

A Molecular Dynamics Study into Annexin A1 Induced Membrane Binding and Aggregation

by

Matthew P. Donohue

July, 2010

Chair: Rickey P. Hicks

Department of Chemistry

Annexins constitute a family of proteins that bind to anionic membranes in a reversible and calcium dependent manner through the unique architecture of their calcium binding sites. In addition, annexins with relatively large N-terminal domains have been identified to cause membrane aggregation and fusion. There is a contradiction between x-ray crystallography and cryo-EM studies as to the proposed mechanism of annexin-induced membrane aggregation.

Molecular dynamics simulations were performed in an effort to study the calcium dependent binding of annexin I to a phospholipid bilayer and to investigate the N-terminus as a possible second membrane binding site. Site specific mutations were created on the N-terminus to study the effects phosphorylation has on the tertiary structure of the protein. Simulation trajectories were analyzed in terms of non-bonded interaction energies of protein residues, root mean square deviations of the protein backbone, root mean square fluctuations of residues and nuclear distances between calcium ions and their oxygen ligands. Calcium coordination with lipid headgroups was observed in repeat IV of the core domain. Two lysine residues located in the N-terminus and speculated to be crucial to membrane aggregation displayed significant electrostatic

attractions to the phospholipid layer based on MM-PBSA calculations. This thesis will present a model for the mechanism of interaction between annexin A1 and membranes.

A Thesis
A Molecular Dynamics Study into Annexin A1 Induced Membrane Binding and
Aggregation

Presented to
the Faculty in the Department of Chemistry
East Carolina University

In Partial Fulfillment
of the Requirements for the Degree
Master of Science

by
Matthew P. Donohue
July, 2010

© Matthew P. Donohue, 2010

A Molecular Dynamics Study into Annexin A1 Induced Membrane Binding and Aggregation

by

Matthew P. Donohue

Approved by:

Director of Thesis: _____
Yumin Li, PhD

Committee Member: _____
Libero Bartolotti, PhD

Committee Member: _____
Andrew L. Sargent, PhD

Committee Member: _____
Ruth Schwalbe, PhD

Chair of the Department
of Chemistry: _____
Rickey P. Hicks, PhD

Dean of the
Graduate School: _____
Paul J. Gemperline, PhD

TABLE OF CONTENTS

LIST OF TABLES.....	v
LIST OF FIGURES.....	vi
LIST OF ABBREVIATIONS.....	ix
CHAPTER 1 – BACKGROUND.....	1
1.1: Annexins - Introduction.....	1
1.2: Annexins - Classification.....	2
1.3: Annexins - Functional Aspects.....	3
1.4: Annexins - Physiological Importance.....	4
1.5: Cryo-Electron Microscopy of Junctions Formed by Annexins	5
(I) Junctions formed by annexin II and annexin I, Ca ²⁺ and DOPG-DOPC liposomes.....	6
1.6: X-ray Crystallography Studies of Annexin 1.....	8
(I) Structure of full length Annexin I.....	8
(II) Observations made in this study.....	9
1.7: Conflict Between X-ray Studies and Cryo-EM Studies.....	11
(I) Problem Statement.....	12
1.8: N-Terminal Domain.....	13
1.9: Core Domain.....	17
CHAPTER 2 - RESEARCH PLAN.....	20
2.1: Approach.....	20
2.2: Theory of Molecular Dynamics Simulation.....	21
2.3: Molecular Dynamics Simulations on Annexins.....	27
CHAPTER 3 - METHODOLOGY.....	34
3.1: Computational Aspects of MD simulations.....	34
3.2: Force Field.....	35
3.3: Amber.....	36
(I) Preparatory Programs	3

(II)	Simulation Programs	38
(III)	Analysis Programs.....	38
3.4:	Systems Used in Molecular Dynamics	
	Simulations.....	40
(I)	Annexin Construction.....	40
(II)	Phospholipid Construction.....	41
3.5:	Simulation Conditions.....	42
(I)	Step 1: Energy Minimization.....	43
(II)	Step 2: MD – Warming.....	44
(III)	Step 3: Constant Pressure Dynamics (NPT).....	44
(IV)	Step 4: Constant Volume Dynamics (NVT).....	45
(V)	Step 5: Analysis of Trajectories.....	45
3.6:	Simulation Protocol.....	47
CHAPTER 4 - RESULTS.....		48
4.1:	Analysis of Systems Used in Simulations.....	48
4.2:	Simulation Times.....	48
4.3:	Steps – MD Simulation.....	49
(I)	Minimization.....	49
(II)	Equilibration.....	49
(III)	Constant Pressure Dynamics (NPT).....	49
(IV)	Constant Volume Dynamics (NVT).....	51
4.4:	Root Mean Square Deviations (RMSD).....	51
(I)	RMSD plot for full length annexin I in setup 1.....	51
(II)	RMSD plot of core domain in setup 1.....	54
(III)	RMSD plot by core repeats in setup 1.....	57
(IV)	RMSD plot of N-terminal in setup 1.....	58
(V)	RMSD plot for full length annexin I in setup 2.....	59
(VI)	RMSD plot of the core domain in setup 2.....	60
(VII)	RMSD plot by core repeats in setup 2.....	61
(VIII)	RMSD plot of N-terminal in setup 2.....	62

4.5:	Conclusions Drawn from RMSD Plots.....	64
4.6:	Root Mean Square Fluctuations (RMSF).....	64
4.7:	Observations Made from RMSF Plots.....	67
4.8:	B Factors.....	68
4.9:	MM-PBSA.....	70
4.10:	Observations Made from MM-PBSA Plots.....	84
4.11:	Inter Atomic Distances.....	86
4.12:	Observations Made in the Inter-Atomic Distance Plots	95
4.13:	Visualization of the MD Simulation Trajectories.....	95
4.14:	Observations Made from the Visualization of MD Simulation Trajectories.....	98
CHAPTER 5 - MUTATION STUDIES.....		98
5.1:	Background.....	98
5.2:	Methods.....	99
5.3:	Results.....	101
5.4:	Visualization of MD simulation trajectories.....	106
5.5:	Implications of Mutation Studies.....	109
CHAPTER 6 - DISCUSSION.....		111
6.1:	Residue Flexibility.....	111
6.2:	Anchor Residues.....	112
6.3:	Calcium Bridging Interactions.....	115
6.4:	Phosphorylation Implications	115
6.5:	Stabilization of the N-terminal.....	116
6.6:	Conclusions.....	117
REFERENCES.....		120

LIST OF TABLES

Table 1.1:	Classification of Human Annexins	2
Table 1.2:	Binding Site Geometries of Annexins.....	18
Table 1.3:	Calcium ions in annexin I and their associated ligands.....	19
Table 3.1:	Dimensions of each system set-up, including total number of atoms used per simulation and timescale of each setup.....	40
Table 4.1:	Simulation Time Scales.....	49
Table 4.2:	Setup 1 MM-PBSA results of charged residues from 10240-11150 ps.....	74
Table 4.3:	Setup 1 MM-PBSA results of charged residues from 11770-13330 ps.....	76
Table 4.4:	Setup 1 MM-PBSA results of charged residues from 14540-15575 ps.....	77
Table 4.5:	Setup 1 MM-PBSA results of charged residues from 17015-17640 ps.....	79
Table 4.6:	Setup 2 MM-PBSA results of charged residues from 10962-12014 ps.....	80
Table 4.7:	Setup 2 MM-PBSA results of charged residues from 13672-14771 ps.....	82
Table 4.8:	Setup 2 MM-PBSA results of charged residues from 16922-18469 ps.....	83
Table 5.1:	Dimensions of each full-length protein system, including total number of atoms used per simulation and timescale of each system.....	101
Table 5.2:	Dimensions of each peptide system, including total number of atoms used per simulation and timescale of each system.....	101

LIST OF FIGURES

Fig 1.1:	Cryo EM junctions of DOPG-DOPC vesicles.....	7
Fig 1.2:	Electron density profile of a six stripe junction.....	7
Fig 1.3:	Plausible fusion mechanisms.....	11
Fig 1.4:	Axial view of Annexin A1.....	20
Fig 3.1:	Basic information flow in AMBER.....	37
Fig 3.2:	Calcium-bound annexin A1 with exposed N-terminal.....	41
Fig 3.3:	DOPC and DOPG used in the simulation.....	42
Fig 4.1:	Density plot for setup 1.....	50
Fig 4.2:	Density plot for setup 2.....	50
Fig 4.3:	RMSD plot of full-length protein setup 1.....	53
Fig 4.4:	Distance plot between Asp-197 and Ala-2.....	54
Fig 4.5:	RMSD plot of core domain setup1.....	55
Fig 4.6:	RMSF plot of α carbon atoms during 15315-16420 ps, Setup 1.....	56
Fig 4.7:	RMSF plot of α carbon atoms during 16420-18430 ps, Setup 1.....	56
Fig 4.8:	RMSD plot of annexin repeats setup 1.....	57
Fig 4.9:	RMSD plot of N-terminal domain setup 1.....	58
Fig 4.10:	RMSD plot of N-terminal helices setup 1.....	59
Fig 4.11:	RMSD plot of full-length protein setup 2.....	60
Fig 4.12:	RMSD plot of core domain setup 2.....	61
Fig 4.13:	RMSD plot of annexin repeats setup 2.....	62
Fig 4.14:	RMSD plot of N-terminal domain setup 2.....	63
Fig 4.15:	RMSD plot of N-terminal helices setup 2.....	63
Fig 4.16:	RMSF plot of backbone atoms setup 1.....	65
Fig 4.17:	RMSF plot of side-chain atoms setup 1.....	65
Fig 4.18:	RMSF plot of backbone atoms setup 2.....	66
Fig 4.19:	RMSF plot of side-chain atoms setup 2.....	66

Fig 4.20:	B factor of setup 1.....	69
Fig 4.21:	B factor of setup 2.....	69
Fig 4.22:	Regions of the RMSD plot of setup 1 in which MM-PBSA calculations were performed.....	72
Fig 4.23:	Regions of the RMSD plot of setup 2 in which MM-PBSA calculations were performed.....	73
Fig 4.24:	MM-PBSA calculation of N-terminus with phospholipid layer at 10240-11150 ps in setup 1.....	73
Fig 4.25:	MM-PBSA calculation of N-terminal with core domain at 10240-11150 ps in setup 1.....	74
Fig 4.26:	MM-PBSA calculation of N-terminus with phospholipid layer at 11770-13330 ps in setup 1.....	75
Fig 4.27:	MM-PBSA calculation of N-terminal with core Domain at 11770-13330 ps in setup 1.....	75
Fig 4.28:	MM-PBSA calculation of N-terminal with phospholipid layer at 14540-15575 ps in setup 1.....	76
Fig 4.29:	MM-PBSA calculation of N-terminus with core domain at 14540-15575 ps in setup 1.....	77
Fig 4.30:	MM-PBSA calculation of N-terminus with phospholipid layer at 17015-17640 ps in setup 1.....	78
Fig 4.31:	MM-PBSA calculation of N-terminus with core domain at 17015-17640 ps in setup 1.....	78
Fig 4.32:	MM-PBSA calculation of N-terminus with phospholipid Layer at 10692-12014 ps in setup 2.....	79
Fig 4.33:	MM-PBSA calculation of N-terminus with core domain at 10692-12014 ps in setup 2.....	80
Fig 4.34:	MM-PBSA calculation of N-terminus with phospholipid layer at 13672-14771 ps in setup 2.....	81
Fig 4.35:	MM-PBSA calculation of N-terminus with core domain at 13672-14771 ps in setup 2.....	81

Fig 4.36:	MM-PBSA calculation of N-terminus with phospholipid layer at 16922-18469 ps in setup 2.....	82
Fig 4.37:	MM-PBSA calculation of N-terminus with core domain at 16922-18469 ps in setup 2.....	83
Fig 4.38:	Calcium 1 – ligand distances.....	87
Fig 4.39:	Calcium 2 – ligand distances.....	88
Fig 4.40:	Calcium 3 – ligand distances.....	89
Fig 4.41:	Calcium 4 – ligand distances.....	90
Fig 4.42:	Calcium 5 – ligand distances.....	91
Fig 4.43:	Calcium 6 – ligand distances.....	92
Fig 4.44:	Calcium 7 – ligand distances.....	93
Fig 4.45:	Calcium 8 – ligand distances.....	94
Fig 4.46:	Snapshots of setup 1.....	96
Fig 4.46:	Snapshots of setup 2.....	97
Fig 5.1:	RMSD plot of wild-type protein.....	102
Fig 5.2:	RMSD plot of S28A mutant.....	102
Fig 5.3 :	RMSD plot of S28E mutant.....	103
Fig 5.4:	RMSD plot of wild-type core domain.....	103
Fig 5.5:	RMSD plot of S28A mutant core domain.....	104
Fig 5.6:	RMSD plot of S28E mutant core domain.....	104
Fig 5.7:	RMSF overlay of proteins.....	105
Fig 5.8:	RMSF overlay of proteins core domain close-up.....	105
Fig 5.9:	Snapshots of full-length proteins in mutation study.....	107
Fig 5.10:	Snapshots of peptides in mutation study.....	108
Fig 6.1:	Orientation of K-26 and K-29 snapshots.....	114

LIST OF ABBREVIATIONS

PLA ₂	Phospholipase A2.....	4
AA	Arachidonic Acid.....	4
COX-2	Cyclooxygenase 2.....	4
tPA	Tissue Plasminogen Activator.....	5
EM	Electron Microscopy.....	5
DOPG	Di-oleoylphosphatidylglycerol.....	6
DOPC	Di-oleoylphosphatidylcholine.....	6
MD	Molecular Dynamics.....	20
BPTI	Bovine Pancreatic Trypsin Inhibitor.....	22
RMSD	Root Mean Square Deviation	34
CHARMM	Chemistry at HARvard Molecular Mechanics.....	36
GROMACS	GRoningen MACHine for Chemical Simulations.....	36
GROMOS	GRoningen MOlecular Simulation.....	36
OPLS	Optimized Potentials for Liquid Simulations	36
AMBER	Assisted Model Building with Energy Refinement.....	36
Sander	Simulated Annealing with NMR-Derived Energy Restrictions.....	38
MMPBSA	Molecular Mechanics Poisson-Boltzmann Surface Area.....	38
RMSF	Root Mean Square Fluctuation.....	39
NPT	Constant pressure dynamics.....	44
NVT	Constant volume dynamics.....	45

CHAPTER 1 - BACKGROUND

1.1: Annexins - Introduction

Annexins constitute a family of proteins that have the ability to bind onto negatively charged phospholipid bilayers in a reversible and calcium dependent manner through the unique architecture of their calcium binding sites. This property places annexins at the nexus of many membrane related events, including certain endo-, exo-, and phagocytic steps as well as membrane trafficking and reorganization. The annexins are expressed throughout eukaryotic phyla, with at least nine members of the family identified in mammalian tissues, although they have been found to be absent in yeasts and prokaryotes.¹

The name annexin is derived from the Greek *annex* meaning “bring/hold together”, which accurately describes nearly all annexins in their ability to bind biological structures, particularly membranes.

