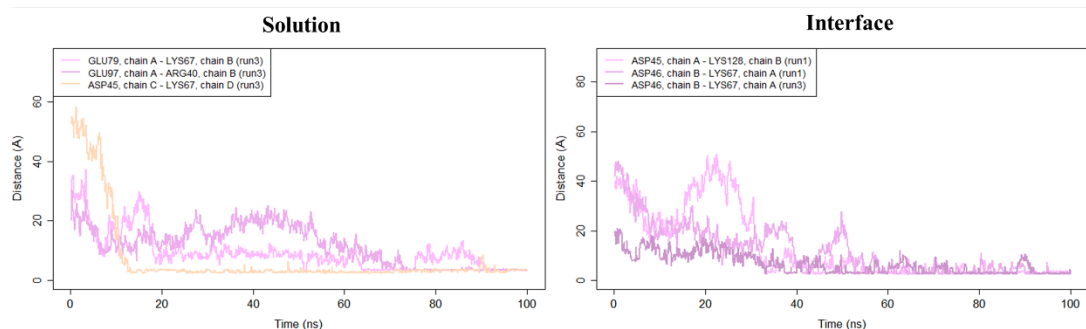


Figure 92. Measurements of conserved interchain salt bridges throughout the trajectories in Orientation 4 simulation systems with positionally RWM for both environments. The distance between residues that form bridges was measured in the time course of the simulations. The chain pairs A & B and C & D were shown as purple, and beige lines in the graphs, respectively. Oxygen-nitrogen distance cutoff was set to 3.2 Å. However, salt bridge distance up to 4 Å was acceptable as conserved interaction.

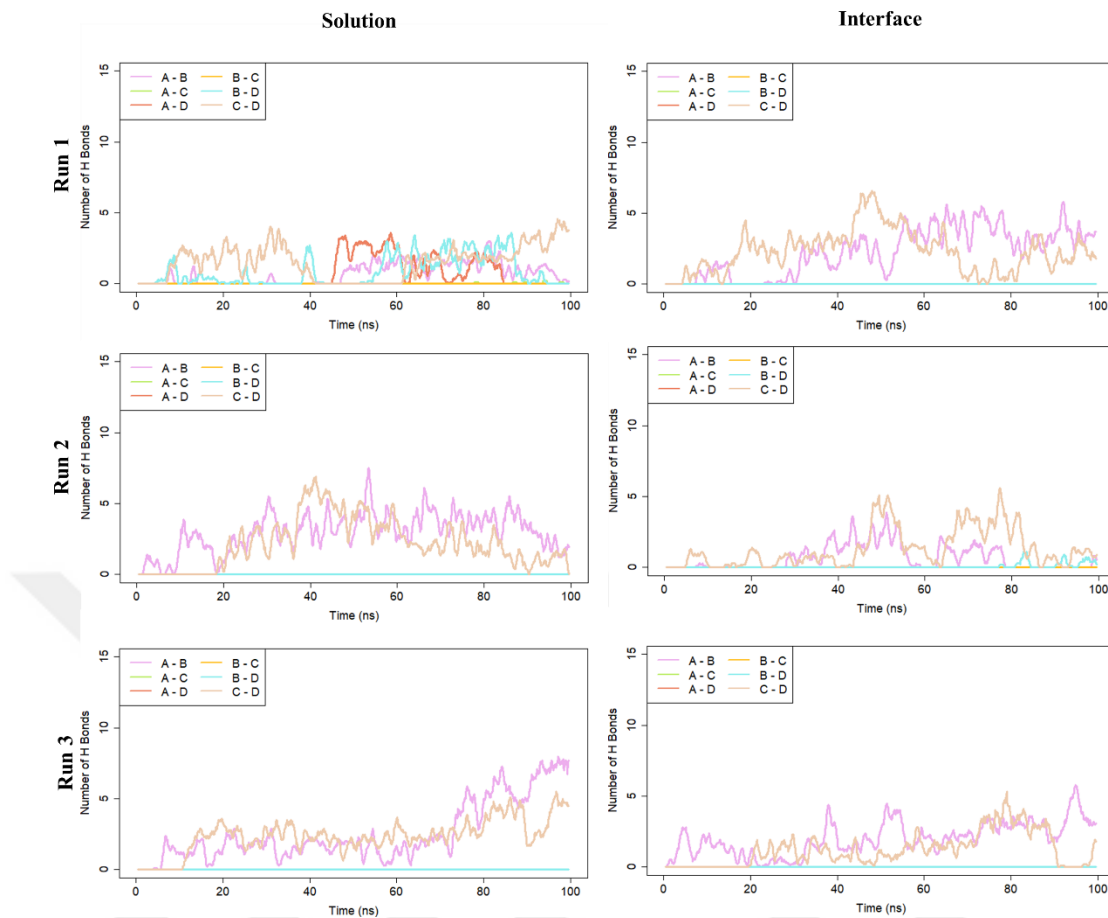


Figure 93. Measurements of the number of hydrogen bonds between chain pairs in Orientation 4 simulation systems with positionally RWM in both environments. The chain pairs A & B, A & C, A & D, B & C, B & D, and C & D were shown as purple, green, orange, yellow, cyan, and beige lines in the graphs, respectively. Donor-acceptor distance and angle cutoff were set at 3.5 Å and 30° respectively.

Orientation 4 monomers were relatively less mobile in all simulations regardless of the simulation environment. N- termini of the monomers in solution simulations were more mobile compared to interface monomers (Figure 94).

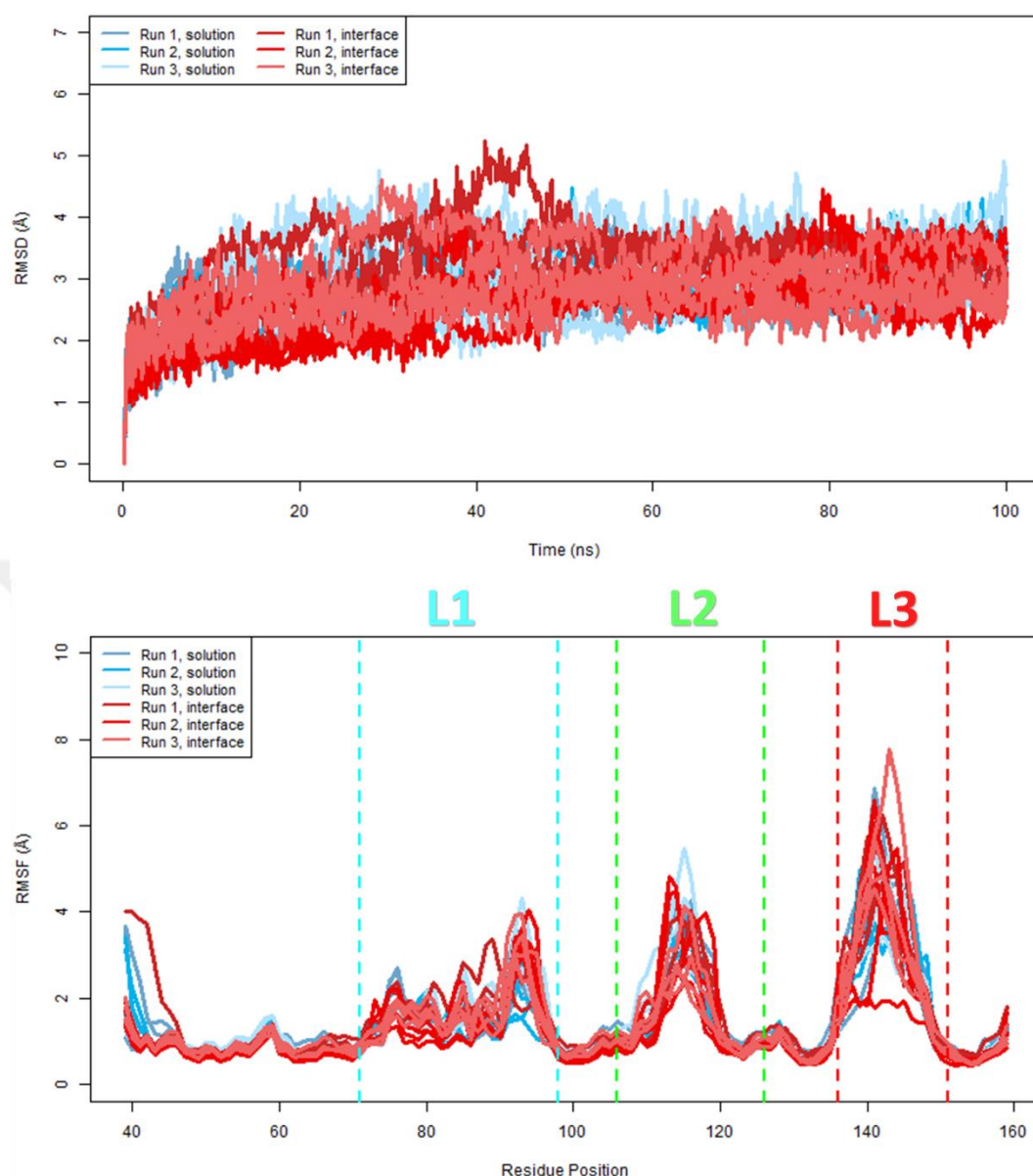


Figure 94. RMSD and RMSF measurements of chains in Orientation 4 simulations with positionally RWM in both environments. Blue and red color gradients (from light to dark) were used to differentiate parallel solution and interface simulations, respectively. Regions between dashed cyan, green, and red lines show L1, L2, and L3 loops, respectively.

5 DISCUSSION

A. fumigatus is an opportunistic fungal pathogen to humans and may cause severe respiratory diseases in the immune-deficient host after invasion which is facilitated by inhaled conidia. Although the cell wall of these spores is rich in polysaccharides which are known to have the potential to initiate the immune response against the pathogen, highly ordered self-assembled Class I RodA layers at the outer surface of conidia masks them from being recognized by the immune system until germination. Also, RodA itself was found to have the potential to trigger an immune response after interacting with a complement system protein in the innate immune system. In the earlier studies, both L2 and L3 loops were *in silico* determined as amyloidogenic and suggested to be involved in rodlet formation (11). In this thesis, it was hypothesized that homogenous L2 loop interactions play a significant role in initial RodA oligomerization. MD simulations were performed with single, two, and four RodA monomers both in bulk solution and at air/water interface. The water molecules near the edges of the simulation boxes were also restrained to prevent the interactions of RodA monomers with the periodic images. It was aimed to get insight into the initial steps of self-assembly at hydrophilic/hydrophobic interfaces by measuring intermolecular interactions that contribute to hydrophobin oligomerization and detecting key regions that play a role in RodA self-assembly and interface adsorption.