Annexins have been extensively studied over the last several decades, and much progress has been made in defining the sequence and structure of many of these proteins from crystallography results. Furthermore, it has recently become clear that abnormalities in annexin expression and activity cause human diseases, and the term *annexinopathies* has been coined.¹ Despite detailed research, specific physiological functions still need to be assigned to individual annexins.

1.2: Annexins - Classification

Since the turn of the century, more than 160 unique annexins have been discovered in over 65 different species. The best-studied subfamily of annexins is from vertebrate animals, designated annexin group A. The following Table 1.1 represents the annexin nomenclature. All annexins are comprised of a core domain which contains four homologous repeats (in annexin VI, eight) of five α -helices and the divergent N-terminus, which precedes the core and is unique for a given member of the family. The N-terminal domains vary in length and sequence, from 19 or fewer residues (Annexins A3, A4, A5, A6, A10, A12, and A13b), to 30-100 residues (Annexins A1, A2, A7, and A11).¹

Table 1.1: Classification of Human Annexins.¹

Name	Synonyms/Former name(s)	Human gene symbol
Annexin A1	Lipocortin 1, annexin I	ANXA1
Annexin A2	Calpactin 1, annexin II	ANXA2
Annexin A3	Annexin III	ANXA3
Annexin A4	Annexin IV	ANXA4
Annexin A5	Annexin V	ANXA5
Annexin A6	Annexin VI	ANXA6
Annexin A7	Synexin, Annexin VII	ANXA7
Annexin A8	Annexin VIII	ANXA8
Annexin A9	Annexin XXXI	ANXA9
Annexin A10		ANXA10
Annexin A11	Annexin XI	ANXA11
Annexin A12	unassigned	
Annexin A13	Annexin XIII	ANXA13

1.3: Annexins - Functional Aspects

Annexins bind in a calcium dependent manner to negatively charged phospholipids. In addition to membrane binding, annexins with relatively larger N-terminal domains (between 30-100 residues) have been identified to cause membrane aggregation and fusion events. The three-dimensional structures solved so far indicate that the N-terminus is located on the concave side of the protein, opposite the calcium binding sites. These N-terminal domains harbor binding sites for S100 proteins and various phosphorylation sites for serine/threonine and tyrosine-specific kinases.¹ S100 proteins are a multigene family characterized by two calcium-binding sites of the EF-hand type conformation.

Annexins III and V have short N-terminal domains and therefore cannot participate in membrane aggregation. In annexin A3, replacing Trp-5 by alanine has showed a direct effect of the N-terminal domain on properties displayed by the core. The W5A mutant protein shows a much stronger phospholipid binding and a more disordered N-terminal domain. This finding suggests that subtle differences in the N-terminal domain result in significantly altered properties of the protein, possibly explaining the functional diversity among otherwise highly conserved annexins. Additionally, a chimera comprising the core domain of annexin A5 (which does not promote membrane fusion activity) fused with the N-terminus of annexin A1 was found to cause membrane fusion by Andree and co-workers. Thus, differences in length and sequence between individual annexins have displayed altered functions between members of this protein family.¹

1.4: Annexins - Physiological Importance

To date, no human disease has been described in which a mutation in the annexin gene is the primary cause. However, there is good evidence to suggest that changes in annexin expression or localization may contribute to the pathogenesis of certain disease phenotypes. Thus, annexins have been implicated in some of the more serious human diseases, including cancer, cardiovascular disease and diabetes. In this way, annexins have significant value in disease prognosis, diagnosis and therapy.

The term “annexinopathies” was coined by J. H. Rand in 1999 to classify diseases related to annexin abnormalities. He and co-workers investigated the over-expression of annexin A2 in the leukocytes of patients having a hemorrhagic form of acute promyelocytic leukemia. Also annexin A5 was found to be under-expressed on placental trophoblasts of patients with antiphospholipid syndrome and in preeclampsia.²

Much work has been done on annexin A1 in eliciting its role as an anti-inflammatory agent by Parente and Solito. They found that annexin A1 inhibits the expression of inducible nitric oxide synthase in macrophages and inhibition of phospholipase A₂ (PLA₂). PLA₂ hydrolyzes the sn-2 ester bond of a phospholipid, releasing a fatty acid (usually arachidonic acid (AA)). AA is oxidized by cyclooxygenase-2 (COX-2) to produce eicosanoids and prostaglandins. Prostaglandins in turn signal the inflammatory cascade. Much of the work in this study used animal models for rheumatoid arthritis and myocardial infarct.³ Although valuable, this work still cannot be directly related with human disease.

Annexin A2, and its binding ligand S100A10, have been reported to be secreted on the surface of peripheral blood monocytes where they regulate fibrinolysis. The

annexin A2 tetramer binds both plasminogen and tissue plasminogen activator (tPA) and facilitates the release of plasmin. Annexin A2 was also found to be competitively inhibited by both thrombin and angiostatin. These roles place the annexin A2 tetramer as a significant part of fibrinolytic pathways, thrombosis, and angiogenesis.⁴

Some annexins are over-expressed in certain tumor types, while other annexins are under-expressed. Annexin A1 has found to be over-expressed in oesophageal adenocarcinoma, gastric adenocarcinoma, colorectal adenocarcinoma, clear cell renal carcinoma, hairy cell leukemia, and others. Down-regulation of annexin A1 was reported in many types of cancers including head and neck, breast, and prostate.⁴

Although the precise functions of annexins at the molecular level are still not understood, several cellular processes in which they are involved in are becoming apparent. A direct correlation between annexin anomalies and disease symptoms is still missing, as laboratory data and clinical data must be closer linked.

1.5: Cryo-Electron Microscopy of Junctions Formed by Annexins

Cryo-electron microscopy (Cryo-EM) studies reported several types of junctions formed connecting liposomes and several annexins, including the annexin II-p11 heterotetramer, monomeric annexin A1 and annexin A2.⁵ The annexin II-p11 tetramer is a stable complex formed between annexin II and p11, a member of the S100 Ca²⁺-binding protein family. S100 proteins are a multigene family characterized by two calcium-binding sites of the EF-hand type conformation. A short amphiphatic α -helix segment located on the 33 residue N-terminus of annexin II provides the binding site for p11. Dimeric interactions (p11-p11) stabilize the tetrameric complex. As my research

involves the use of monomeric annexin I, I will summarize the results of the cryo-EM experiment pertaining to monomeric annexins only.

(I) Junctions Formed by Annexin II and Annexin I, Ca^{2+} and DOPG-DOPC Liposomes

Monomeric annexin I and annexin II were observed to aggregate liposomes *via* similar junctions. The liposomes were negatively charged and composed of DOPG (dioleoylphosphatidylglycerol) –DOPC (dioleoylphosphatidylcholine) lipids in a 1:4 molar ratio. Two types of junctions were reported: the 160 Å thick six stripe junctions were more frequently observed than the 125 Å thick five stripe junctions, as shown in Figure 1.1. Monomeric annexin II formed junctions less frequently than the annexin II-p11 tetramer. Aggregates of liposomes were observed at 1.5 mM Ca^{2+} . In regards to the six stripe junctions, Lambert and co-workers interpreted the two outer stripes of the junction as the inner leaflets of the liposome, the two inner stripes were each assigned as an annexin monomer, and the remaining stripes between the inner leaflet and the protein were due assigned as the outer leaflets of the liposome, as shown in Figure 1.2. The experimenters suggested that each annexin molecule interacts with one liposome by its convex face and with the other annexin by its N-terminus. The distances between the centers-of-mass of the outer bilayer leaflets were shown to be 80 Å (six-stripe junctions), which after accounting for the width of the lipid headgroups corresponds to a protein thickness of 60 Å. Two interpretations were proposed for the five striped junctions: 1) each annexin molecule interacts with two liposomes, by both its convex and concave faces, and 2) annexins interact with liposomes only through their convex faces and junctions are stabilized by annexin-annexin lateral interactions.⁵

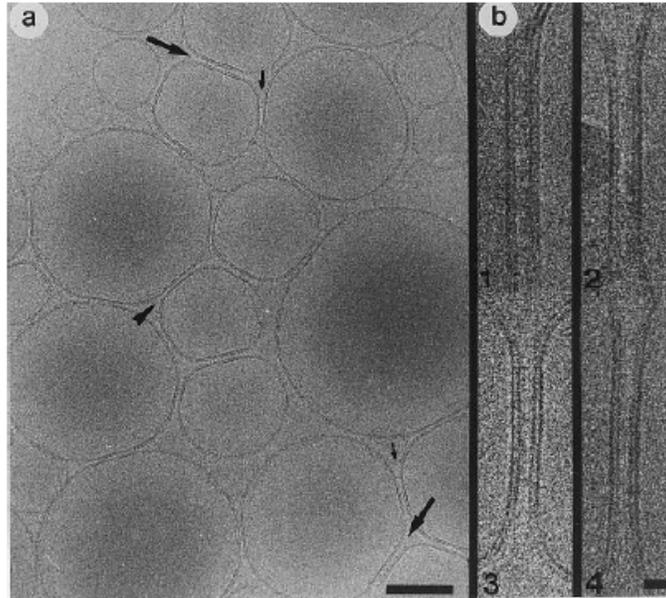


Fig 1.1: Ca^{2+} -induced junctions formed between DOPG-DOPC liposomes and monomeric annexin I. a, aggregates of liposomes; large arrows point to the six striped junctions, smaller arrows point to the five striped junctions. The scale bar represents 1000 Å. b, Junctions 1, 2 and 3 present six high density stripes, junction 4 presents five stripes. The scale bar represents 100 Å.⁵

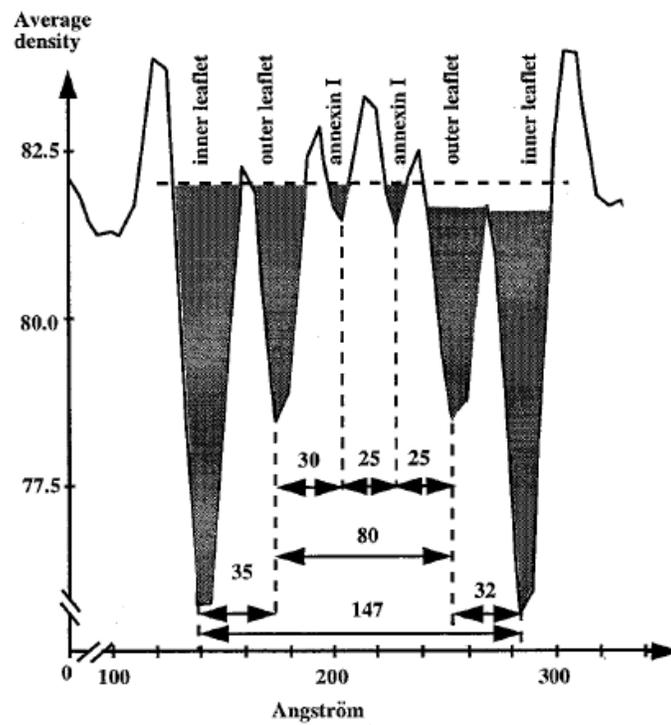


Fig 1.2: Electron density profile of a six stripe junction.⁵

1.6: X-ray Crystallography Studies of Annexin I

Anja Rosengarth and co-workers conducted X-ray studies on annexin I which provided the full-length structure of annexin 1 as well as valuable implications into its mechanism of membrane aggregation. They reported that annexins are comprised of two principal domains: namely the protein core and the N-terminal domain, also known as the 'head'. The former contains the Ca^{2+} binding sites and is responsible for mediating initial membrane binding. The core domain comprises four (in annexin A6 eight) homologous repeats of about 75 amino acid residues that fold into five alpha-helices (A-E) and form an anti-parallel bundle. The crystallized core domain revealed Ca^{2+} type II and type III binding sites on its convex face, which faces the membrane when the protein is peripherally associated with phospholipids. The bound calcium ions in the type II and type III binding sites serve as a bridge to enable annexin interaction with the phospholipid membrane. A central hydrophilic pore proposed to function as a Ca^{2+} ion channel was also revealed in the core domain.⁶

(I) Structure of Full Length Annexin I

Knowledge of the overall structure of annexin I is paramount to understanding the mechanism of how this protein interacts with other macromolecules inside the cell. The previously determined structure of human annexin I lacking the first 32 residues (pdb code 1AIN) is compared to the full length annexin structure (pdb code 1HM6) elucidated in this experiment. The full length structure was obtained from crystal grown in the absence of calcium, and attempts to obtain crystals in the presence of calcium were unsuccessful. An explanation of the failed attempts to crystallize full length annexin in the presence of calcium shall be explained in the following discussion.

(II) Observations Made in the Study

Rosengarth and co-workers identified changes in conformation of the N-terminus and repeat III of the core domain upon comparison of the two annexin I core domains, that is with and without calcium bound. In the inactive form of annexin A1, which is represented in the absence of calcium, the amphipathic alpha helix formed by residues 2-12 of the N-terminal domain is buried inside repeat III of the protein core, which replaces the D-helix. Examination of the electron density for residues 2-8 reveals favorable packing of hydrophobic residues of the N-terminal domain (Met3, Val4, Phe7) into a hydrophobic pocket formed by residues Phe221, Leu225, Phe237, and Val268 of repeat III of the core domain. The D-helix of repeat III unwinds to form a flap over the inserted N-terminal helix. The type II calcium binding site in repeat III is destroyed by the inserted N-terminal helix, because the “cap” residue (Glu255), 39 residues downstream of the calcium-coordinating backbone carbonyl oxygen atoms in the AB loop, is no longer in the proper position for calcium coordination. Upon calcium-mediated membrane binding, the D-flap folds into the proper conformation in repeat III, forming a Ca^{2+} binding site and ejecting the N-terminus so that it becomes solvent accessible. A comparison between the structures of full-length annexin I in the presence and absence of calcium demonstrates that calcium binding in repeat III causes the N-terminus to be expelled from the hydrophobic pocket located in the core domain. The exposed N-terminus is then much more susceptible to proteolysis during crystallization, and this is the reason why attempts to crystallize the full-length protein in the presence of calcium have failed thus far. Such movement could be the precursor for membrane aggregation activity described for annexin A1. It is tempting to speculate that each Ca^{2+} /membrane-

activated annexin 1 molecule could bind to a second membrane *via* its exposed N-terminal domain. Helical wheel analysis showed that an amphipathic helix is formed by the first 11 amino acids. Since this helix is formed largely by the hydrophobic effect, less specificity would be expected explaining the lack of head group specificity for the second lipid binding site of annexin I. This provides a structural explanation of the many studies that describe the activation of a second, distinct phospholipid binding site upon calcium-dependent membrane binding. The model proposes a topology by which a single annexin molecule could simultaneously interact with two membrane bilayers, providing a mechanism for membrane aggregation.⁶

Several mechanisms were proposed for the annexin A1-membrane interaction leading to aggregation. These include: interaction of the N-terminal helix with a second bilayer, dimerization of two annexins *via* their N-terminal domains, and connection of two annexins *via* an S100A11 dimer, which contains two interaction regions for the N-terminal helix, as shown in Figure 1.3.

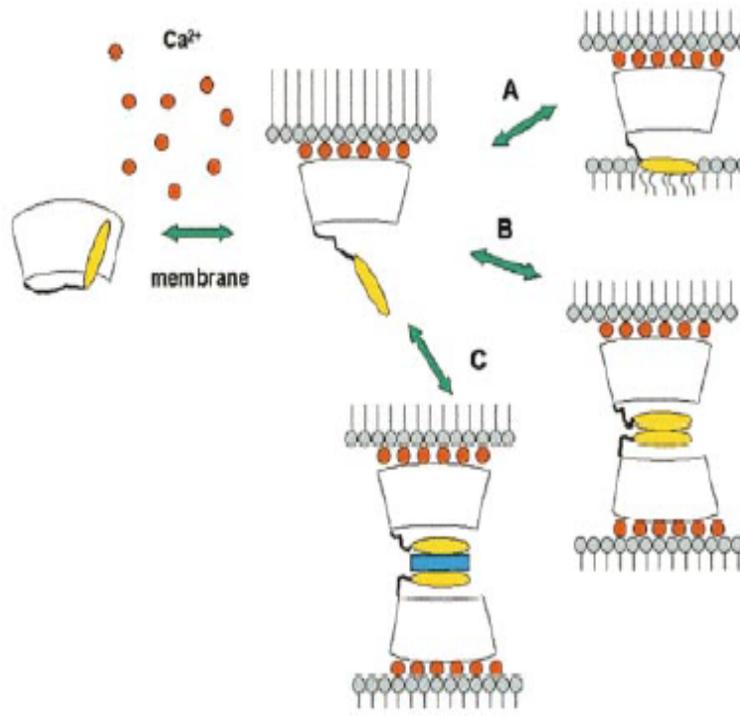


Fig 1.3: Pictorial representation of plausible fusion mechanisms proposed in X-ray studies; (A) Interaction of the N-terminal helix with a second bilayer; (B) Dimerization of two annexins *via* their N-terminal domains; (C) Connection of two annexins *via* an S100A11 dimer, which contains two interaction regions for the N-terminal helix.⁶

1.7: Conflict Between X-ray Studies and Cryo-EM Studies

In vitro cryo-electron microscopy experiments have produced detailed images of annexin-membrane junctions. There is, however, a dispute between cryo-EM and crystallography results regarding the proposed mechanism of annexin-induced membrane aggregation. According to cryo-EM studies, the distance between the lipid headgroups in the six-striped junction of an annexin I induced liposome aggregate measured 60 Å. Lambert *et al.* proposed that an annexin I dimer was responsible for the inner two stripes of the six-striped junction.⁵ Based on crystal structure, an annexin double layer between two lipid bilayers would be at least 70 Å thick (not accounting for the N-terminal domain) and about 90 Å in the case of full length annexin I with a thickness of 45 Å per

monomer. Explanation of the six stripe junctions proposed by Rosengarth described that the two central stripes were due to annexin molecules randomly attached to one or the other bilayer *via* their calcium-binding convex faces, and consequently with their exposed N-terminal domains interacting with the opposing bilayer, resulting in an average appearance of two layers.⁶ The thinner five-stripe junctions could then be interpreted as sections where the majority of the convex faces of annexin are attached to one to the bilayers, resulting in a single highly asymmetric high-density feature between the phospholipid bilayers.^{5,6}

(I) Problem Statement

Two recent studies conducted using Cryo electron microscopy and X-ray crystallography techniques proposed conflicting mechanisms for annexin A1 induced membrane aggregation. My research aims to elucidate the calcium-bridging mechanism and secondary binding site of annexin A1 through analysis of molecular dynamics simulations of an annexin/phospholipid-bilayer system.

1.8: N-Terminal Domain

The unique N-terminal domains vary in length and sequence in different annexins. Due to the structural similarity of the core domains of annexins, the N-terminus of annexin I might be involved in membrane aggregation. The N-terminus of porcine annexin I comprises 41 residues. The three-dimensional structures solved so far indicate that the N-terminus is located on the concave side of the protein, opposite the calcium binding sites. The full-length annexin I crystal structure in the absence of calcium refined at 1.8 Å showed a 42 residue alpha helical N-terminal domain connected to the core domain by a flexible linker. The first 26 amino acid residues form two alpha

helices, which are tilted with respect to each other by a kink of 60° at Glu17. The first helix (2-16) inserts into repeat III of the of the core domain while replacing the D-helix. Residues 27-33 were viewed as an unstructured strand.⁶

A single tryptophan residue is located at position 12 in the N-terminus of annexin I. The crystal structure revealed that the nitrogen atom on the indole ring of the residue is hydrogen-bonded to a water molecule, therefore the residue is solvent accessible. Fluorescence spectra of annexin I in solution showed a maximum emission wavelength of 345 nm for Trp12, which is close to the maximum of free tryptophan in solution, again suggesting that the residue is solvent accessible. Annexins with relatively large N-terminal domains (between 30-100 residues) harbor binding sites for S100 proteins and various phosphorylation sites for serine/threonine and tyrosine-specific kinases. Annexin 1 has been shown to be phosphorylated at Tyr21 by the epidermal growth factor receptor (EGF-R) kinase *in vitro* in the presence of calcium, while the N-terminus is exposed. Interestingly, EGF-R is located at multivesicular bodies, and therefore annexin I is associated with those lipid structures. Protein kinase C also phosphorylates annexin 1 within the N-terminus at Thr24, Ser27, Ser28, and Thr41 and in the core domain by PKA at Thr216.⁷

Residues 10-14 of annexin I represent the binding site for a protein ligand of the S100 family, S100A11. As previously stated, in the absence of calcium the N-terminus of annexin I is found to be buried inside repeat III of the core domain, replacing the D helix. The conformational changes the N-terminus undergoes upon calcium binding have important mechanistic effects. Upon calcium-dependent membrane binding, the D helix is speculated to be forced back into position in the core, while the N-terminal helix is

expelled from the core and is free to interact with an S100 ligand. Furthermore, the exposed N-terminal helix, reasoned to be the second membrane binding site on the protein, is free to interact with another membrane surface or to provide a homophilic interaction site for another annexin I molecule. Therefore, calcium appears to have a dual regulatory role in that it triggers the attachment of the protein to a membrane surface *via* the convex face of annexin I as well as ejecting the N-terminus from the core domain so that it is able to interact with cellular protein ligands, and/or a second membrane surface. In its exposed position the N-terminus is also accessible to its respective kinases, thereby regulating the activity of annexin I. Phosphorylations are thought to regulate some annexin I functions, such as liposome aggregation, by increasing calcium requirement.⁸

The structure of residues 1-14 of annexin I in complex with S100A11 ligand has been solved, revealing a 1:1 stoichiometry between the two peptides. The structure is very similar to that of the N-terminus sequence of annexin II in complex with its S100 ligand, S100A10. For both annexin I and II, the first 14 residues of the N-terminus provide the binding sites for two ligands of the S100 family of proteins. In annexin II, the formation of a heterotetrameric complex containing an S100A10 dimer and two annexin II molecules has been shown to significantly alter the biochemical properties of the protein in cells in *in vitro* and *in vivo* studies. The A2-S100A10 heterotetramer is able to aggregate membrane vesicles at micromolar Ca^{2+} concentrations, a property which is unique to this particular annexin complex. From high-resolution images of the junctions formed between adjacent membranes and the annexin A2-S100A10 complex, it appears that the two annexin A2 subunits are bound to the two separate bilayers while the

S100A10 dimer connects the two subunits through binding to the N terminal domains. A similar scenario could hold true for the A1-S100A11 complex.¹

Annexin I has been shown to exhibit two distinct membrane binding sites. One is calcium dependent and is specific for a negatively charged membrane, while the other is calcium independent and has not shown charge specificity. The calcium dependent binding has clearly been shown to take place on the convex face of the annexin monomer. The secondary membrane binding site is thought to be located on the concave face of the annexin monomer, in particular on the exposed N-terminus.^{9,10,11}

According to a study involving immunochemical analysis annexin I was shown to associate with early endosomes and multivesiculated bodies. A truncated version of annexin I lacking the first 26 residues did not bind to early endosomes; rather it was associated with late endosomes and multivesicular bodies. This finding is an example of the importance of the N-terminus in regulating annexin I function and its possible role in endocytosis.¹²

Additionally, a chimera comprising the core domain of annexin V (which does not promote membrane fusion activity and has a shorter N-terminus with only 19 residues) fused with the N-terminus of annexin I was found to cause membrane aggregation by Andree and co-workers. The protein bound to a single phospholipid bilayer composed of 80% phosphatidylcholine (PC)/ 20% phosphatidylserine (PS) at identical Ca^{2+} levels as annexin V, but aggregated and fused vesicles unlike annexin V. The chimera protein was unable to bind to pure PC vesicles. This suggests upon calcium-dependent binding to a negatively charged membrane, the chimera exposes a second

membrane binding site not present in annexin V. This underscores the regulatory role played by the N-terminal domain of annexin I in membrane aggregation.¹³

Studies conducted by Eduard and Cho generated truncated mutants of annexin I lacking various parts of the N-terminus in an attempt to discern the specific residues responsible for the membrane aggregating activity. They first truncated the entire N-terminus (i.e., Δ^{1-41}), which showed no detectable membrane aggregation activity under normal assay conditions, yet at high protein concentrations was shown to cause aggregating activity. Upon addition of residues 29-41, the protein lost aggregating activity of the core domain, whereas further addition of residues 25-29 fully restored the wild-type activity. They speculated that residues 25-29 played an important role in membrane aggregation. This region contains two cationic residues (Lys-26 and Lys-29) and two serines (Ser-27 and Ser-28), whose *in vitro* phosphorylation by protein kinase C has been shown to decrease the vesicle aggregating property of annexin I. They speculated the two lysines form the interaction site for aggregation. Phosphorylation of the two serine residues would add a negative charge on those residues and thereby electrostatically neutralize the positive charge on the lysine side-chains, thus interfering with the proposed interaction site of aggregation.¹⁴

The previous discussion emphasized and strengthened the fact that the N-terminus plays a vital part as the plausible second membrane binding site on annexin I and must be further investigated to determine its precise role in the mechanism of annexin induced membrane aggregation.

1.9: Core Domain

The core domains of annexins exhibit a sequence homology of up to 80% among different members of the family, and are composed of four repeats (eight in annexin VI) made up of five alpha-helices (A-E) each. Four of the helices (A, B, D and E) form a coiled coil structure, with helix C capping the structure. The core domain of porcine annexin I comprises 305 residues. The core takes the shape of a slightly curved disc with loops connecting helices A and B and helices D and E on the convex face of the disc. The calcium binding sites are located within these loops. The bound calcium ions serve as a hypothetical 'bridge' between protein and membrane by simultaneously coordinating ligands from acidic side-chains of the protein and from phosphoryl moieties of the lipids. The convex face of the core domain is thus able to peripherally dock in a reversible manner onto the surface of negatively charged membranes. The concave side of the protein faces the cytoplasm and is available for further interactions.¹⁵

To date, many of the cores have been crystallized, revealing calcium type II or type III binding sites on the convex side, and a central hydrophilic pore proposed to function as a Ca^{2+} ion channel, as depicted in Figure 1.4. In the presence of phospholipids, the Ca^{2+} affinity of these sites is in the low micromolar range, although exact affinities vary between different annexins.

In annexins, the calcium ion coordination complex adopts a pentagonal bipyramidal molecular geometry. Calcium ions in crystal structures generally have coordination numbers from 6 to 8, and rarely adopt ligands other than oxygen atoms. More specifically, the calcium ion binding sites of annexins are of three particular geometries, listed below with their oxygen ligands.¹⁶

Table 1.2: Binding Site Geometries of Annexins.