5.1 RodA Monomers were More Mobile in Bulk Solution Depending on the Initial Orientation

Class I hydrophobins are known to have larger loops and the contribution of this regions to hydrophobin solubility were investigated by De Simone *et al.* for L1 loop of EAS. Additionally, L1 loop of EAS was found to be less mobile after the adsorption to the air/water interface in 300 ns long REMD simulations (29). Unfolded SC3 folded into a β -sheet rich structure and was found to become more compact at the hexane/water interface in 100 ns (7). In case of RodA, all three loops (L1, L2 and L3) were mobile with the dominance of L3 loop, although L2 loop was the longest and unstructured one. However, the mobility of general structure of RodA and local

regions (loops) did not appear to be affected after the adsorption in 50 ns long monomer simulations.

In simulations with two monomers, the only difference in monomer stability between two simulation environments was observed in Orientation 1. L3 loops that were not involved in intermolecular interactions were in fluctuation on a larger scale in solution simulations. In the case of four-monomer simulations with positionally UWM, monomer mobilities were different only between solution and interface simulations of Orientation 2; L3 loops were more mobile in solution. There was no difference in terms of monomer stability between Orientation 2 and 4 simulations with positionally RWM in two different simulation environments, but general mobility decreased at Orientation 1 and Orientation 3 interface simulations; only L3 loops were more mobile in solution simulations. Although the interface adsorption was mostly formed through L2 loops, their flexibility was similar to L2 loops in bulk solution in which they participated in intermolecular interactions. Thus, longer simulation time may be required to monitor the changes in the hydrophobin structure and mobility at the air/water interface.

5.2 Compact Small Clusters were Formed Regardless of Initial Orientation

The clusters with smaller sizes formed in early time steps in general. The solution RodA dimers formed through anti-parallel monomers proposed by Zykwincka *et al.* was not observed in trajectories obtained in our study (24). Instead, clusters of two monomers with small COM–COM distances were dominant in two and four-monomers simulations, except Orientation 2 four-monomer simulations in which elongated clusters were observed. However, intermolecular distances between monomers of the clusters tend to decrease through a compacting process during most interface simulations. Clusters with higher compactness were detected in two- and four-monomer simulations with positionally RWM. This distinction was not observed in four-monomer simulations with UWM. In bulk solution, the clusters were mostly disorganized in large clusters of Orientation 1 and 3 simulations compared to small clusters. However, clusters of two monomers with close β -barrels were dominant in

four-monomer Orientation 4 simulations regardless of the simulation environment and water restraints. These clusters formed in the much earlier time steps with smallest intermolecular distances.

It can be claimed that the air/water interface positively affects the compactness of small clusters compared to the bulk solution since lower distances were measured between monomers. This orientation-dependent difference in clustering behavior was not observed in simulations with UWM, except Orientation 4. These findings indicate that longer simulation times can contribute to better understanding of the effect of the interface on hydrophobin mobility and clustering behavior.

5.3 Initial Orientation and RWM Affect Cluster Size

In literature, the self-assembly of several Class I hydrophobins was claimed to be facilitated through different homogenous loop interactions by forming stacked β -sheets of extended loops (33,34). In this thesis, four different initial orientations were generated for MD simulations of multiple RodA monomers by positioning the same loop types in proximity to each other to observe their interaction. By using these orientations, it was aimed to shorten the time that takes for the loop regions of the monomers to find each other in the water box in order for them to interact and form oligomers. Random orientation was also considered to increase conformation sampling (Orientation 1) and Orientation 4 was generated to test if the monomers find their interaction partners in the case of distant loop orientations.

In the case of two-monomer simulations, a number of RodA clusters were observed only in solution and interface simulations of Orientation 1 and Orientation 4. Clustering behavior was not significant in Orientation 2 and Orientation 3 simulations in which aggregation prone L2 and L3 loops of the monomers were in proximity of each other, respectively.

Regardless of the simulation environment and conditions of water molecules, clusters of two monomers formed in the first half of most simulation trajectories of

four-monomer simulations, except Orientation 1 where higher order clusters were dominant. Clusters of three monomers were mostly observed in the Orientation 3 interface simulations with positionally RWM. On the other hand, Orientation 4 simulations mostly resulted in the formation of clusters of the two monomers. Large clusters were mostly formed through a sequential clustering behavior in the later time steps of the trajectories and were mostly observed in simulations with positionally RWM.

The RodA cluster size seem to highly dependent on the initial orientation, more than the simulation environment, although simulations with RWM allowed us to observe that interface mediated the formation of relatively larger clusters.

5.4 N-termini and L2 Loops were Highly Involved in Intermolecular Interactions

The clusters were formed by intermolecular interactions through the N-termini and L2 loops in most simulations with high atomic contacts. However, the dominance of these regions in the intermolecular interactions varied between different simulations. In Orientation 3 simulations, L3 loop also played a role in cluster formation. Although these regions were highly in interaction, inter-loop regions, especially L2-L3 region, were making contributions to cluster stability with high number and long-lasting atomic contacts in the interface simulations with positionally RWM.

The duration of the intermolecular interactions was longer at Orientation 1 and Orientation 2 interface simulations with RWM. The interaction regions were relatively large in solution simulations, but transient interactions were also pronounced throughout the simulation time. Longer interactions were dominant at interface simulations with UWM.

To check if homogenous loop interactions had a role in RodA oligomerization, L2-L2 interactions were monitored since L2 loop is longer than L3 even though both

L2 and L3 were predicted as amyloidogenic. At the interface simulations with positionally RWM, high number of L2-L2 interactions were observed, except Orientation 3. There were also a number of heterogeneous loop interactions in clusters. Contrary to the models proposed in the literature for hydrophobin self-assembly, heterogeneous loop interactions (L1-L2, L1-L3 and L2-L3) were observed in addition to interactions of the N-termini.

5.5 Interactions Contributing to the Cluster Stability Differ Between the Two Simulation Environments

In most simulations performed with multiple RodA monomers, hydrophobic interactions were found to play a significant role in intermolecular interactions. This finding was in line with the study of the Kershaw and Talbot in which short-range hydrophobic interactions were claimed to be crucial in hydrophobin assemblies (45). The contribution of these interactions was mostly observed in the clusters formed in Orientation 1 and 2 simulations in four-monomer simulations with RWM. In general, the dominance of these interactions varied depending on the simulation environment. In simulations with positionally RWM and two-monomer simulations, hydrophobic interactions were measured more dominantly in L2 loops of solution clusters contrary to four-monomer simulations with UWM. In the case of Orientation 4, the contribution of the hydrophobic interactions to the cluster stability was not pronounced instead; a relatively high number of hydrogen bonds formed in the early time steps of both solution and interface simulations and were conserved. The reason of this decrease in hydrophobic interactions after interface adsorption probably resulted from the tendency of hydrophobic regions to remain in contact with air at the interface after adsorption as observed in the interface simulations of SC3 L1 dimers (27). In the simulations with UWM, hydrogen bonds formed in clusters either in later time steps or largely fluctuated throughout the trajectories. In the case of two-monomer simulations, the number of hydrogen bonds increased only in the second half of the trajectories of Orientation 1 interface simulations.

Salt bridges were observed in RodA monomers both in solution and interface simulations, but only the one that formed on the N-terminal α -helix was measured as stable (E54-K50). In addition to the ionic interactions (E54-K50 and K107-D109) stated in the study of Valsecchi *et al.* (11) for RodA monomers, five additional interactions (Figure 10) were observed in monomer simulations that may affect the flexibility of L1 and L2 loops in bulk solution since they were located at both N- and C- terminals of the loops. In RodA clusters, the contribution of ionic interactions to the cluster stability was only observed in Orientation 3 simulations regardless of simulation environment.

5.6 RodA Monomers Adsorbed to the Interface Mostly Through L2 Loops

In several studies, it was observed that hydrophobic surfaces of hydrophobins face towards the air at the air/water interfaces (36). In this thesis, both L2 and L3 loops of RodA that are known to contain hydrophobic stretches were highly surface active since they adsorbed to the interface within 50 ns individually or together in different runs in monomer simulations regardless of the initial orientation. Although the parallel runs were started with the same RodA orientations, short NPT runs performed prior to the interface simulations resulted in varied initial orientations and distances to the interfaces in monomer interface simulations. In the case of SC3, homology model of the monomer was transiently interacting with the air/water interface in 100 ns simulations and could not completely adsorb to the interface (27).

In the case of multiple monomers simulations, both clusters and individual monomers reached the interface in the first half of the simulations and L2 loops were dominantly interacting with the air. The reason of the dominance of L2 in surface adsorption may be explained as follows: Although the hydrophobic region (residues 136-148) is larger in L3 loops, the total length of the loop is relatively shorter and therefore could be more easily buried in the protein-protein interfaces in solution clusters. On the other hand, L2 loops were longer, and it was probably difficult to prevent them from being exposed in the RodA clusters.

Orientation 4 simulations provided significant clues about RodA self-assembly at the air/water interface. The clusters formed in these simulations were highly stable compared to others, especially the clusters in which both L2 and L3 loops of a monomer adsorbed to the interface. In the clusters in which L2 loops aligned in the same side and were interacting with the interface, the cluster stability was lower compared to former cluster. The reason of this relative instability was probably due to the fluctuations in L3 loops that were trying to reach the interface. With the help of these findings, a model could be proposed for self-assembly mechanism of RodA at the air/water interface: RodA monomers adsorbed to the surface both through their L2 and L3 loops and the loops probably aligned side-by-side in the oligomers in an open form without loop stacking unlike stacked loops proposed for EAS L3 loops in the study of Macindoe *et al.* (33) and with a close beta barrel distance between monomers. After surface adsorption, conformational rearrangements may result in extended loop conformations, but longer simulation time is needed to test this hypothesis. Both loops were found involve rodlet formation in previous studies. In our study, L2 loops were highly involved in intermolecular interactions in short simulations. I was also hypothesized that L2 loops were more available for amyloid like fibril formation. The surface of the hydrophobin which was claimed as to be hydrophobic faces to the air in this adsorption behavior.