Binding Site	Location	Ligands
Type II	AB loop	-3 carbonyl oxygen atoms (from protein backbone) -1 bidentate acidic side-chain (40 residues downstream of AB loop) -2 H ₂ O
Type III	DE loop	-2 carbonyl oxygen atoms -1 acidic side-chain from the E helix -3 H ₂ O
AB'	AB loop	-1 carbonyl oxygen atom -1 acidic side-chain -5 H ₂ O

In a manner of housekeeping, Type I binding sites are the canonical EF-hand motif (helix-loop-helix) found in many proteins. Type I sites have more carboxylate ligands, fewer water ligands, and a longer calcium binding loop. Table 1.3 lists the specific ligands with which the calcium ions of annexin I are coordinated with. It should be noted that in annexin I calcium ions 1 and 4 coordinate with eight oxygen atoms, whereas the remainder of the ions coordinate with seven oxygen atoms.

Table 1.3: Calcium ions in annexin I and their associated ligands.¹⁷

Calcium ion	Binding site	Coordination number	Ca – O distance (Å)	
1	Type II, repeat I	8	O(59) = 2.27	H ₂ O1 = 2.59
			O(60) = 2.34	H ₂ O2 = 2.33
			OE1(62) = 2.42	H ₂ O3 = 2.60
			OE2(62) = 2.46	H ₂ O4 = 2.36
2	Type III, repeat I	7	O(97) = 2.29	H ₂ O1 = 2.49
			O(100) = 2.53	H ₂ O2 = 2.42
			OE2(105) = 2.44	H ₂ O3 = 2.37
			OD2(196) = 2.30	
3	Type II, repeat II	7	O(127) = 2.26	OD2(171) = 2.59
			O(129) = 2.32	H ₂ O1 = 2.34
			O(131) = 2.35	H ₂ O2 = 2.42
			OD1(171) = 2.42	
4	AB', repeat II	8	O(132) = 2.36	H ₂ O2 = 2.57
			OE1(134) = 2.63	H ₂ O3 = 2.54
			OE2(134) = 2.53	H ₂ O4 = 2.45
			H ₂ O1 = 2.29	H ₂ O5 = 2.53
5	Type II, repeat III	7	O(210) = 2.44	OE2 (255) = 2.70
			O(213) = 2.38	H ₂ O1 = 2.36
			O(215) = 2.38	H ₂ O2 = 2.41
			OE1(255) = 2.61	
6	Type II, repeat IV	7	O(286) = 2.37	OE2(330) = 2.49
			O(288) = 2.31	H ₂ O1 = 2.46
			O(290) = 2.46	H ₂ O2 = 2.49
			OE1(330) = 2.55	
7	Type III, repeat IV	7	O(328) = 2.27	H ₂ O2 = 2.28
			O(331) = 2.26	H ₂ O3 = 2.29
			O(336) = 2.24	H ₂ O4 = 2.41
			H ₂ O1 = 2.87	
8	Type III, repeat III	7	O(253) = 2.45	H ₂ O1 = 2.55
			O(256) = 2.60	H ₂ O2 = 2.90
			OE2(261) = 3.03	H ₂ O3 = 2.80
			OD1(253) = 3.03	

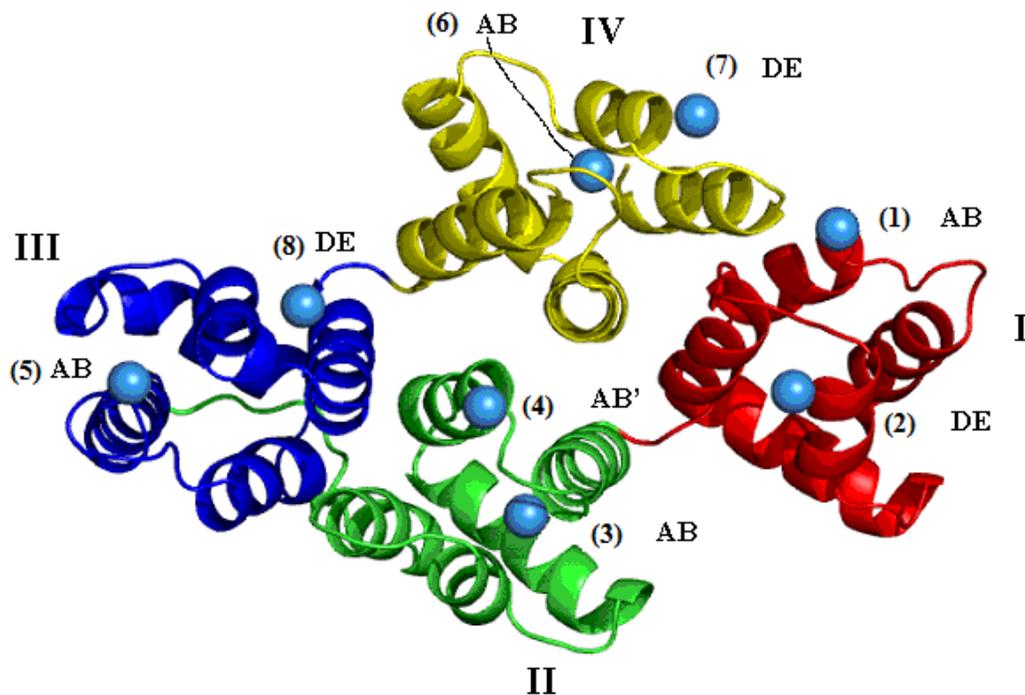


Fig 1.4: Axial view of annexin A1 taken from 1MCX.pdb. Repeats are color coded: Repeat I (red), Repeat II (green), Repeat III (blue), Repeat IV (yellow). Calcium ions are shown as blue spheres, labeled (1-8) with their binding site geometries. It should be noted that calciums 1 and 4 are octa-coordinated, while the remaining calciums are hepta-coordinated.

CHAPTER 2 - RESEARCH PLAN

2.1: Approach

To solve the discrepancy of the model of membrane aggregation between cryo EM and X-ray studies, more structural studies are needed. However, crystallization of full-length calcium-bound annexin I in the presence of phospholipids has not been successful so far. Molecular dynamics (MD) simulations appear to be the best approach to studying the complex environment of the lipid protein interface at the molecular level and to elucidate the mechanism of annexin-induced membrane aggregation.

The molecular dynamics simulation can be performed on a system including the annexin I monomer positioned in between two phospholipids bilayers composed of dioleoylphosphatidylcholine (DOPC) and dioleoylphosphatidylglycerol (DOPG) in a 4:1 molar ratio, comparable to the liposome composition used in the cryo EM studies. The trajectory of the simulation will be analyzed, and after the expected interactions between the protein and phospholipids, the averaged distance between the bilayers will be compiled and compared with experimental results. Cryo-EM experiments were also performed to study the annexin I dimer induced membrane fusion mechanism, and simulations can be performed on this system as well. However, the dimer simulation is considerably more complex and requires longer computational time and resources in order to properly study. This study will therefore focus on simulations comprised of the annexin monomer positioned between two bilayers. Specifically, the annexin monomer has calcium ions bound in all eight binding sites on its convex face, as well as its N-terminal helix positioned outside the core domain where it is solvent accessible. We are therefore able to analyze the initial membrane binding site on the convex face of the protein and the importance of the speculated secondary membrane binding site located at the N-terminal head. The discussion in the coming chapters provides the fundamental approaches of MD simulations and a summary of the simulations performed on annexin I in close proximity to phospholipids bilayers.

2.2: Theory of Molecular Dynamics Simulation

Our research involved the use of molecular dynamics simulations. Molecular dynamics (MD) is a computational technique which allows atoms and molecules to interact in a temporal evolution under the known laws of physics. MD was originally

devised within theoretical physics in the 1950's; however, it wasn't until the 1970's that people began using MD simulations routinely due to the increased availability of computers. The first biomolecule simulated using MD was bovine pancreatic trypsin inhibitor (BPTI), a small globular protein.^{18,19,20}

With the advent of modern supercomputers, it has become possible to run a MD simulation on the nanosecond timescale. A 1 microsecond all-atom simulation was conducted on a villin headpiece, a 36-residue peptide, in 1998 by Duan and Kollman.²¹ Long MD simulations are mathematically ill-conditioned however, and result in a culmination of numerical integration errors.

Classical MD bases the temporal evolution on Newton's second law, $F = ma$, where the forces are obtained as gradients of the potential energy. In classical MD, the energy of the system is a function of the nuclear positions only (Born-Oppenheimer approximation).²²

The force fields used in classical MD tend to be relatively simple, such as treating covalent bonds as springs using Hooke's law. There are a variety of force fields in use today. Force fields are a collection of atomic interactions, and their parameters define the potential energy of the system. A simple form of such a force field is represented in the following function:

$$V(r) = V_l + V_\theta + V_{\text{tors}} + V_{\text{vdw}} + V_{\text{elect}}$$

In this equation, the potential energy, $V(r)$, is a function of the positions r of N particles, V_l is the bond deformation energy, V_θ is the deformation energy of the bond angles, V_{tors} is the energy associated with the dihedral angles, V_{vdw} is the van der Waals non-bonded energy and V_{elect} is the non-bonded electrostatic energy.²³

Where the bond length fluctuation is close to the natural bond lengths, a simple harmonic approximation is sufficient. Similarly, a deviation of bond angles from their reference value, the angle potential, can be represented in a formula similar to Hooke's law.

$$V_l = 0.5 \sum k_l (l - l_0)^2$$

$$V_\theta = 0.5 \sum k_\theta (\theta - \theta_0)^2$$

Proper torsions, which describe the rotation around the central chemical bond in a series of three bond vectors, are usually expressed in the form of a cosine expression. The most commonly used function for proper torsion potential is the Pitzer potential:

$$V_{\text{tors}} = \sum k_\phi [1 + \cos(n\phi - \delta)]$$

In this equation, k_ϕ is the force constant, δ is the reference angle at which the potential energy is at a maximum, and n is the multiplicity which is the number of potential minima in one full rotation. Improper torsions describe the out-of-plane bending of an atom found in the center of three surrounding atoms bonded to it.²³

The electrostatic non-bonded potentials can be represented a number of different ways. If fractional point charges are assigned to the nuclear centers of atoms in a molecule, they are referred to as partial or net atomic charges. Using Coulomb's law, the interactions are calculated as a sum of interactions between pairs of point charges:

$$V_{\text{elect}} = \sum_i \sum_{j \neq i} \frac{q_i q_j}{4\pi\epsilon_0\epsilon_r r_{ij}}$$

where q_i and q_j are the pair of point charges i and j , r_{ij} is the distance between those two charges, ϵ_0 is the permittivity of free space and ϵ_r is the relative dielectric constants which may be function of r_{ij} .

Non-bonded interactions other than those caused by electrostatic forces were described by Dutch scientist Johannes Diderick van der Waals while studying deviations from ideal gas behavior, including terms representing dispersion V_{disp} and repulsion V_{rep} . Dispersion forces (sometimes referred to as London forces) are weak intermolecular forces that arise from instantaneous dipoles caused by fluctuations in electron clouds. A dipole can provoke other dipoles in neighboring atoms, giving rise to an inductive effect. The attractive force potential is expressed in the form of a power series:

$$V_{\text{disp}} = \frac{C_{ij}}{r_{ij}^6} + \frac{C_{ij}}{r_{ij}^8} + \frac{C_{ij}}{r_{ij}^{10}} + \dots$$

where r_{ij} is the distance between two atoms i and j and C_{ij} is the coefficient which depends on atomic polarizabilities (and is always negative, implying an attraction). Repulsive forces are generated when atomic wavefunctions overlap in conjunction with the Pauli exclusion principle, in that it is impossible for two electrons with the same spin to occupy the same space. The potential of this force is represented in the following equation:

$$V_{\text{rep}} = Ae^{-br_{ij}}$$

where A and b are constants and r_{ij} is the distance between the two atoms. This term is often replaced and is commonly modeled with the Lennard-Jones function V_{LJ} which is represented in the following equation:

$$V_{\text{LJ}} = 4\varepsilon_{ij} \left[\left(\frac{\sigma}{r_{ij}} \right)^{12} - \left(\frac{\sigma}{r_{ij}} \right)^6 \right]$$

where σ is the collision diameter (the separation for which the energy is zero) and ε_{ij} is the depth of the energy well.²³

All the aforementioned interactions are calculated between pairs of atoms. A significant proportion of interactions involving three or more atoms can be incorporated into the pairwise model with the correct parameterization.

Furthermore, the simulation of the protein-membrane interaction will occur in its natural, physiological environment, an aqueous medium. The most commonly used models of water are the TIP3P and SPC.^{24,25} The models employ a rigid geometry and use Coulombic and Lennard-Jones expressions to define atomic pair-wise interactions.

Once a suitable force field is established, an optimization algorithm may be implemented to generate minimum energy conformations. A variety of such algorithms exist, many of which require the calculation of the first derivative. The first derivative is useful because it can convey information regarding the shape of the energy surface, which aids significantly in locating a minimum. Iterative minimization algorithms are used to relieve local stresses of a structure due to overlap of non-bonded atoms, bond length distortions, etc. The steepest descent method (often called the saddle-point approximation) is commonly used to calculate integrals in the form:

$$\int_a^b e^{-Mf(x)} dx$$

where $f(x)$ is some twice differentiable function, M is a large number, and the limits of integration a and b could possibly be infinite. This method moves in a direction parallel to the net force. For $3N$ Cartesian coordinates, this direction is represented by a $3N$ -dimensional unit vector, s_k .²³

$$s_k = \frac{\mathbf{g}_k}{|\mathbf{g}_k|}$$

where g_k is the gradient for each iteration k . New coordinates of the system are obtained by taking a step of arbitrary length, x_k , along the unit vector s_k . The new set of coordinates after step k is given by the equation:

$$x_{k+1} = x_k + \lambda_k s_k$$

where λ_k is the step size.