5.7 There is No Common Structural Region Involved in Class I Hydrophobin Assembly

In addition to RodA, there are a number of hydrophobins in different pathogenic and non-pathogenic filamentous fungi known to function in the surface coverage of conidia and other aerial structural components and have ability to form amphipathic layers after secretion at hydrophilic/hydrophobic interfaces. However, the morphology of the layers, the number and pattern of secondary structural contents in monomeric structures (except four-stranded central β -barrels) and sequence lengths between cysteine residues mostly differ between Class I and Class II hydrophobins, although the pattern of disulfide bonds is conserved. The amino acid sequence and structures of the loop regions are highly divergent even within Class I hydrophobins including the

loops that were determined or predicted to involve in self-assembly. For example, L3 loop of EAS is unstructured and was claimed to form amyloid structures by adopting β -sheets in the hydrophobin assemblies, while structured L2 loop of DewA was suggested as the key player in amyloid formation (14,33). In the case of RodA monomer where both L2 and L3 were found to be critical regions in RodA self-assembly, the only significant secondary structural element (a long α -helix) was determined in the L1 loop; L2 and L3 loops were mostly unstructured in bulk solution as described in the study of Valsecchi *et al.* (11). Furthermore, to our knowledge, there was two loop regions that were suggested to be involve in RodA rodlet formation unlike remaining Class I hydrophobins. Due to these reasons, it was unlikely to make a consensus about the behavior of Class I hydrophobin self-assembly and use this information in determining the mechanism of RodA self-assembly.

5.8 Limitations

A full-length, experimentally determined 3D structure of RodA was not available, instead, the first model from NMR assemble of RodA missing 20 residues (19-39) was used in the MD simulations (PDB ID: 6GCJ). Although the model of the full-length structure was deposited in the AlphaFold Protein Structure Database recently, the confidence score of the predictions were low for the N-terminal region (91).

There was no experimental study about the self-assembly mechanism of RodA at the air/water interface or the multimeric state of the hydrophobin in solution to compare the results obtained in this thesis. There are assumptions about the solution state and the only available data was related to the morphology of layers on the conidia surface and Monte Carlo simulations were performed to investigate the layer formation on a 2D plate (24).

To reduce the computing time and size of the trajectory files, 50 (single and two-monomer simulations) and 100 ns-long (four-monomer simulations) simulations were run, and analyses were performed by loading the trajectory files by skipping several time steps. In addition, a queued system was used in TRUBA, and the queued jobs

were prioritized depending on the run time and user. Therefore, it was a time-consuming process to run a simulation even for 50 or 100 ns-long simulations with minimized atom numbers.

Last but not least, the environment that hydrophobins are secreted are much more crowded and molecules other than hydrophobins may also be present. It is unlikely to mimic the exact environment in MD simulations. The salt concentrations or ion types used in this thesis may not compensate for the ions present in its natural environment.



6 CONCLUSION

A. fumigatus RodA hydrophobins were studied by performing ns-long all-atom MD simulations in both bulk solution and air/water interfaces to get insights into the initial steps of hydrophobin self-assembly. In stability measurements, orientation-dependent decreases in the hydrophobin mobility and local fluctuations were observed at the interface simulations. Small hydrophobin clusters were mostly more compact at the interface and formed in earlier time steps of the trajectories regardless of the simulation condition. On the other hand, monomers clustered into larger structures sequentially and required more time to stabilize. The size of a cluster formed in simulations was also highly dependent on initial orientation; randomly distributed RodA monomers (Orientation 1) formed larger clusters, while clusters of two monomers were dominant in simulations in which the loops of RodA monomers were distant while central β -barrels were in close (Orientation 4). N-termini and L2 loops were involved in intermolecular interactions with high rates. However, the dominance of these regions varied in different simulations. Surprisingly, inter-loop regions contributed to cluster stability with several atomic contacts, especially in the interface simulations with positionally RWM. Unlike the consensus in the literature, heterogeneous loop and non-loop interactions were observed in the clusters. Depending on the initial orientation, the type of the interaction contributing to cluster stability varied. While hydrophobic interactions are common, a number of salt bridges was only observed in Orientation 3, especially at the interface. Both L2 and L3 loops were surface active and, with the help of mostly Orientation 4 simulations, a model for interface self-assembly mechanism was proposed. In the future studies, longer simulations of the selected initial orientations will be performed to observe the effect of the interface on RodA self-assembly and conformational changes adopted by the adsorbed hydrophobins. A higher number of monomers will be used in simulations to monitor the clustering in a more crowded environment. The self-assembling behavior of RodA on solid hydrophobic surface can be studied by MD simulations. Lastly, the interaction between complement system protein CD3 and RodA will be studied by using molecular docking methods and MD simulations.

7 REFERENCES

1. Wösten HAB. Hydrophobins: Multipurpose Proteins. *Annu Rev Microbiol.* 2001 Oct;55(1):625–46.
2. Bayry J, Aïmanianda V, Guijarro JI, Sunde M, Latgé JP. Hydrophobins—Unique Fungal Proteins. Heitman J, editor. *PLoS Pathog.* 2012 May 31;8(5):e1002700.
3. Wong SSW, Daniel I, Gangneux JP, Jayapal JM, Guegan H, Dellièvre S, et al. Differential Interactions of Serum and Bronchoalveolar Lavage Fluid Complement Proteins with Conidia of Airborne Fungal Pathogen *Aspergillus fumigatus*. Noverr MC, editor. *Infect Immun.* 2020 Aug 19;88(9):e00212-20.
4. Ley K, Christofferson A, Penna M, Winkler D, MacLaughlin S, Yarovsky I. Surface-water Interface Induces Conformational Changes Critical for Protein Adsorption: Implications for Monolayer Formation of EAS Hydrophobin. *Front Mol Biosci* [Internet]. 2015 Nov 16 [cited 2022 Jul 21];2. Available from: <http://journal.frontiersin.org/Article/10.3389/fmolb.2015.00064/abstract>
5. de Vocht ML, Scholtmeijer K, van der Vegte EW, de Vries OMH, Sonveaux N, Wösten HAB, et al. Structural Characterization of the Hydrophobin SC3, as a Monomer and after Self-Assembly at Hydrophobic/Hydrophilic Interfaces. *Biophysical Journal.* 1998 Apr;74(4):2059–68.
6. De Vocht ML, Reviakine I, Ulrich WP, Bergsma-Schutter W, Wösten HAB, Vogel H, et al. Self-assembly of the hydrophobin SC3 proceeds via two structural intermediates. *Protein Sci.* 2002 May;11(5):1199–205.
7. Zangi R, de Vocht ML, Robillard GT, Mark AE. Molecular Dynamics Study of the Folding of Hydrophobin SC3 at a Hydrophilic/Hydrophobic Interface. *Biophysical Journal.* 2002 Jul;83(1):112–24.
8. Kwan AH, Macindoe I, Vukašin PV, Morris VK, Kass I, Gupte R, et al. The Cys3–Cys4 Loop of the Hydrophobin EAS Is Not Required for Rodlet Formation and Surface Activity. *Journal of Molecular Biology.* 2008 Oct;382(3):708–20.
9. Butko P, Buford JP, Goodwin JS, Stroud PA, McCormick CL, Cannon GC. Spectroscopic Evidence for Amyloid-like Interfacial Self-Assembly of Hydrophobin Sc3. *Biochemical and Biophysical Research Communications.* 2001 Jan;280(1):212–5.
10. Zykwincka A, Guillemette T, Bouchara JP, Cuenot S. Spontaneous self-assembly of SC3 hydrophobins into nanorods in aqueous solution. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics.* 2014 Jul;1844(7):1231–7.
11. Valsecchi I, Lai JI, Stephen-Victor E, Pillé A, Beaussart A, Lo V, et al. Assembly and disassembly of *Aspergillus fumigatus* conidial rodlets. *The Cell Surface.* 2019 Dec;5:100023.
12. Mackay JP, Matthews JM, Winefield RD, Mackay LG, Haverkamp RG, Templeton MD. The Hydrophobin EAS Is Largely Unstructured in Solution and Functions by Forming Amyloid-Like Structures. *Structure.* 2001 Feb;9(2):83–91.

13. Morris VK, Ren Q, Macindoe I, Kwan AH, Byrne N, Sunde M. Recruitment of Class I Hydrophobins to the Air:Water Interface Initiates a Multi-step Process of Functional Amyloid Formation. *Journal of Biological Chemistry*. 2011 May;286(18):15955–63.
14. Morris VK, Kwan AH, Sunde M. Analysis of the Structure and Conformational States of DewA Gives Insight into the Assembly of the Fungal Hydrophobins. *Journal of Molecular Biology*. 2013 Jan;425(2):244–56.
15. Pham CLL, Rey A, Lo V, Soulès M, Ren Q, Meisl G, et al. Self-assembly of MPG1, a hydrophobin protein from the rice blast fungus that forms functional amyloid coatings, occurs by a surface-driven mechanism. *Sci Rep*. 2016 Jul;6(1):25288.
16. Paananen A, Vuorimaa E, Torkkeli M, Penttilä M, Kauranen M, Ikkala O, et al. Structural Hierarchy in Molecular Films of Two Class II Hydrophobins. *Biochemistry*. 2003 May 1;42(18):5253–8.
17. Szilvay GR, Paananen A, Laurikainen K, Vuorimaa E, Lemmetyinen H, Peltonen J, et al. Self-Assembled Hydrophobin Protein Films at the Air–Water Interface: Structural Analysis and Molecular Engineering. *Biochemistry*. 2007 Mar 1;46(9):2345–54.
18. Ren Q, Kwan AH, Sunde M. Solution structure and interface-driven self-assembly of NC2, a new member of the Class II hydrophobin proteins: Q. Ren et al. *Proteins*. 2014 Jun;82(6):990–1003.
19. Magarkar A, Mele N, Abdel-Rahman N, Butcher S, Torkkeli M, Serimaa R, et al. Hydrophobin Film Structure for HFBI and HFBII and Mechanism for Accelerated Film Formation. Nilges M, editor. *PLoS Comput Biol*. 2014 Jul 31;10(7):e1003745.
20. Chang HJ. Predicting the self-assembly film structure of class II hydrophobin NC2 and estimating its structural characteristics. 2020;10.
21. Latgé JP, Chamilos G. *Aspergillus fumigatus* and Aspergillosis in 2019. *Clin Microbiol Rev*. 2019 Dec 18;33(1):e00140-18.
22. Croft CA, Culibrk L, Moore MM, Tebbutt SJ. Interactions of *Aspergillus fumigatus* Conidia with Airway Epithelial Cells: A Critical Review. *Front Microbiol* [Internet]. 2016 Apr 7 [cited 2022 Dec 20];7. Available from: <http://journal.frontiersin.org/Article/10.3389/fmicb.2016.00472/abstract>
23. Aïmanianda V, Latgé JP. Fungal hydrophobins form a sheath preventing immune recognition of airborne conidia. *Virulence*. 2010 May;1(3):185–7.
24. Zykwincka A, Pihet M, Radji S, Bouchara JP, Cuenot S. Self-assembly of proteins into a three-dimensional multilayer system: Investigation of the surface of the human fungal pathogen *Aspergillus fumigatus*. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*. 2014 Jun;1844(6):1137–44.
25. Wösten HAB, Scholtmeijer K. Applications of hydrophobins: current state and perspectives. *Appl Microbiol Biotechnol*. 2015 Feb;99(4):1587–97.
26. Chiti F, Dobson CM. Protein Misfolding, Functional Amyloid, and Human Disease. *Annu Rev Biochem*. 2006 Jun 1;75(1):333–66.