Molecular dynamics includes temperature (kinetic energy) through atomic velocities. The temperature T of the system is measured by the mean kinetic energy, as described in the following equation:

$$\frac{1}{2} \sum_{i=1}^N m_i (v_i)^2 = \frac{3}{2} N k_B T$$

where N is the number of atoms in the system, v_i is the average velocity and k_B is Boltzmann's constant.²³

Atoms are assigned velocities by finding the acceleration a_i of atom i from Newton's law $F_i = \mathbf{m}_i a_i$, where the force on the atom is computed from the derivative of the potential energy function with respect to atomic position, and \mathbf{m} is the mass of the atom. The total force on each particle is calculated as the vector sum of its interactions with other particles. Because MD incorporates a vast number of particles and continuous potentials, it is impossible to solve the properties of the system under study analytically; instead, MD uses numerical methods, implemented within computer algorithms, to circumvent this problem. The method is deterministic; once the positions and velocities of each atom are known, the state of the system can be predicted at any time in the future. The Verlet algorithm is one of the most popularly used methods. The atom's

acceleration at time t and at position r_i are then used to calculate its new position at $t + \delta t$ using the following equation.²⁶

$$r_i(t + \delta t) = 2r_i(t) - r_i(t - \delta t) + \delta t^2 a(t)$$

The leap-frog algorithm, a modification of the Verlet algorithm, is used in a number of simulation packages for the integration of the equations of motion.

Periodic boundary conditions are used, in which a box containing the system under study is repeated infinitely in every direction to give a periodic array. As particles leave the box they are replaced by identical imaged particles, thus keeping the number of particles in the box constant. The Particle Mesh Ewald method is efficient in handling the long-range forces of the system.^{27,28} The SHAKE algorithm is commonly used to constrain the degrees of freedom of covalent bonds involving hydrogen atoms, thus reducing the amount of computer time needed for the simulation.

2.3: Molecular Dynamics Simulations on Annexins

Molecular dynamics simulations have been reported on full-length annexins as well as on individual annexin repeats in attempt to elucidate structural and conformational changes occurring between the calcium-free and calcium-bound forms of the protein. The following studies are described in chronological order and comprise the background for my research.

In 1997, G.V. Musat and co-workers published results in which they applied homology modeling to predict the structure of domain II of annexin I from the known crystallized structure of domain II of annexin V. Homology modeling refers to the construction of a target protein from an evolutionary-related template protein, using the primary sequence of the target protein taken from genome research projects. The

availability of the crystal structure of annexin I allowed for an assessment of the accuracy of their model using molecular dynamics simulation in explicit solvent.

The domain II which they created using homology modeling was placed in a cubic water box with a volume of 125000 \AA^3 , and periodic boundary conditions applied using the CHARMM software, version 23. The experimenters performed two simulations in the microcanonical ensemble, namely a ‘constrained’ and an ‘unconstrained’ simulation. The primary difference between the two is that harmonic constraints were only applied to the ‘constrained’ simulation during the equilibration phase. Significant deviations of the backbone structure from the initial structure were observed in both simulations. The RMS deviations of the backbone heavy atoms were 3.5 \AA in the unconstrained simulation and 3.0 \AA in the constrained simulation after production.²⁹

The study by G.V. Musat employed homology modeling only on core residues and excluded other annexin domains as well as the solvent environment. A sequence identity of 41% was reported in the core region between the two crystal structures, and the RMS deviation of corresponding α -carbons was 0.96 \AA .

D. Cregut *et al.* (1998) used molecular dynamics and essential dynamics (ED) to study the hinge-bending motions in apo-annexin V and calcium-bound annexin V and annexin I. In this study, three simulations were performed to investigate the influence of calcium binding by monitoring the changes in conformation and dynamics in the two modules of annexins. An approximate twofold symmetry exists in annexins, and the protein can be divided into two modules (repeats I/IV and II/III). The following two types of hinge-bending motions were postulated for annexins: the opening of the α -angle

between the modules might be related to calcium binding; the rotation of the *tor* dihedral angle between the modules is constrained upon membrane binding on the convex face of annexin V.

This study obtained the initial starting coordinates of apo-annexin V and calcium-bound annexin V from their X-ray structures in the protein data bank. The calcium-bound annexin V crystal structure included 316 residues, five Ca^{2+} ions and 202 water molecules. Calcium-bound annexin I contained 314 residues and included six Ca^{2+} ions and 377 water molecules. It should be noted that the first 32 residues of the N-terminus were unable to be crystallized in the annexin I X-ray structure (pdb code 1ain). X-ray crystallographic waters as well as a 9 Å shell of TIP3P waters were used in the simulations. MD simulations were carried out using four processors on an IBM SP2. The Assisted Model Building with Energy Refinement (AMBER), version 4.1, all atoms force field was used to conduct the calculations for the MD simulations. The water shell was first minimized, and then the system was subjected to a 10 ps trajectory at 300 K with the protein and ion positions fixed. After equilibration, several minimization steps were performed, the first of which had force constraints applied to backbone atoms and side-chain atoms as well as the Ca^{2+} ions. During the next step, side-chain constraints were removed, and finally, the last minimization step was performed without any constraints. For each system, a 1050 ps trajectory was calculated at a constant 300 K.³⁰

The results of the study by D. Cregut study showed that, on average, the α -angle between modules was larger, by approximately 5°, in the calcium-bound form of annexin V than in the calcium-free form. This is in good agreement with experimental results. The average *tor* dihedral angle did not differ by much between the two annexin V

simulations (a difference of 1.8° was reported). However, the simulations were performed without a membrane present, so it is difficult to analyze the importance of the hinge-bending motion on membrane binding. The average α angle and *tor* dihedral angles for the annexin I simulation were significantly different from the annexin V simulations, and were attributed to the different biological functions of the proteins. Furthermore, this study found that repeat I was the most mobile repeat on annexins.

Tru Huynh *et al.* (1999) conducted MD simulations on the isolated domain I of annexin I. Nuclear magnetic resonance (NMR) studies on isolated annexin domains showed that domain I retains its tertiary structure whereas domain II unfolds. This behavior is interesting since the four domains in annexins have a virtually identical structure, even though there is only approximately 30% sequence homology between the domains.³¹

Two simulation protocols were used in the study conducted by Tru Huynh. The initial coordinates of both were taken from the X-ray structure, with domain I consisting of residues Thr 41- Lys 113. The first protocol embedded domain I in a 46 \AA cubic TIP3P water box and was run for 1100 ps using the CHARMM version 23 software. Bonds containing hydrogen were constrained using the SHAKE algorithm. Domain I was found to unfold within 400 ps at a temperature of 300 K. The second protocol embedded domain I in a 62 \AA cubic water box using the CHARMM program version 26b1 producing a 1600 ps simulation. The second protocol led to the native-like conformation of domain I. This work demonstrated the sensitivity of the conformation of annexin I upon varying MD conditions.

Jana Sopkova-De Oliveira Santos et al. (2000) studied the conformational changes upon calcium binding to domain III of annexin V, particularly the surface exposure of Trp187. A complex, low energy pathway for the conformational change was created using the conjugate peak refinement method. The pathway is presented as a sequence of molecular events starting from Trp187 exposed (reactant) to Trp187 buried (product). It was found that the burial of Trp187 caused an increase in conformational strain, which was compensated by improved protein-protein interaction energies. All calculations were performed using the CHARMM version 23 software. Five acidic residues were found to have a crucial impact on the conformational changes of domain III via hydrogen bonds with the indole ring of Trp187. Thus they concluded that protonation and deprotonation of the acidic residues, dependent on the pH, had a large-scale conformational influence on the protein.³²

Jana Sopkova-De Oliveira Santos et al. (2001) conducted MD simulations to further investigate a molecular switch for the large-scale change in conformational dynamics of domain III of annexin V. Asp226, one of the key acidic residues which forms a hydrogen bridge with the hydroxyl group of Thr226 and thus is speculated to stabilize the IIID-IIIE loop, was mutated to a lysine. They showed that the single point mutation D226K was sufficient to invoke a large scale conformational change in domain III, and verified the mutation results experimentally by analyzing X-ray crystallography and fluorescence spectroscopy results.³³

Tru Huynh *et al.* (2002) performed multiple MD simulations on the unfolding transitions of domain II of annexin I and compared the results with NMR data. The simulations were performed with CHARMM version c26b2 using the PARM22 all atom

parameter set. Two pre-equilibrated cubic TIP3P water boxes were used to embed domain II in; one measuring 62.04 Å side, and the second for the unfolded state measuring 80 Å side.³⁴

After minimization and equilibration steps, the production phase was performed in the NVT ensemble. Six simulations were performed; one each at 300 K, 350 K and 400 K and three at 450 K (labeled A, B, U). The 450 KA simulation started from the crystal structure of domain II and ran for ~10ns, and no unfolding was observed, contrary to NMR data. The 450 KB simulation was started at 4.2ns of the initial simulation, thus bifurcating the 450 KA trajectory. The 450 KB simulation had unfolded more than 450 KA with an increased radius of gyration, but still did not reach the equilibrium unfolded state found from NMR data. The 450 KU simulation placed domain II in a larger box and applied constraints so that no significant contact occurred between helices A, B, C, and E for 145 ps. At this point the constraints were removed and the domain was left to relax for ~3ns. Thus their goal was achieved in that the simulated domain of the 450 KU reproduced in detail the native unfolded state observed in NMR experiments.

Pierre Montaville *et al.* (2002) conducted MD simulations and mutagenesis studies to identify a novel phosphatidylserine (PS) binding site in domain I of annexin V. Radiocrystallography as well as NMR data hardly provide information concerning annexin binding to phospholipids at atomic detail, so the experimenters turned to modeling and simulation methods. The PS molecule was first drawn and then placed in close proximity to the supposed binding site, which is between helices A and D of domain I. The lipid and the six-residue side-chains located between helices A and D were allowed to move, and the rest of the protein was held fixed. The final structure was

obtained by a short dynamic run at 100 K, followed by minimization and simulated annealing utilizing the Sybyl program. The docking of a PS molecule on annexin V domain I revealed a conserved PS binding sequence. The affinity of the protein for a calcium ion at the AB site was shown to depend on the presence and number of lipid phosphate groups in the calcium coordination sphere.³⁵

Franci Merzel *et al.* (2005) developed force field parameters describing the interaction of a calcium ion in an annexin with its environment. They used quantum mechanical calculations to describe the potential energy surface of the calcium ion within the three different binding sites of domain I of annexin V. They were able to quantify the partial charges experienced by the atoms in the binding sites as well as describing the geometry and harmonic restraints between the calcium ion and its oxygen atom ligands. Finally, they applied their force field to an MD simulation using CHARMM and compared the results to a simulation using a standard force field. They conducted the following two ~3ns simulations on domain I of annexin V: Annexin V domain I in complex with three PS molecules in water solution, and annexin V domain I without PS molecules in solution. The refined force field led to a better overall structural stability of the domain during dynamics. Moreover, the analyzed RMS deviations of both simulations agree well with electron paramagnetic resonance (EPR) results reported for calcium dependent membrane binding of annexin B12.³⁶

Rohini Sesham *et al.* (2008) performed MD simulations on three protein systems to elucidate conformational changes during annexin I induced membrane aggregation. These systems were: 1) Ca^{2+} free annexin I with the N-terminus buried inside the core, 2) Ca^{2+} bound annexin I lacking the N-terminus, and 3) Ca^{2+} bound annexin I with the

N-terminus in an exposed position outside the core comprised the three systems under study. All computations were performed on a 32-processor SGI origin 350. All calculations were performed using AMBER's all atom force field as implemented in AMBER 8 program. The results indicated that the calcium coordinating residues on the convex face of the protein showed relatively higher fluctuation values than non-coordinating residues. RMSD (Root Mean Square Deviation) analysis indicated that the N-terminus is the most flexible region of the protein when in an exposed position, implicating its possible role as a second membrane binding site. The exposed N-terminus lost its secondary structure over the course of the simulation, leading the experimenters to postulate that it adopts a favorable conformation to bind to a second membrane. The measured dimensions of the annexin system with the exposed N-terminus supported the mechanism of membrane aggregation proposed by X-ray studies.³⁷

It should be noted that a majority of the previously described MD studies on annexins were conducted on a single domain of an annexin. The goal of the present study is to conduct an MD simulation on full-length annexin I in the presence of a phospholipids bilayer and to establish a mechanism for the observed aggregation property of this protein. The aforementioned studies will provide a strong background to help achieve this goal.

CHAPTER 3 – METHODOLOGY

3.1: Computational Aspects of MD Simulations

Molecular dynamics simulations were implemented for the purpose of investigating the mechanism of annexin I induced membrane aggregation at atomic level detail. MD is a computational technique which allows atoms and molecules to interact in

a temporal evolution under the known laws of physics. Classical MD bases the evolution on Newton's second law, $F=ma$, where the forces are obtained as gradients of the potential energy.

$$U = \sum_{bonds} K_b(b-b_0)^2 + \sum_{angles} K_\theta(\theta-\theta_0)^2 + \sum_{dihedrals} K_\chi[1+\cos(n\chi-\sigma)] \\ + \sum_{electrostatic} \frac{q_i q_j}{\epsilon_d r_{ij}} + \sum_{vdw} \left(\epsilon_{ij} \left[\left(\frac{R_{min,ij}}{r_{ij}} \right)^{12} - \left(\frac{R_{min,ij}}{r_{ij}} \right)^6 \right] \right)$$

The first and second terms use a formula similar to Hooke's law. In the case of bonds, K_b is the stretching constant obtained from spectroscopy and b_0 is the reference bond length. Similarly, K_θ is the spring constant and θ_0 is the reference bond angle. Dihedral rotations take the form of a cosine series expansion. Partial charges are assigned to nuclear centers, and electrostatic interactions are calculated using Coulomb's law. The partial charges are generated using *ab initio* quantum mechanical calculations. The last term takes into account the remaining non-bonded interactions, modeled with the Lennard-Jones function, where r_{ij} is the distance between two atoms, $R_{min, ij}$ is the distance at which the energy is zero, and ϵ_{ij} is the depth of the energy well.

3.2: Force Field

The force fields used in molecular dynamics simulations are a collection of atomic interactions, or parameter sets for each atom type, which describe the potential energy of a system. These parameter sets are derived from experimental work as well as high-level quantum mechanical calculations. The functional form of a force field includes both bonded and non-bonded terms related to atoms.