27. Fan H, Wang X, Zhu J, Robillard GT, Mark AE. Molecular dynamics simulations of the hydrophobin SC3 at a hydrophobic/hydrophilic interface. *Proteins*. 2006 Jun 12;64(4):863–73.
28. Wang X. Oligomerization of hydrophobin SC3 in solution: From soluble state to self-assembly. *Protein Science*. 2004 Feb 6;13(3):810–21.
29. De Simone A, Kitchen C, Kwan AH, Sunde M, Dobson CM, Frenkel D. Intrinsic disorder modulates protein self-assembly and aggregation. *Proc Natl Acad Sci USA*. 2012 May;109(18):6951–6.
30. Scholtmeijer K, de Vocht ML, Rink R, Robillard GT, Wösten HAB. Assembly of the Fungal SC3 Hydrophobin into Functional Amyloid Fibrils Depends on Its Concentration and Is Promoted by Cell Wall Polysaccharides. *Journal of Biological Chemistry*. 2009 Sep;284(39):26309–14.
31. Cheung DL. Molecular Simulation of Hydrophobin Adsorption at an Oil–Water Interface. *Langmuir*. 2012 Jun 12;28(23):8730–6.
32. Mereghetti P, Wade RC. Diffusion of hydrophobin proteins in solution and interactions with a graphite surface. *BMC Biophys*. 2011 Dec;4(1):9.
33. Macindoe I, Kwan AH, Ren Q, Morris VK, Yang W, Mackay JP, et al. Self-assembly of functional, amphipathic amyloid monolayers by the fungal hydrophobin EAS. *Proc Natl Acad Sci USA* [Internet]. 2012 Apr 3 [cited 2022 Nov 7];109(14). Available from: <https://pnas.org/doi/full/10.1073/pnas.1114052109>
34. Pennacchio A, Cicatiello P, Notomista E, Giardina P, Piscitelli A. New clues into the self-assembly of Vmh2, a basidiomycota class I hydrophobin. *Biological Chemistry*. 2018 Jul 26;399(8):895–901.
35. Gandier JA, Langelaan DN, Won A, O'Donnell K, Grondin JL, Spencer HL, et al. Characterization of a Basidiomycota hydrophobin reveals the structural basis for a high-similarity Class I subdivision. *Sci Rep*. 2017 Apr;7(1):45863.
36. Kwan AHY, Winefield RD, Sunde M, Matthews JM, Haverkamp RG, Templeton MD, et al. Structural basis for rodlet assembly in fungal hydrophobins. *Proc Natl Acad Sci USA*. 2006 Mar 7;103(10):3621–6.
37. Wessels JGH. Hydrophobins: Proteins that Change the Nature of the Fungal Surface. In: *Advances in Microbial Physiology* [Internet]. Elsevier; 1996 [cited 2022 Nov 21]. p. 1–45. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S006529110860154X>
38. Mulder GH, Wessels JGH. Molecular cloning of RNAs differentially expressed in monokaryons and dikaryons of *Schizophyllum commune* in relation to fruiting. *Experimental Mycology*. 1986 Sep;10(3):214–27.
39. Wösten HA, Schuren FH, Wessels JG. Interfacial self-assembly of a hydrophobin into an amphipathic protein membrane mediates fungal attachment to hydrophobic surfaces. *The EMBO Journal*. 1994 Dec;13(24):5848–54.
40. Wösten HA, de Vries OM, van der Mei HC, Busscher HJ, Wessels JG. Atomic composition of the hydrophobic and hydrophilic membrane sides of self-assembled SC3p hydrophobin. *J Bacteriol*. 1994 Nov;176(22):7085–6.

41. Stringer MA, Timberlake WE. *dewA* encodes a fungal hydrophobin component of the *Aspergillus* spore wall. *Mol Microbiol.* 1995 Apr;16(1):33–44.
42. Bell-Pedersen D, Shinohara ML, Loros JJ, Dunlap JC. Circadian clock-controlled genes isolated from *Neurospora crassa* are late night- to early morning-specific. *Proc Natl Acad Sci USA.* 1996 Nov 12;93(23):13096–101.
43. Talbot NJ, Ebbole DJ, Hamer JE. Identification and characterization of MPG1, a gene involved in pathogenicity from the rice blast fungus *Magnaporthe grisea*. *The Plant Cell.* 1993 Nov 1;5(11):1575–90.
44. Nakari-Setälä T, Aro N, IlmeN M, Munoz G, Kalkkinen N, Penttilä M. Differential Expression of the Vegetative and Spore-Bound Hydrophobins of *Trichoderma Reesei* Cloning and Characterization of the Hfb2 Gene. *Eur J Biochem.* 1997 Sep;248(2):415–23.
45. Kershaw MJ, Talbot NJ. Hydrophobins and Repellents: Proteins with Fundamental Roles in Fungal Morphogenesis. *Fungal Genetics and Biology.* 1998 Feb;23(1):18–33.
46. Wösten HAB, van Wetter MA, Lugones LG, van der Mei HC, Busscher HJ, Wessels JGH. How a fungus escapes the water to grow into the air. *Current Biology.* 1999 Jan;9(2):85–8.
47. Temple B, Horgen PA. Biological roles for cerato-ulmin, a hydrophobin secreted by the elm pathogens, *Ophiostoma ulmi* and *O. novo-ulmi*. *Mycologia.* 2000 Jan;92(1):1–9.
48. Brown GD, Denning DW, Gow NAR, Levitz SM, Netea MG, White TC. Hidden Killers: Human Fungal Infections. *Sci Transl Med [Internet].* 2012 Dec 19 [cited 2022 Dec 1];4(165). Available from: <https://www.science.org/doi/10.1126/scitranslmed.3004404>
49. Valsecchi I, Stephen-Victor E, Wong SSW, Karnam A, Sunde M, Guijarro JI, et al. The Role of RodA-Conserved Cysteine Residues in the *Aspergillus fumigatus* Conidial Surface Organization. *JoF.* 2020 Aug 26;6(3):151.
50. Tamura K, Stecher G, Kumar S. MEGA11: Molecular Evolutionary Genetics Analysis Version 11. Battistuzzi FU, editor. *Molecular Biology and Evolution.* 2021 Jun 25;38(7):3022–7.
51. Okonechnikov K, Golosova O, Fursov M. Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics.* 2012 Apr;28(8):1166–7.
52. Sunde M, Kwan AHY, Templeton MD, Beever RE, Mackay JP. Structural analysis of hydrophobins. *Micron.* 2008 Oct;39(7):773–84.
53. Hakanpää J. Two crystal structures of *Trichoderma reesei* hydrophobin HFBI--The structure of a protein amphiphile with and without detergent interaction. *Protein Science.* 2006 Sep 1;15(9):2129–40.
54. Hakanpää J, Linder M, Popov A, Schmidt A, Rouvinen J. Hydrophobin HFBII in detail: ultrahigh-resolution structure at 0.75 Å. *Acta Crystallogr D Biol Crystallogr.* 2006 Apr 1;62(4):356–67.
55. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al. UCSF Chimera—A visualization system for exploratory research and analysis. *J Comput Chem.* 2004 Oct;25(13):1605–12.

56. Wang X, Shi F, Wösten HAB, Hektor H, Poolman B, Robillard GT. The SC3 Hydrophobin Self-Assembles into a Membrane with Distinct Mass Transfer Properties. *Biophysical Journal*. 2005 May;88(5):3434–43.
57. Wosten HAB. Interfacial Self-Assembly of a Fungal Hydrophobin into a Hydrophobic Rodlet Layer. 1993;8.
58. de Vocht ML, Reviakine I, Wösten HAB, Brisson A, Wessels JGH, Robillard GT. Structural and Functional Role of the Disulfide Bridges in the Hydrophobin SC3. *Journal of Biological Chemistry*. 2000 Sep;275(37):28428–32.
59. Wang X, Permentier HP, Rink R, Kruijtz JAW, Liskamp RMJ, Wösten HAB, et al. Probing the Self-Assembly and the Accompanying Structural Changes of Hydrophobin SC3 on a Hydrophobic Surface by Mass Spectrometry. *Biophysical Journal*. 2004 Sep;87(3):1919–28.
60. Vergunst KL, Langelaan DN. The N-terminal tail of the hydrophobin SC16 is not required for rodlet formation. *Sci Rep*. 2022 Jan 10;12(1):366.
61. Gravagnuolo AM, Longobardi S, Luchini A, Appavou MS, De Stefano L, Notomista E, et al. Class I Hydrophobin Vmh2 Adopts Atypical Mechanisms to Self-Assemble into Functional Amyloid Fibrils. *Biomacromolecules*. 2016 Mar 14;17(3):954–64.
62. Askolin S, Linder M, Scholtmeijer K, Tenkanen M, Penttilä M, de Vocht ML, et al. Interaction and Comparison of a Class I Hydrophobin from *Schizophyllum commune* and Class II Hydrophobins from *Trichoderma reesei*. *Biomacromolecules*. 2006 Apr 1;7(4):1295–301.
63. Nakari-Setälä T, Aro N, Kalkkinen N, Alatalo E, Penttilä M. Genetic and Biochemical Characterization of the *Trichoderma reesei* Hydrophobin HFBII. *Eur J Biochem*. 1996 Jan;235(1–2):248–55.
64. Torkkeli M, Serimaa R, Ikkala O, Linder M. Aggregation and Self-Assembly of Hydrophobins from *Trichoderma reesei*: Low-Resolution Structural Models. *Biophysical Journal*. 2002 Oct;83(4):2240–7.
65. Hakanpää J, Paananen A, Askolin S, Nakari-Setälä T, Parkkinen T, Penttilä M, et al. Atomic Resolution Structure of the HFBII Hydrophobin, a Self-assembling Amphiphile. *Journal of Biological Chemistry*. 2004 Jan;279(1):534–9.
66. Lienemann M, Gandier JA, Joensuu JJ, Iwanaga A, Takatsuji Y, Haruyama T, et al. Structure-Function Relationships in Hydrophobins: Probing the Role of Charged Side Chains. *Appl Environ Microbiol*. 2013 Sep 15;79(18):5533–8.
67. Cheng Y, Wang B, Wang Y, Zhang H, Liu C, Yang L, et al. Soluble hydrophobin mutants produced in *Escherichia coli* can self-assemble at various interfaces. *Journal of Colloid and Interface Science*. 2020 Aug;573:384–95.
68. Sallada ND, Dunn KJ, Berger BW. A Structural and Functional Role for Disulfide Bonds in a Class II Hydrophobin. *Biochemistry*. 2018 Feb 6;57(5):645–53.
69. Hall BG. Building Phylogenetic Trees from Molecular Data with MEGA. *Molecular Biology and Evolution*. 2013 May 1;30(5):1229–35.