To begin a molecular dynamics simulation, an initial set of atomic coordinates is required. With the set of coordinates, a calculation is performed to equilibrate the system. This relieves local stresses due to non-bonded orbital overlap, as well as bond length and angle distortions. Next, velocities are assigned to atoms, starting from a low temperature, and successively increased to our desired temperature (300K). A potential energy function is calculated as the minimization and warming steps occur, thus a system trajectory is obtained. During MD simulations, Newton's equations of motions are continuously integrated to generate atomic coordinates and velocities as a function of time. The velocities and coordinates are recorded in an output file after each time step. The time step should be very small so that the potential energy does not change significantly during each step. Typical time steps are on the order of one femtosecond (10^{-15} s). The SHAKE algorithm can be used to extend the value of the time step by constraining the bond geometry of hydrogen atoms.

The most popular classical force fields used are AMBER, CHARMM, GROMACS, GROMOS, and OPLS. The AMBER ff03 force field was used in this study since it is widely used and accredited for MD simulations of biological macromolecules. The AMBER software package also contains an extensive suite of programs that allow for thorough post simulation analysis.

3.3: AMBER

The present study utilized AMBER (Assisted Model Building with Energy Refinement) to perform MD simulations. AMBER refers to the suite of programs used to carry out MD simulations and perform common calculations, and is also used in reference to the empirical force fields that are implemented in MD simulations. AMBER

9 was the version of the software used in the present research. It is designed to be run on a Unix based platform.

The flow of information in the software package is illustrated in the following Figure 3.1.

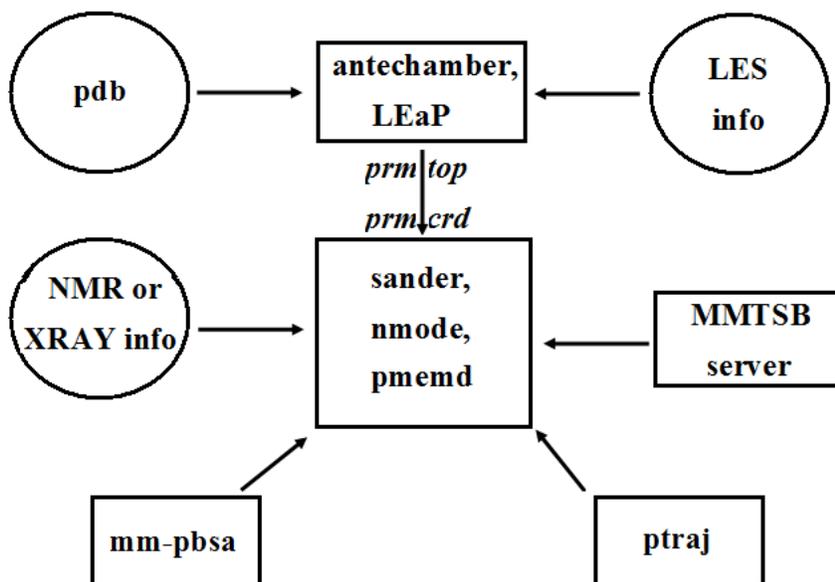


Fig 3.1: Basic information flow in AMBER.³⁸

The three main steps in the MD simulations over the course of the present research include system preparation, simulation, and trajectory analysis.

(I) Preparatory Programs

LEaP is the primary program to create a new system in Amber. The name LEaP is an acronym for prep, link, edit, and parm. It is a graphical builder of input files used in other modules of AMBER. LEaP generates two very important input files: a coordinate

file (.crd) containing the Cartesian coordinates of the system (obtained from X-ray or NMR data), and a topology file (.top) which define the atom connectivity within a molecule through chemical bonds. In the present study, the LEaP module was used to determine the overall charge of the system, to create a disulphide bond within the protein, to add sodium and chloride ions in order to provide system neutralization and physiological ion concentrations, and then explicitly solvate the system.

(II) Simulation Programs

Sander (Simulated Annealing with NMR-Derived Energy Restraints) is the basic energy minimization and molecular dynamics program. Sander provides direct support for force fields for proteins, nucleic acids, water molecules and organic solvents. The MD portion of Sander generates configurations of the system under study by integrating Newtonian equations of motion. It uses MPI (message passing interface) programming for communication among multiple processors. Each processor ‘owns’ certain atoms, yet all processors know the coordinates of every atom. Each processor computes its portion of the potential energy and the corresponding gradients, after which the force vector is summed and reported to each processor.

The particle-mesh Ewald (PME) procedure was used in Sander to handle long-range electrostatic interactions. Long-range van der Waals interactions are estimated by a continuum model.

(III) Analysis Programs

The trajectories generated during the MD simulations were analyzed using the ptraj program and the MM-PBSA (Molecular Mechanics Poisson-Boltzmann Surface Area) script. Ptraj is a general purpose utility for analyzing and processing trajectory or

coordinate files to extract information including bond, angle, dihedral values, atomic positional fluctuations, and other information. For each coordinate set read in, a sequence of actions can be performed on each of the configurations read in, and after processing all the configurations, a trajectory file is generated as output. To use the program, the following steps are performed:

1. Read in a parameter/topology file.
2. Set up a list of input coordinate files.
3. Specify an output file.
4. Specify a series of actions to be performed on each coordinate set read in.

The series of actions used in the present research included calculations of RMSD, RMSF, B-factors, snapshots of the system in pdb format and inter-atomic distance values.

Another analysis script used over the course of the research was MM-PBSA. This method combines molecular mechanics and continuum solvent approaches to estimate binding energies. The electrostatic contribution to the solvation free energy is solved using the Poisson-Boltzmann equation. The initial sets of structures will be collected from MD simulation and then used in the AMBER program to calculate the total binding energy of the system as shown in the equation:

$$\Delta G_{\text{bind}} = \hat{G}_{\text{complex}} - [\hat{G}_{\text{protein}} + \hat{G}_{\text{ligand}}]$$

The specific interactions we analyzed using MM-PBSA include:

- Amino terminal residues – bottom layer phospholipid interactions
- Core domain residues – amino terminal residue interactions.

3.4: Systems Used in Molecular Dynamics Simulation

In order to investigate the mechanism by which annexin I induces membrane aggregation and fusion, MD simulations were performed on two setups, summarized in the following Table 3.1. Both setups contain the calcium-bound annexin I monomer with an exposed N-terminus positioned between two phospholipid monolayers made up of 180 DOPC molecules and 45 DOPG molecules. The difference between the two lies in the initial distance between the two phospholipid layers prior to running any calculations.

Table 3.1: Dimensions of each system set-up, including total number of atoms used per simulation and timescale of each setup.

	Setup 1	Setup 2
Box Dimensions (Å)	122.75 x 122.00 x 112.94	122.75 x 118.67 x 112.94
Total Number of atoms	142,364	127,344
Simulation length (ps)	26702.2	26344

(I) Annexin Construction

There are no crystal structures available for full-length, calcium-bound annexinA1 with its N-terminus in an exposed position outside of the core domain. Therefore, our first task was to construct this protein under study. The calcium bound core domain starting coordinates were obtained from its X-ray structure resolved at 1.80 Å (1MCX.pdb).⁸ The 41 residue N-terminus coordinates were taken from the X-ray structure of annexin A1 in the absence of calcium (1HM6.pdb), and fused to the core domain in an exposed position using Insight II software. This annexin construction was part of the work of previous molecular dynamics experiments, and therefore was readily available at the start of the present research.³⁷ The constructed protein has a total of 351 residues, eight of which are calcium ions, as shown below.

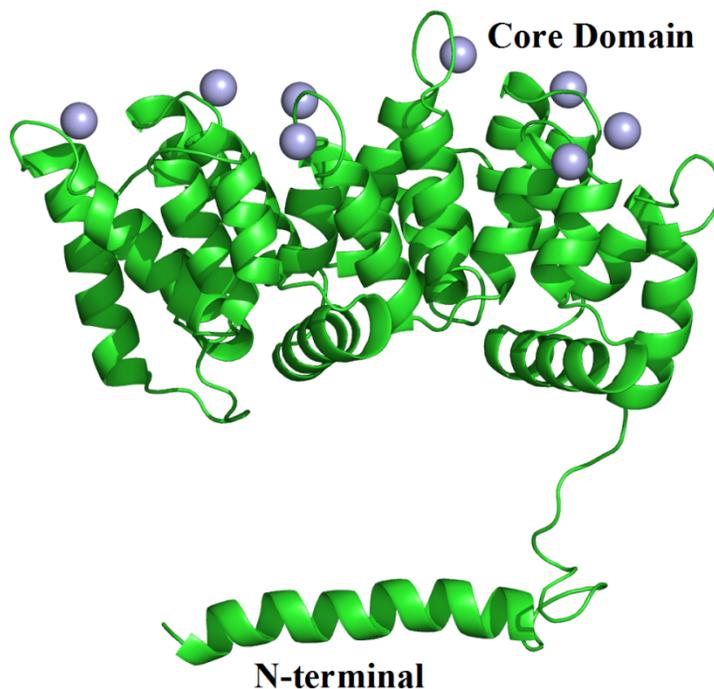


Fig 3.2: Calcium-bound annexinA1 with exposed N-terminus. Calcium ions are shown as light-blue spheres.

(II) Phospholipid Construction

The basic structural element of a membrane is a phospholipid, which contains long hydrocarbon chains and a variety of polar groups. DOPG (dioleoylphosphatidylglycerol), an anionic phospholipid, and DOPC (dioleoylphosphatidylcholine), a zwitterionic phospholipid, were also constructed using the Insight II software, and optimized with DMol3 using a 6-31G basis set. AMBER parameters for the phospholipids were created using the Antechamber program. A total of 225 phospholipids were used per monolayer (180 DOPC: 45 DOPG), creating a 15 x 15 array of phospholipids. This phospholipid layer is better thought of as a viscous fluid rather than a permanent structure, with a viscosity about 100 times that of water. Biological membranes exist as liquid crystals at physiological temperatures. The ratio of

4:1 DOPC:DOPG was chosen to model after the phospholipid composition used in the cryo-EM experiment.⁷ Three separate pdb files were made; the top layer of phospholipids, the annexin protein, and the bottom layer of phospholipids. All three were loaded into Insight II at the origin (0, 0, 0) in coordinate space. Then, each pdb file (top, protein, or bottom) was translated along the y-axis until the annexin protein was positioned between the two constructed monolayers at the desired distance. The system could then be exported as a single pdb file.

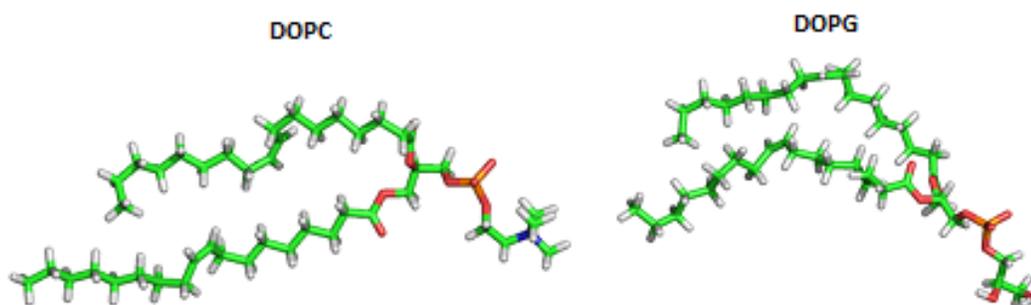


Fig 3.3: Optimized structures of DOPC (left) and DOPG (right) used in the simulation.

3.5: Simulation Conditions

The following steps were performed in LEaP to prepare the coordinate and topology files used in the MD simulations.

1. The AMBER ff03 force field was loaded.
2. The constructed pdb file, stripped of hydrogen atoms due to naming incompatibility between the protein data bank and Amber, was loaded.
3. Hydrogen atoms were added where needed.
4. The disulphide bond was created between cysteine 324 and cysteine 343.

5. The systems were neutralized with counter-ions, and additional counter-ions were added to achieve the experimental concentrations.
6. The systems were solvated in the space between the phospholipid layers using the TIP3P water model.
7. The topology and coordinate files were saved for later use in the simulations.

All computations were performed on a 128-processor SGI Altix. Molecular displays were rendered on Silicon Graphics, Incorporated octane workstation using Pymol and InsightII software. All calculations were carried out using AMBER's force field (ff03) in the AMBER 9 software package.

(I) Step 1: Energy Minimization

The initial structures were energy minimized twice before the MD simulation. The energy minimizations utilized information from gradient codes. A restrained minimization was first performed on the solvent and the counter ions while keeping the protein and phospholipids fixed. The atoms specified in this group were restrained using a harmonic potential of $20 \text{ kcal/mol/\AA}^2$. This minimization was performed for 1000 steps, where every 25th step was printed in the form of output. For 500 cycles, the steepest descent method was used, after which the conjugate gradient algorithm was switched on. The dielectric constant was set at the default value of 1.0. The non-bonded cutoff was set to 12 \AA to truncate those interactions. Next, the entire system was minimized for a total of 2500 steps. Again, the steepest descent method was used for the first 500 cycles. The final coordinates were written to a restart file.

(II) Step 2: MD - Warming

The minimized system was then warmed for 20 ps to 300 K from an initial temperature of 10 K. A 0.02 ps time step was used, and the SHAKE algorithm was used to constrain all bonds involving hydrogen atoms with a geometric tolerance of 0.00001 Å. A 12 Å non-bonded cutoff was applied, and periodic boundary conditions were applied at constant volume. Every 25th step was printed to the output file. The trajectory was written to the “mdcrd” file every 500 steps, and the final coordinates were written to the restart file.

(III) Step 3: Constant Pressure Dynamics (NPT)

Constant pressure dynamics were performed on the system for a total of 200 ps. NPT is representative of the isothermal-isobaric ensemble, where moles (N), pressure (P) and temperature (T) are held fixed, and the volume of the unit cell is adjusted by small amounts on each step. Constant pressure dynamics was performed for 100,000 steps, with every 100th step printed to the output file. The time step was set at 0.02 ps, and the SHAKE algorithm was applied. Periodic boundary conditions were used, and a 12 Å non-bonded cutoff was set to truncate non-bonded interactions. The flag for constant pressure was used with anisotropic pressure scaling, and the pressure relaxation time was set at 2 ps. Simulation trajectories were written to a “mdcrd” file, and for every 100 steps, energy information was written to an output file. The density of the system was plotted over the course of the NPT run, and stabilization of the density as a function of time was an indication that the constant pressure dynamics step was complete.

(IV) Step 4: Constant Volume Dynamics (NVT)

Constant volume dynamics, the isothermal-isochoric ensemble, comprised the bulk of the simulations. The input file specified use of a 0.02 ps time step while employing the SHAKE algorithm to fix the hydrogen geometries. The temperature of the system was set at 300 K. Periodic boundary conditions were applied at constant volume, with a non-bonded cutoff distance of 12 Å. For every 100 steps energy information was written to an output file, while for every 500 steps the coordinates were written to a “mdcrd” file.

(V) Step 5: Analysis of Trajectories

The simulation trajectories were analyzed in part by using the PTRAJ module in AMBER. The RMSD and B-factor values were estimated for the simulation trajectories using ptraj, as were inter-atomic distance variations.

RMSD calculations were performed on the alpha-carbons (C α) in the protein backbone. RMSD measures the average distance of displacement between the C α atomic coordinates and the initial reference structure as function of time. An RMSD value is expressed in units of length, most commonly the Ångström (Å).

$$RMSD(t) = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i^t - x_i^I)^2 + (y_i^t - y_i^I)^2 + (z_i^t - z_i^I)^2}$$

The x, y and z terms represent the Cartesian coordinates of every n atom in the system at time initial (*I*) and time = *t*. Calculation of the RMSD is a good quantitative measure for evaluating the stability of the protein.

RMSF (Root Mean Square Fluctuations) calculations were performed on all backbone and side chain atoms as a whole for each residue in the protein. This is a measure of deviation of each residue from a reference position.

$$RMSF = \sqrt{\frac{1}{t} \sum_{t_j=1}^t (x_i(t_j) - x_i)^2 + (y_i(t_j) - y_i)^2 + (z_i(t_j) - z_i)^2}$$

In the above equation, t is the time over which the calculation is averaged, and x_i , y_i and z_i are reference coordinates. This data is useful for analyzing the flexibility of the system.

The B factors, also known as the Debye-Waller factor, describes the degree to which atoms fluctuate about the time averaged molecular structure. In X-ray crystallography, it describes the magnitude to which the electron density is spread out, and indicates how static or dynamic mobility of each atom. It can indicate if there are any errors in model building, and is a multiple of the RMSF by a factor of $8/3\pi^2$.

$$B_i = 8/3\pi^2 RMSF_i^2$$

Non-bonded interaction energies were analyzed using the MM-GBSA script in AMBER. A set of structures were taken from the MD trajectories, and from these structures the electrostatic contribution to the free energy is calculated with either the Poisson-Boltzmann (PB) method, or by the generalized Born (GB) method implemented in Sander. My calculations used the generalized Born (GB) method, which is an approximation to the Poisson-Boltzmann equation, and takes the functional form:

$$G = \frac{1}{8\pi} \left(\frac{1}{\epsilon_0} - \frac{1}{\epsilon} \right) \sum_{i,j}^N \frac{q_i q_j}{f_{GB}}$$

where

$$f_{GB} = \sqrt{r_{ij}^2 + a_{ij}^2} e^{-D}$$

and

$$D = \left(\frac{r_{ij}}{2a_{ij}} \right)^2, a_{ij} = \sqrt{a_i a_j}$$

where ϵ_0 is the permittivity of free space, ϵ is the dielectric constant of the solvent, q_i is the point charge on particle i , r_{ij} is the distance between particles i and j and a_{ij} is the effective Born radius. A hydrophobic solvent accessible surface area term (SA) is included in the calculation. This method was used to calculate the following interactions:

- Amino terminal residues – phospholipid residues
- Amino terminal residues – core domain residues

3.6: Simulation Protocol

All computations were performed on a 128-processor SGI Altix 4700. All calculations were carried out using AMBER's force field (ff03) in the AMBER 9 software package. The Sander module carried out the energy minimizations and molecular dynamics computations. Periodic boundary conditions were applied to the system with a non-bonded cutoff of 12 Å to truncate VDW interactions. The particle-mesh Ewald method was used to treat long range electrostatic interactions, with a cubic B-spline interpolation. The SHAKE algorithm was used to constrain all bonds involving hydrogen atoms with a geometric tolerance of 0.00001 Å.

The system was energy minimized twice before the MD simulation. The energy minimizations utilized information from gradient codes. A restrained minimization was

first performed on the solvent, the counter ions and the phospholipid layers while keeping the protein fixed. Next, the entire system was minimized. The minimized system was then warmed for 20 ps to 300 K from an initial temperature of 10 K. Constant pressure dynamics were then performed on the warmed system for 100 ps. A time step of 0.02 ps was used for the molecular dynamics simulations. Finally, constant volume dynamics (NVT) was performed on the system.

CHAPTER 4 – RESULTS

4.1: Analysis of Systems Used in Simulation

Two systems were used in running the molecular dynamics simulations following the defined protocol. The two systems differed in the initial distance between the centers-of-mass of the phosphate groups of the top and bottom monolayers. This resulted in different volumes for the boxes of the two systems that were subjected to periodic boundary conditions. The volume difference corresponds to a different amount of explicit water molecules used as solvent and a different amount of sodium and chloride counter-ions used in order to keep to ionic concentration consistent between the two systems.

4.2: Simulation Times

Time scales used for the various steps in the MD simulations are reported in the following Table 4.1.

Table 4.1: Simulation Time Scales.

	Setup 1	Setup 2
Restrained Minimization	1000 steps	1000 steps
Minimization	2500 steps	2500 steps
Equilibration	20 ps	20 ps
NPT	200 ps	200 ps
NVT	26702.2 ps	26344 ps

4.3: Steps – MD simulation

(I) Minimization

Both restrained and unrestrained minimizations were performed on both systems. The restrained minimizations only minimized solvent molecules and counter-ions, whereas the unrestrained minimizations included the protein and lipid molecules.

(II) Equilibration

During this step, the systems were warmed to a target temperature of 300 K. This temperature was held constant for the remainder of the simulation.

(III) Constant Pressure Dynamics (NPT)

Constant pressure dynamics were performed on the equilibrated systems for a total of 200 ps. Over this period of time, the density of the systems eventually equilibrated to approximately 1.02 g/cm³. Figures 4.1 and 4.2 depict the density profiles for both systems.

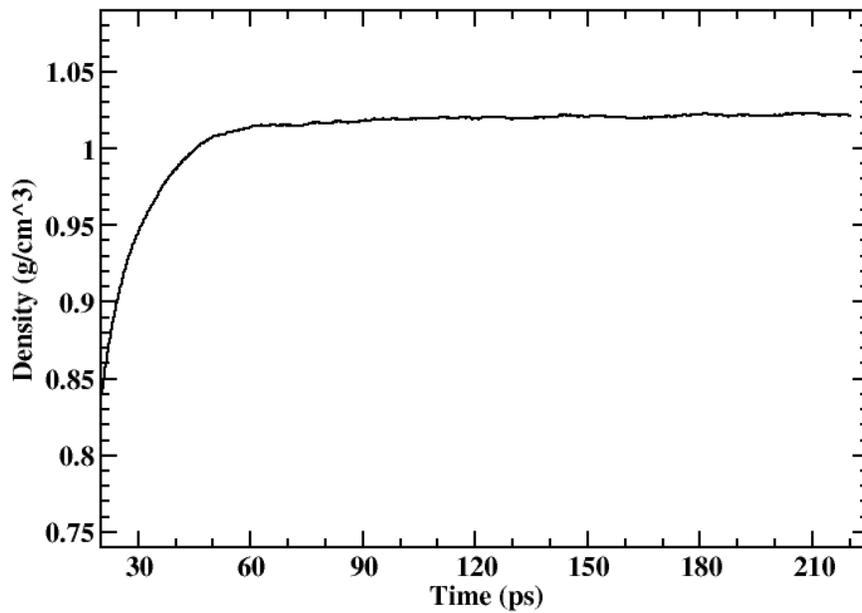


Fig 4.1: Density Plot for setup 1.

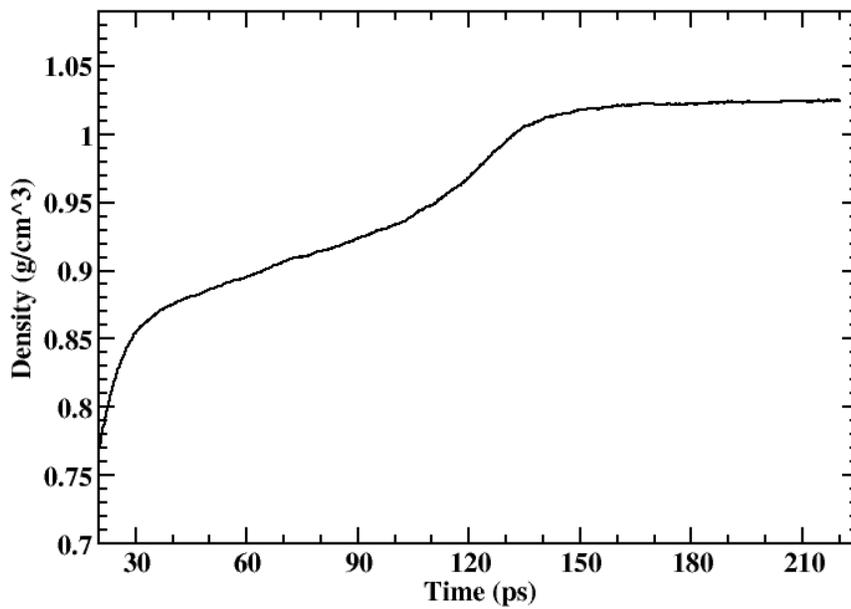


Fig 4.2: Density Plot for setup 2.

(IV) Constant Volume Dynamics (NVT)

Constant volume dynamics were performed on the two systems for the indicated time scales. The trajectories were analyzed upon completion of the NVT dynamics using various programs in AMBER, such as PTRAJ, MM-PBSA and AMBMASK.

4.4: Root Mean Square Deviations (RMSD)

Root mean square deviations of the alpha-carbon atoms with respect to the starting structure were analyzed over the course of the simulation using the PTRAJ program in AMBER. The RMSD plots were constantly monitored and were used to indicate the stability of the systems. Multiple RMSD plots were created to target specific areas of interest in the protein, including the core domain, each individual repeat of the core domain, the N-terminus, the helical regions of the N-terminus, the full-length protein, and the full-length protein lacking the coiled region of residues connecting the N-terminus and core domain. This analysis aided in understanding the conformational changes annexin I underwent during the simulation.

(I) RMSD Plot for Full-Length Annexin I Protein in setup 1

Figure 4.3 illustrates the RMSD plot for the full length protein, meaning all 343 alpha carbon atoms were used in calculating the deviations from the NVT trajectory of setup 1. The protein showed an overall continual rise in RMSD values from the initial frame until 3953 ps, corresponding to a value of 4.018 Å. Between 3953 ps and 9739 ps (at 4.313 Å) the values fluctuated between 2.893-4.399 Å with a mean value of 3.585 Å. During this time, the N-terminus appears to be ‘bouncing’ off of the phospholipids, and its first helix appears to remain parallel to the lipid surface. After 9739 ps, the values fall

rapidly, and from 10108 ps to 11188 ps, the values range from 2.448-3.402 Å, with a mean value of 2.860 Å. During this time, the first eight residues of the N-terminus make somewhat of a kink at the second turn of the alpha helix, and they appear oriented near perpendicular to the lipid surface. The octapeptide primary sequence is as follows: A-M-V-S-E-F-L-K. The majority of this sequence is made up of hydrophobic residues, and the terminal alanine's amino group is protonated ($-\text{NH}_3^+$). Therefore, the driving force behind this conformational change could be that the hydrophobic residues attempt to reach the non-polar interior of the phospholipids, and that an electrostatic attraction exists between the terminal alanine and the negatively charged layer. Further analysis using the MM-PBSA program over this time period revealed a relatively high electrostatic attraction between the terminal alanine and the entire phospholipid layer, with a calculated value of -172.22 kcal/mol.

Shortly after 11188ps the RMSD plot climbs steeply, until a peak is reached at 11770 ps. Between 11770-13332 ps, the deviations ranged between 3.158-4.206 Å, with an average value of 3.756 Å. During this time, the helices of the N-terminus appear to rise away from the phospholipid layer and towards repeat III of the core domain. At the carboxy-terminus of the second helix, Lys-26 has its sidechain fixed in close proximity with the phospholipid layer, so the N-terminus forms an approximate 45 degree angle from the surface of the lipid layer.

After 13332 ps, the plot drops rapidly again until a minimum is reached at 14545 ps. The N-terminus moves away from the core domain and back towards the phospholipid surface over this time duration. The plot rises slightly until 17016 ps is reached. From this point to the end of the simulation, at 26992.2 ps, the RMSD plot

appears to stabilize. The deviation values range between 2.366-3.766 Å, with a mean value of 3.090 Å. The N-terminus also appears to stabilize, and although it is not as close to repeat III of the core domain as it was between 11770-13332 ps, it remains pointed towards repeat III and also retains its secondary structure. As shown in Figure 4.4, by calculating the distance between Asp195, which is located in the A-helix of repeat III, and Ala2, it is possible to see that Ala2 approaches an equilibrium distance with Asp195 at approximately 17000 ps, just as the RMSD plot stabilizes. Asp195 was chosen because it consistently produced the strongest electrostatic attraction to Ala2 of all core domain residues throughout multiple MM-PBSA calculations.