70. Badar MS, Shamsi S, Ahmed J, Alam MdA. Molecular Dynamics Simulations: Concept, Methods, and Applications. In: Rezaei N, editor. Transdisciplinarity [Internet]. Cham: Springer International Publishing; 2022 [cited 2022 Dec 20]. p. 131–51. (Integrated Science; vol. 5). Available from: https://link.springer.com/10.1007/978-3-030-94651-7_7
71. Alegre KO, Paknejad N, Su M, Lou JS, Huang J, Jordan KD, et al. Structural basis and mechanism of activation of two different families of G proteins by the same GPCR. *Nat Struct Mol Biol*. 2021 Nov;28(11):936–44.
72. Grubmüller H, Heller H, Windemuth A, Schulten K. Generalized Verlet Algorithm for Efficient Molecular Dynamics Simulations with Long-range Interactions. *Molecular Simulation*. 1991 Mar;6(1–3):121–42.
73. González MA. Force fields and molecular dynamics simulations. *JDN*. 2011;12:169–200.
74. Huang J, Rauscher S, Nawrocki G, Ran T, Feig M, de Groot BL, et al. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat Methods*. 2017 Jan;14(1):71–3.
75. Ponder JW, Case DA. Force Fields for Protein Simulations. In: *Advances in Protein Chemistry* [Internet]. Elsevier; 2003 [cited 2022 Dec 20]. p. 27–85. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S006532330366002X>
76. Scott WRP, Hünenberger PH, Tironi IG, Mark AE, Billeter SR, Fennel J, et al. The GROMOS Biomolecular Simulation Program Package. *J Phys Chem A*. 1999 May 1;103(19):3596–607.
77. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, et al. Scalable Molecular Dynamics with NAMD. 2008;43.
78. de Souza ON, Ornstein RL. Effect of periodic box size on aqueous molecular dynamics simulation of a DNA dodecamer with particle-mesh Ewald method. *Biophysical Journal*. 1997 Jun;72(6):2395–7.
79. Pitera JW. Expected Distributions of Root-Mean-Square Positional Deviations in Proteins. *J Phys Chem B*. 2014 Jun 19;118(24):6526–30.
80. Lobanov MYu, Bogatyreva NS, Galzitskaya OV. Radius of gyration as an indicator of protein structure compactness. *Mol Biol*. 2008 Aug;42(4):623–8.
81. Humphrey W, Dalke A, Schulten K. VMD: Visual molecular dynamics. *Journal of Molecular Graphics*. 1996 Feb;14(1):33–8.
82. Tsolis AC, Papandreou NC, Iconomidou VA, Hamodrakas SJ. A Consensus Method for the Prediction of ‘Aggregation-Prone’ Peptides in Globular Proteins. Vorberg IM, editor. *PLoS ONE*. 2013 Jan 10;8(1):e54175.
83. Rousseau F, Schymkowitz J, Serrano L. Protein aggregation and amyloidosis: confusion of the kinds? *Current Opinion in Structural Biology*. 2006 Feb;16(1):118–26.
84. de Vries SJ, Bonvin AMJJ. CPORT: A Consensus Interface Predictor and Its Performance in Prediction-Driven Docking with HADDOCK. Fernandez-Fuentes N, editor. *PLoS ONE*. 2011 Mar 25;6(3):e17695.

85. Boonstra S, Onck PR, van der Giessen E. CHARMM TIP3P Water Model Suppresses Peptide Folding by Solvating the Unfolded State. *J Phys Chem B*. 2016 Apr 21;120(15):3692–8.
86. Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., et al. The Protein Data Bank. *Nucleic Acids Research*. 2000;28:235–42.
87. Brooks BR, Brooks CL, Mackerell AD, Nilsson L, Petrella RJ, Roux B, et al. CHARMM: The biomolecular simulation program. *J Comput Chem*. 2009 Jul 30;30(10):1545–614.
88. Kräutler V, van Gunsteren WF, Hünenberger PH. A fast SHAKE algorithm to solve distance constraint equations for small molecules in molecular dynamics simulations. *J Comput Chem*. 2001 Apr 15;22(5):501–8.
89. Paul A, Samantray S, Anteghini M, Khaled M, Strodel B. Thermodynamics and kinetics of the amyloid- β peptide revealed by Markov state models based on MD data in agreement with experiment. *Chem Sci*. 2021;12(19):6652–69.
90. RStudio Team. RStudio: Integrated Development Environment for R [Internet]. [Internet]. Boston, MA; 2015. Available from: <http://www.rstudio.com/>
91. Varadi M, Anyango S, Deshpande M, Nair S, Natassia C, Yordanova G, et al. AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Research*. 2022 Jan 7;50(D1):D439–44.

8 APPENDIX

Appendix 1 Intermolecular distance measurements by “dist_bw_chains.tcl”.

```
set outfileAB [open dist_bw_AB_results_ori1_run1_WR.txt w]
set nf [molinfo top get numframes]
puts $outfileAB "cl_1_AB_ori1_run1_WR cl_2_AB_ori1_run1_WR"
for {set i 0} {$i < $nf} {incr i} {
    set chA [atomselect top "protein and chain A" frame $i]
    set comA [measure center $chA weight mass]
    set chB [atomselect top "protein and chain B" frame $i]
    set comB [measure center $chB weight mass]
    set comdistAB [vecdist $comA $comB]
    puts $outfileAB "$i \t $comdistAB"
}
close $outfileAB
```

Appendix 2 Atomic contact measurements by “contacts_for_chains.tcl”.

```
set num_steps [molinfo top get numframes]
set rodChains {A B C D}
foreach x $rodChains {

set fp [open "ori1_run1_WR_chain_$x _other_chains.txt" w]
puts $fp "fr_ori1_run1_WR_$x  num_ori1_run1_WR_$x "

for {set res_chain 39} {$res_chain < 160} {incr res_chain} {

for {set frame 0} {$frame < $num_steps} {incr frame} {

set sel1 [atomselect top "protein and chain $x and resid $res_chain" frame
$frame]
set sel2 [atomselect top "protein and not chain $x" frame $frame]

set cont [llength [lindex [measure contacts 4.0 $sel1 $sel2] 0]]

puts $fp "$frame  $cont"

}
}
close $fp
}
```

Appendix 3 Hydrophobic contact measurements by “hydrophobic_contacts.tcl”.

```
set num_steps [molinfo top get numframes]
set rodChains {A B C D}
foreach x $rodChains {

  set fp [open "ori1_run1_WR_chain_$x
_other_chains_hydrophobic_contacts.txt" w]
  puts $fp "fr_ori1_run1_WR_$x  num_ori1_run1_WR_$x "

  for {set res_chain 39} {$res_chain < 160} {incr res_chain} {

    for {set frame 0} {$frame < $num_steps} {incr frame} {

      set sel1 [atomselect top "protein and hydrophobic and chain $x and resid
$res_chain" frame $frame]
      set sel2 [atomselect top "protein and hydrophobic and not chain $x" frame
$frame]

      set cont [llength [lindex [measure contacts 4.0 $sel1 $sel2] 0]]

      puts $fp "$frame  $cont"

    }
  }
  close $fp
}
```


Appendix 4. Hydrogen bond measurements by “hbonds.tcl”.

```
set sel_chA [atomselect top "protein and chain A"]
set sel_chB [atomselect top "protein and chain B"]
set sel_chC [atomselect top "protein and chain C"]
set sel_chD [atomselect top "protein and chain D"]

hbonds -sel1 $sel_chA -sel2 $sel_chB -writefile yes -dist 3.5 -ang 30 -
outfile hbonds_AB_ori1_run1_WR.txt -type all

hbonds -sel1 $sel_chA -sel2 $sel_chC -writefile yes -dist 3.5 -ang 30 -
outfile hbonds_AC_ori1_run1_WR.txt -type all

hbonds -sel1 $sel_chA -sel2 $sel_chD -writefile yes -dist 3.5 -ang 30 -
outfile hbonds_AD_ori1_run1_WR.txt -type all

hbonds -sel1 $sel_chB -sel2 $sel_chC -writefile yes -dist 3.5 -ang 30 -
outfile hbonds_BC_ori1_run1_WR.txt -type all

hbonds -sel1 $sel_chB -sel2 $sel_chD -writefile yes -dist 3.5 -ang 30 -
outfile hbonds_BD_ori1_run1_WR.txt -type all

hbonds -sel1 $sel_chC -sel2 $sel_chD -writefile yes -dist 3.5 -ang 30 -
outfile hbonds_CD_ori1_run1_WR.txt -type all
```

Appendix 5 RMSD measurements by “RMSD_protein.tcl”.

```
set rodChains {A B C D}
foreach x $rodChains {

set fl_5 [open "rmsd_four_mons_ori4_int_run3_WR_100ns_Ch_$x .txt" w]
puts $fl_5 "fr_four_mons_ori4_int_run3_100ns_Ch_$x
rmsd_four_mons_ori4_int_run3_WR_100ns_Ch_$x"
# use frame 0 for the reference
set reference [atomselect top "protein and name CA and chain $x" frame 0]
# the frame being compared
set compare [atomselect top "protein and name CA and chain $x"]

set num_steps [molinfo top get numframes]

for {set frame 0} {$frame < $num_steps} {incr frame} {

    # get the correct frame
    $compare frame $frame
    # compute the transformation
    set trans_mat [measure fit $compare $reference]
    # do the alignment
    $compare move $trans_mat
    # compute the RMSD
    set rmsd [measure rmsd $compare $reference]
    # print the RMSD
    puts $fl_5 "$frame \t $rmsd"

}
close $fl_5
}
```

Appendix 6 RMSF measurements by “RMSF_protein.tcl”.

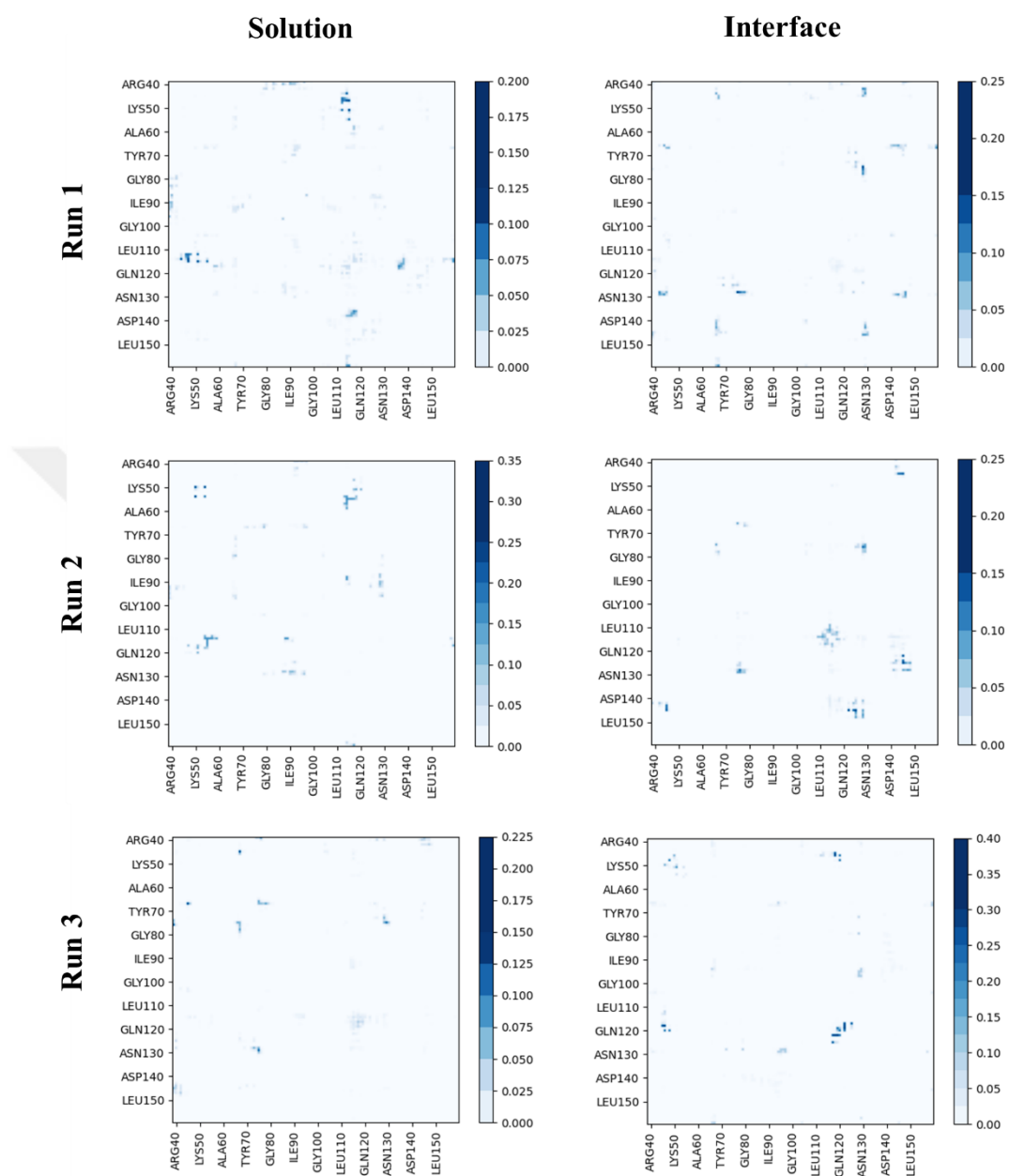
```
# use frame 0 for the reference
set reference [atomselect top "protein and name CA and chain A" frame 0]
# the frame being compared
set compare [atomselect top "protein and name CA and chain A"]
set num_steps [molinfo top get numframes]
set fp_5 [open "RMSF_TWO_Mon_ori2_int_run3_ChainA.txt" w]
puts $fp_5 "fr_TWO_Mon_ori2_int_run3_ChainA
RMSF_TWO_Mon_ori2_int_run3_chainA"

for {set frame 0} {$frame < $num_steps} {incr frame} {

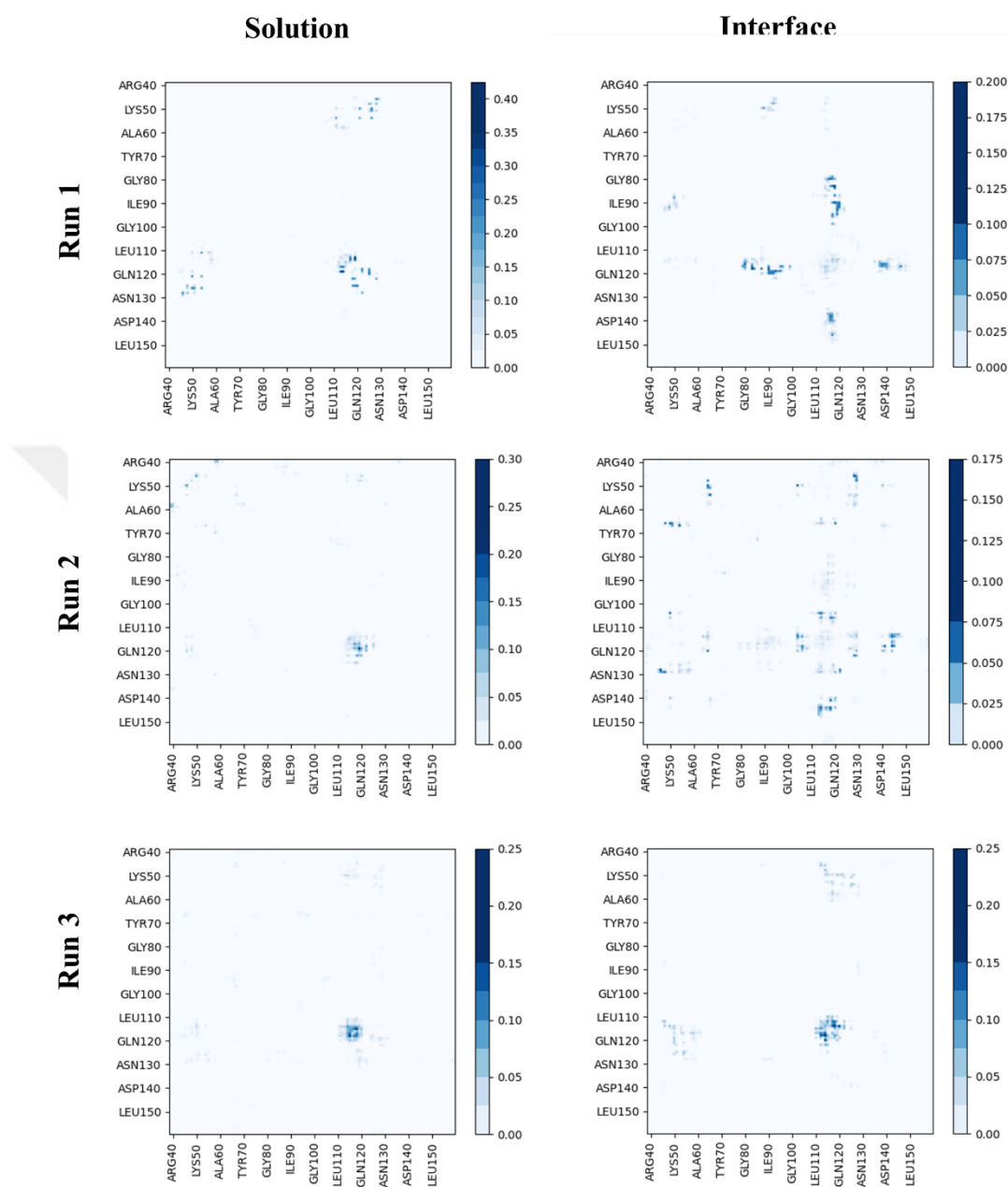
    # get the correct frame
    $compare frame $frame

    # compute the transformation
    set trans_mat [measure fit $compare $reference]
    # do the alignment
    $compare move $trans_mat
}
for {set i 0} {$i < [$compare num]} {incr i} {
    set rmsf [measure rmsf $compare first 0 last 499 step 1]
    puts $fp_5 "[expr {$i+1}]    [lindex $rmsf $i]"
}
close $fp_5
```

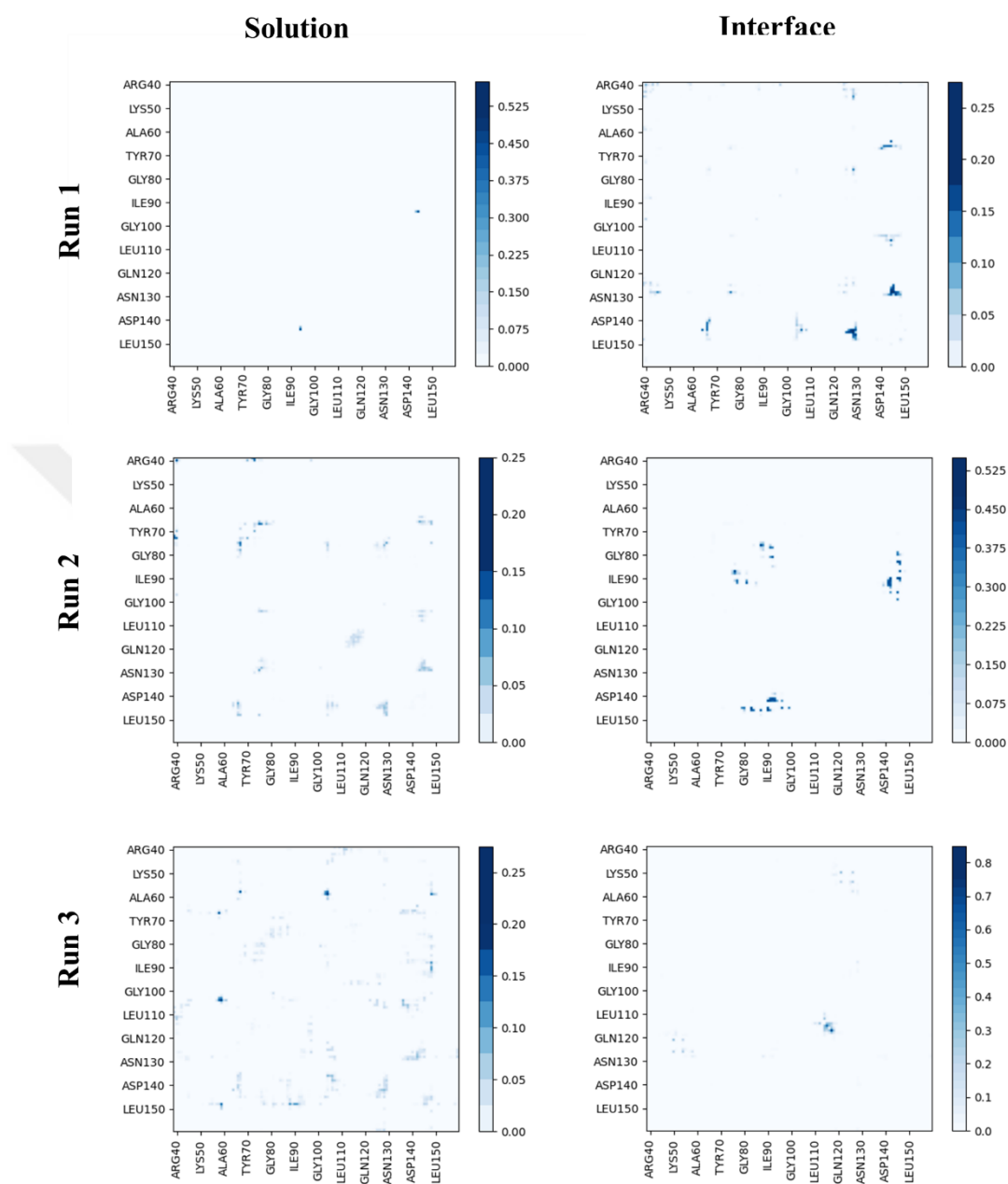
Appendix 7 Contact maps of four monomers in Orientation 1 simulations.



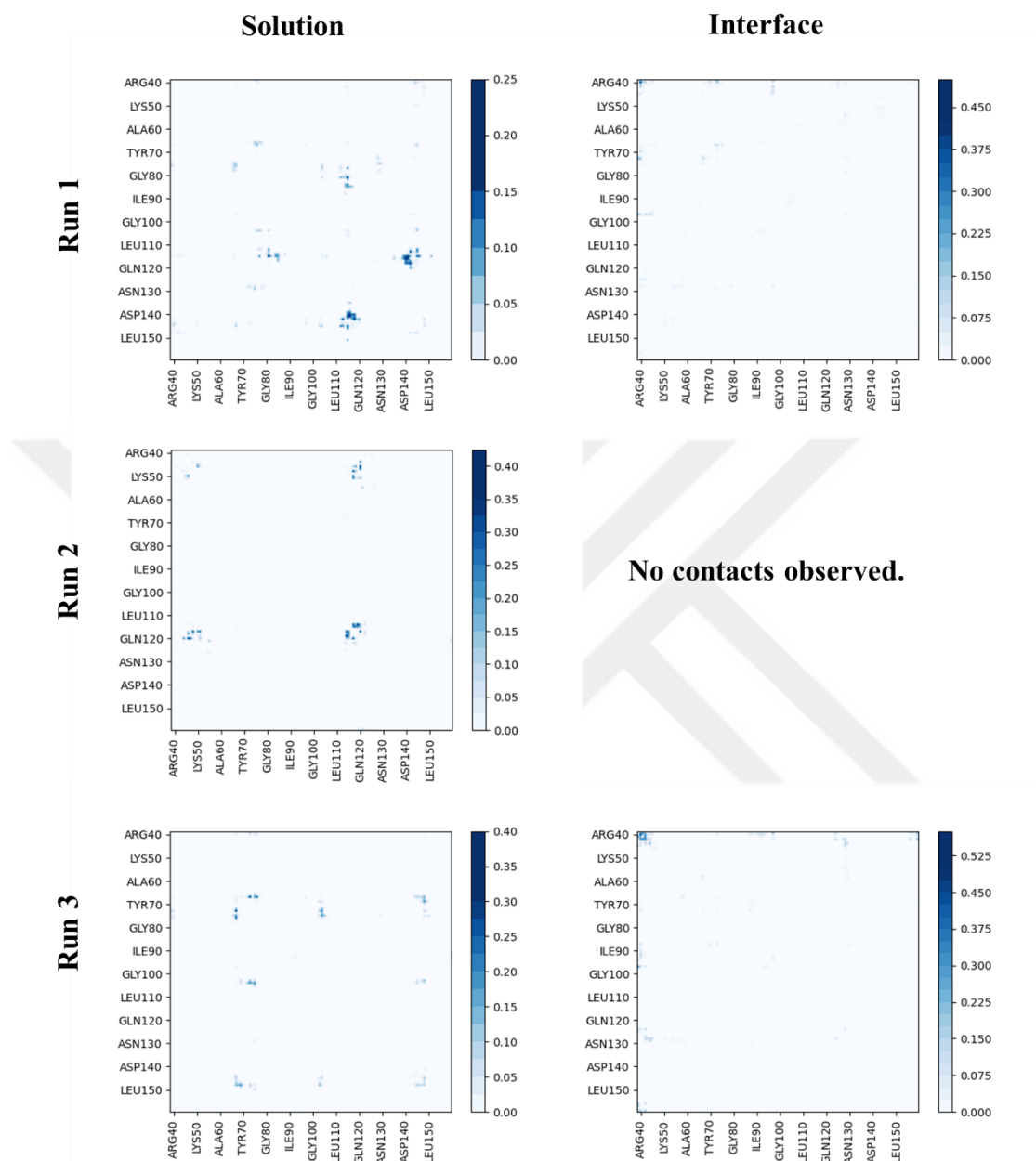
Appendix 8 Contact maps of four monomers in Orientation 2 simulations.



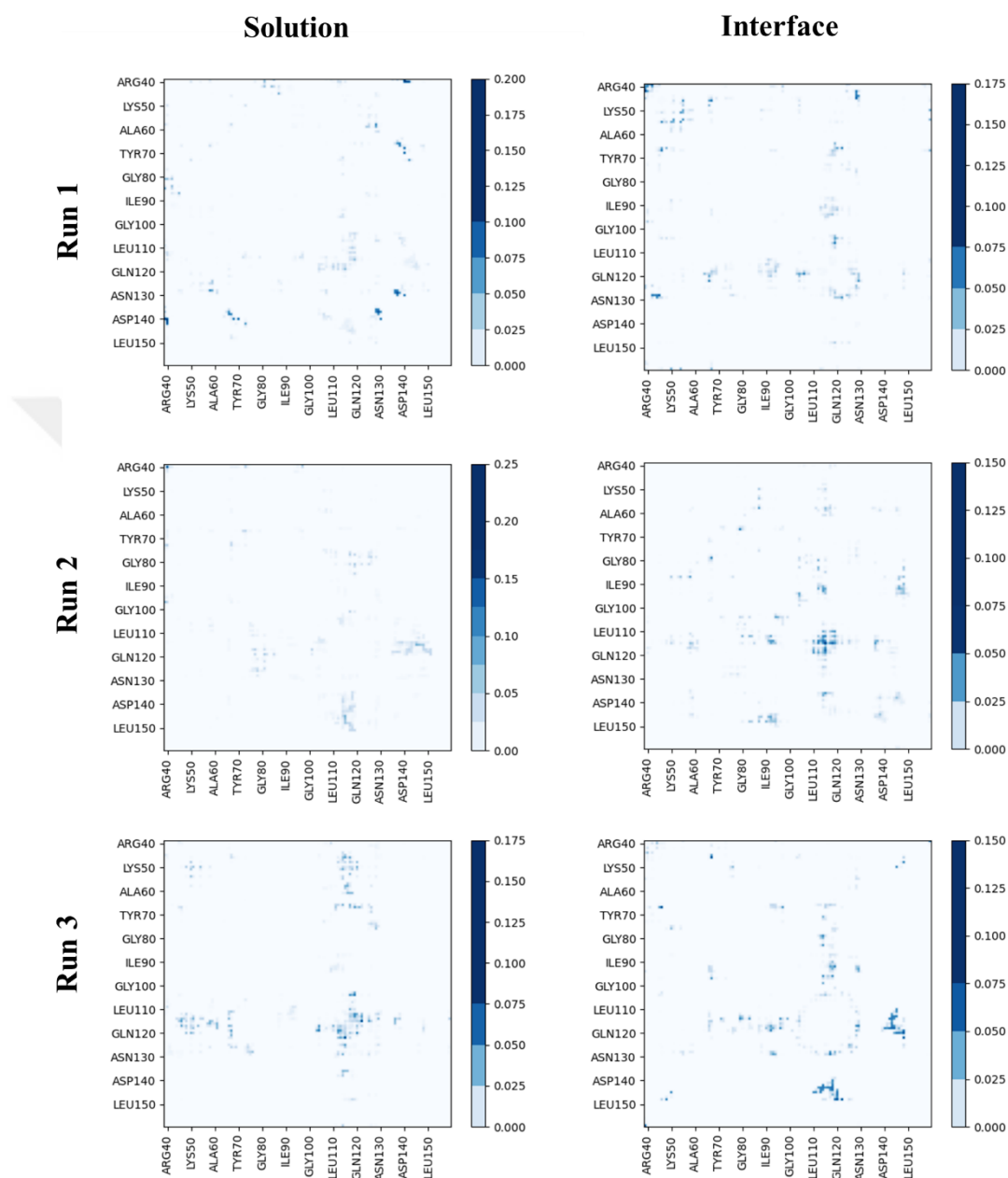
Appendix 9 Contact maps of four monomers in Orientation 3 simulations.



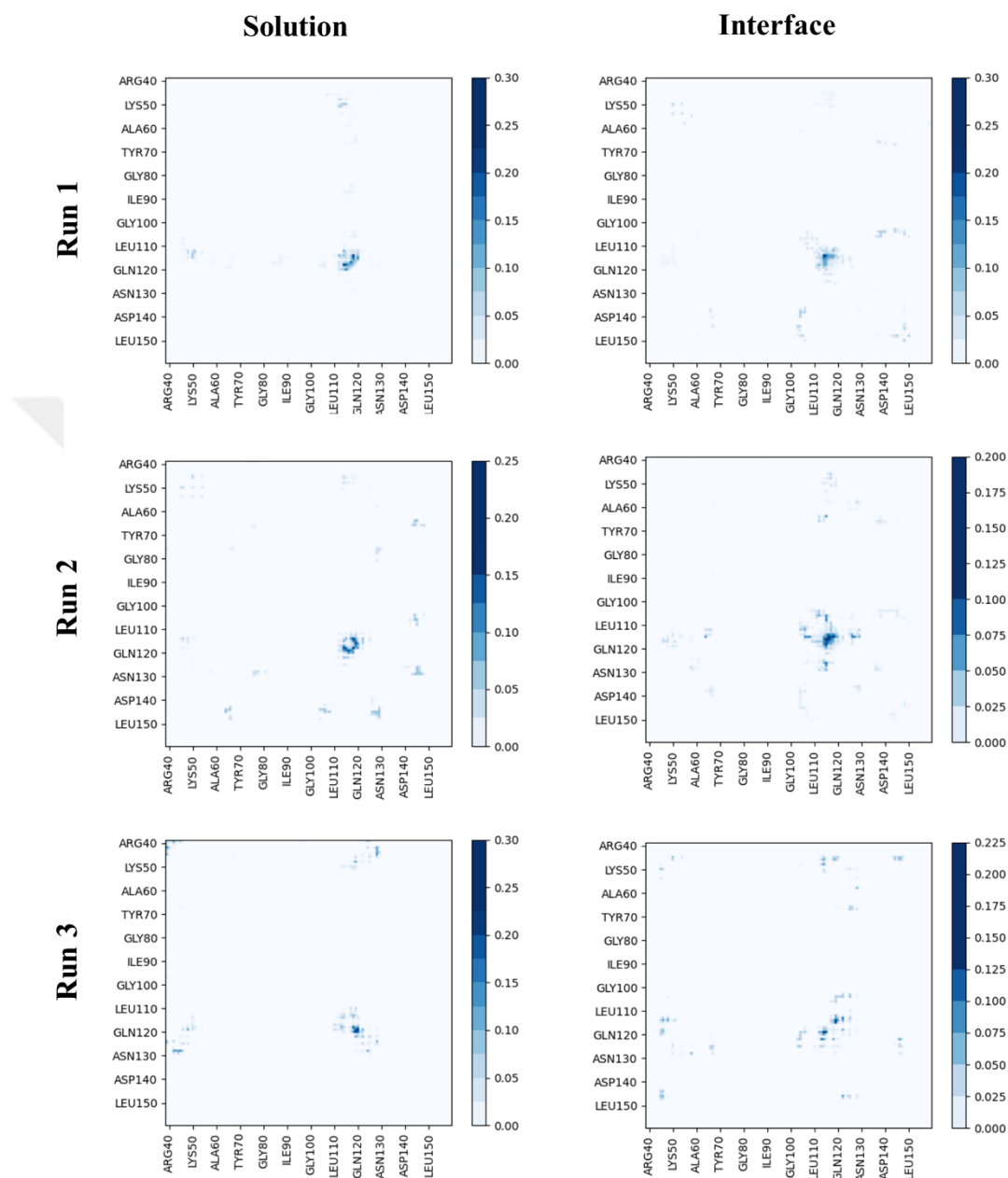
Appendix 10 Contact maps of four monomers in Orientation 4 simulations.



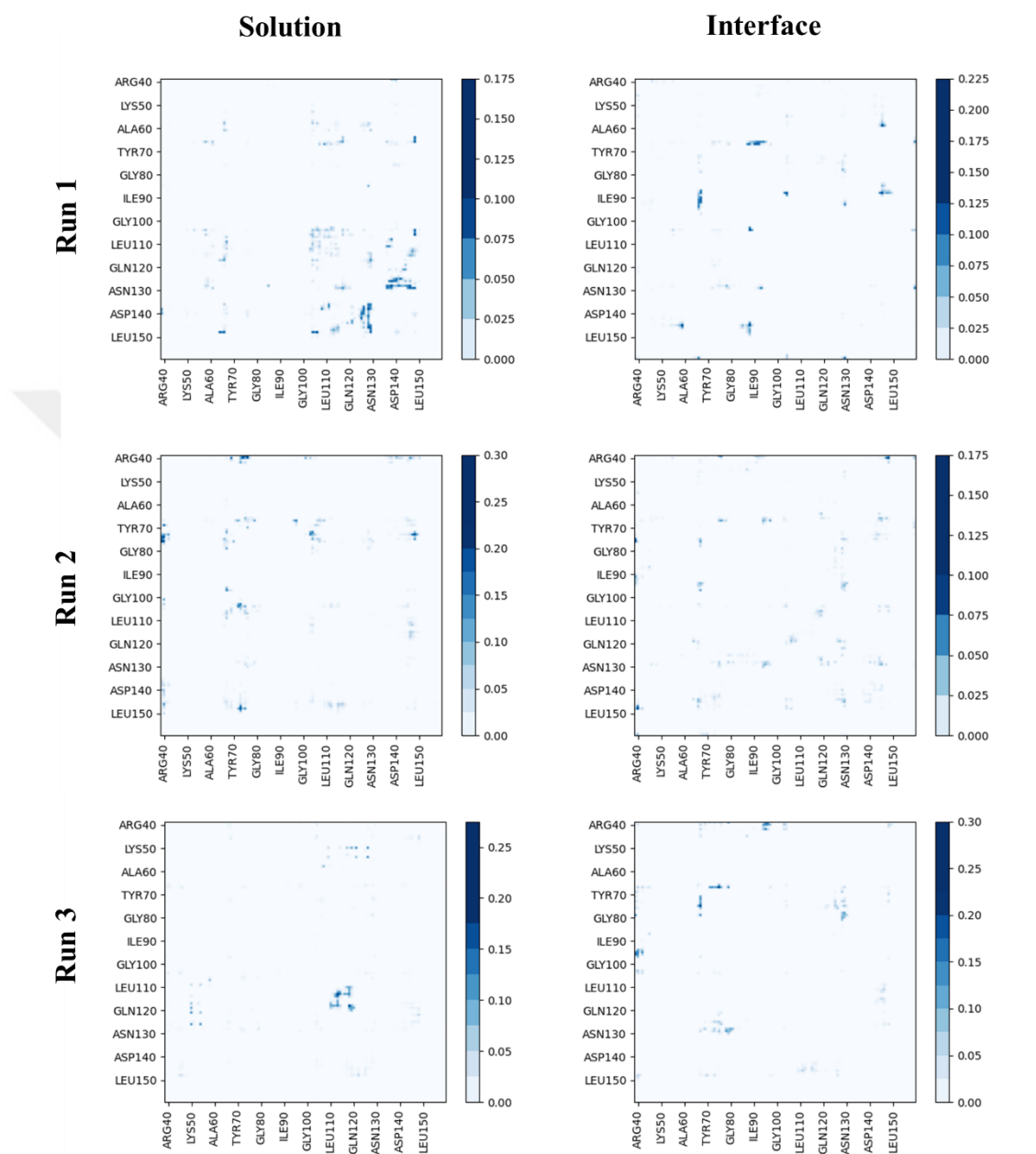
Appendix 11 Contact maps of four monomers in Orientation 1 simulations with positionally RWM.



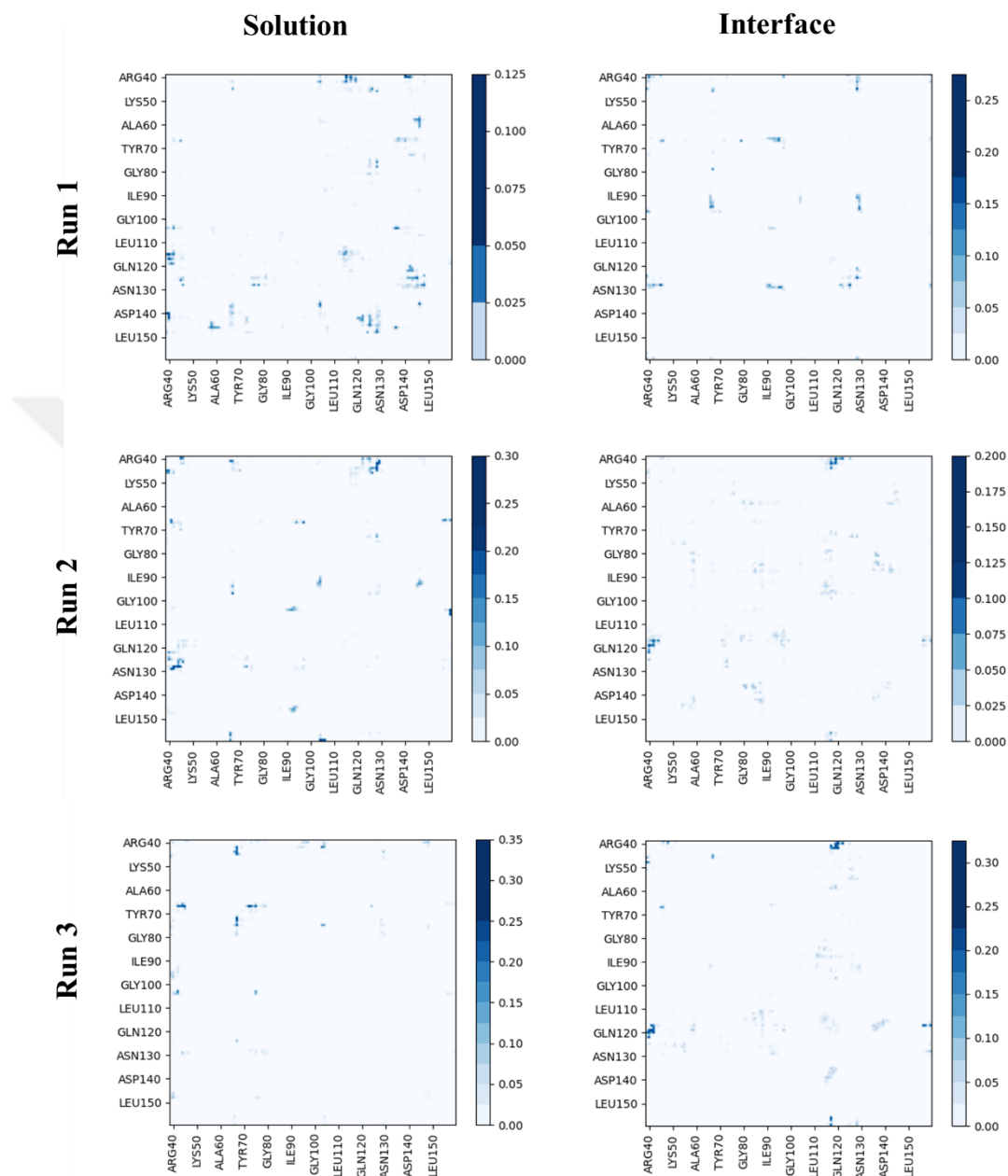
Appendix 12 Contact maps of four monomers in Orientation 2 simulations with positionally RWM.



Appendix 13 Contact maps of four monomers in Orientation 3 simulations with positionally RWM.



Appendix 14 Contact maps of four monomers in Orientation 4 simulations with positionally RWM.



9 CURRICULUM VITAE