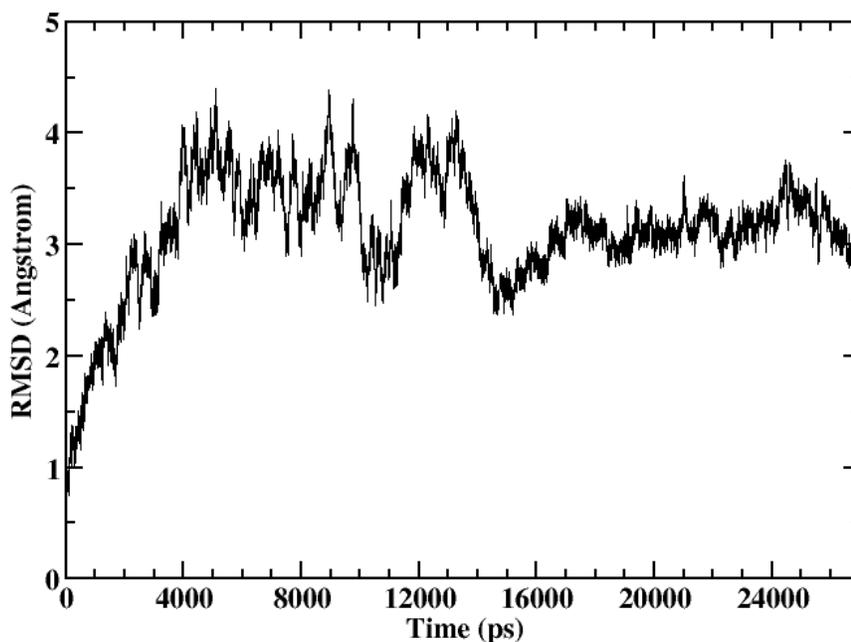


Fig 4.3: RMSD plot of α carbon atoms for the full-length protein in setup 1.

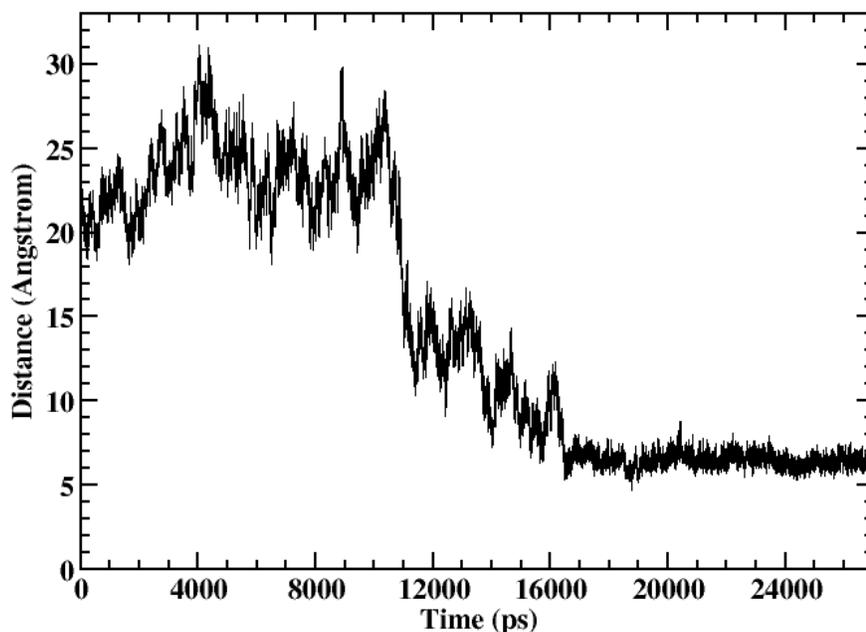


Fig 4.4: Distance plot between the centers-of-mass of alanine-2 and aspartate-195 over the course of setup 1.

(II) RMSD plot of core domain in setup 1

Figure 4.5 depicts the RMSD plot of only the core domain of annexin I in setup 1. The fluctuations are relatively low in this plot after 1000 ps, with a maximum value of 1.860 Å at 16632 ps, a minimum value of 0.783 Å at 2227 ps and a mean value of 1.180 Å. There is one obvious bump in the plot, from 15134-18429 ps. This can be attributed to a period of relatively strong interaction between the core domain and the N-terminus. Figure 4.6 depicts the RMSF values from the time period of 15315-16420 ps, during which there is a rise in the RMSD plot of the core domain. Figure 4.7 depicts the RMSF values from 16420-18430 ps, during which there is a drop in the RMSD plot of the core domain. Both Figures 4.6 and 4.7 show relatively higher values at residues 196-273, comprising repeat III of the core domain. Of particular interest, residues 190-197 make

up a loop region on the concave face of the core domain connecting repeat II to repeat III and also show relatively high fluctuation values. Based on MM-PBSA analysis shown later, aspartic acids 190, 195, 196, 199 and 201 consistently display the strongest electrostatic attractions to Ala2. From Figure 4.4, Ala2 comes to within an equilibrium distance of Asp195 at around 16000 ps, which is approximately the same time that the RMSD plot of the core domain (Figure 4.5) peaks. Therefore, we speculate that the rise in the RMSD plot of the core domain is in fact due to the flexibility of residues from repeat III interacting with the N-terminal domain.

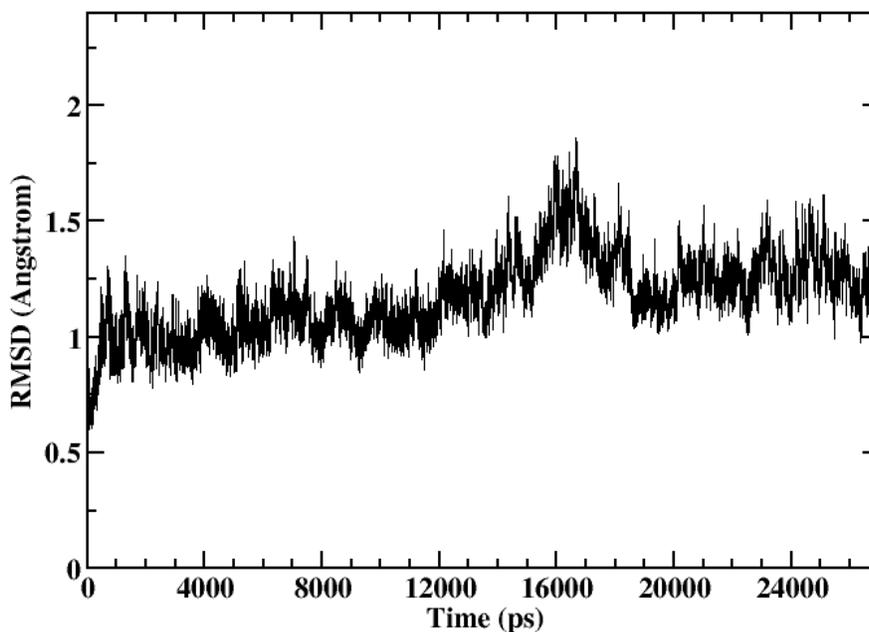


Fig 4.5: RMSD plot of α carbon atoms of the core domain in setup 1.

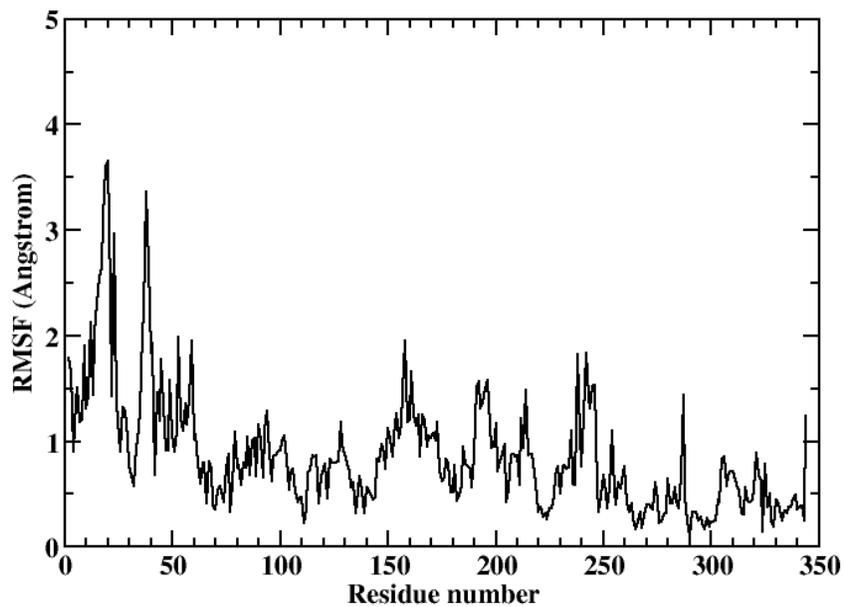


Fig 4.6: RMSF plot of α carbon atoms during 15315-16420 ps, the period in which there is a rise in the RMSD plot of the core domain in setup 1 from Figure 4.5.

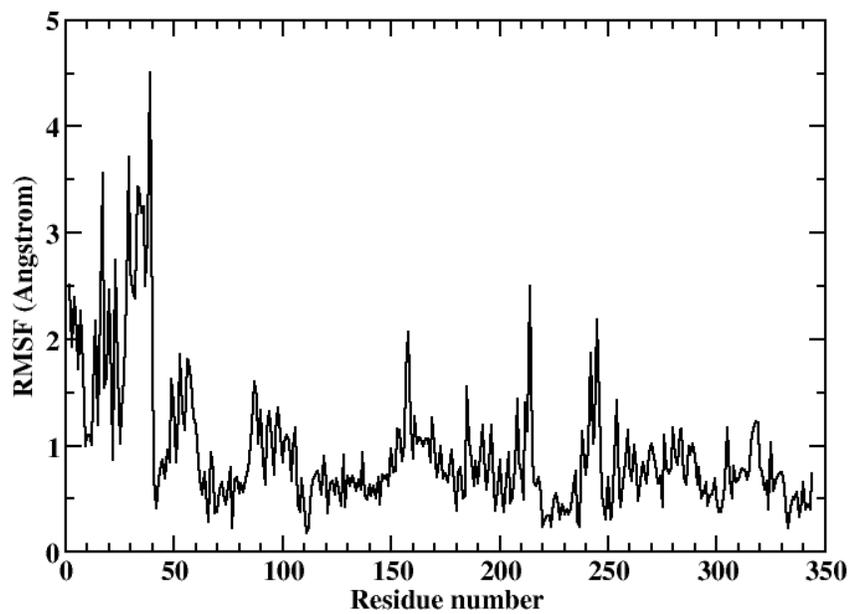


Fig 4.7: RMSF plot of α carbon atoms during 16420-18430ps, the period in which there is a drop in the RMSD plot of the core domain in setup 1 from Figure 4.5.

(III) RMSD Plot by Core Repeats in setup 1

Figure 4.6 depicts the RMSD values of the alpha carbons for each repeat (I-IV) of the core domain of annexin I during the simulation of setup1. Each repeat is made up of approximately 75 amino acids. It is clear that repeat III has the highest overall deviation from the starting structure, with a maximum value of 1.649 Å. This seems logical since the N-terminus was shown to come within close proximity to repeat III, especially the A-helix, as there appears to be some non-bonded interaction between repeat III and the N-terminus.

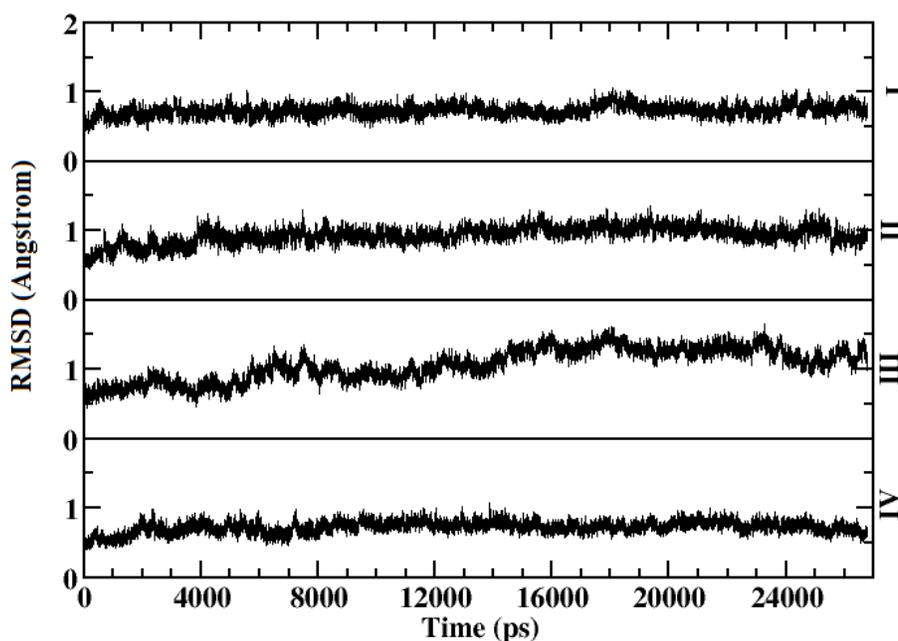


Fig 4.8: RMSD plots of α carbon atoms of annexin repeats I-IV in setup 1.

(IV) RMSD Plot of N-terminus in setup 1

Figure 4.7 depicts the RMSD plot of the 41 residues of the N-terminus in setup 1. The plot is consistent with the full-length RMSD, because the period of 11770-13332 ps experiences a rapid rise and decline in the plot. Again, during this time the N-terminus rises from the phospholipid surface to make an approximate 45 degree angle with repeat III of the core domain, and then falls back towards the surface where it appears to stabilize. The entire N-terminus has a range of 2.119-6.691 Å, and a mean value of 4.413 Å. Figure 4.8 depicts only the helical regions of the N-terminus, residues 2-26, and serves as a comparison to Figure 4.7 to illustrate the impact of the unstructured coiled of residues 27-41. The helical regions have a range of 0.368-2.891 Å, with a mean value of 1.168 Å.

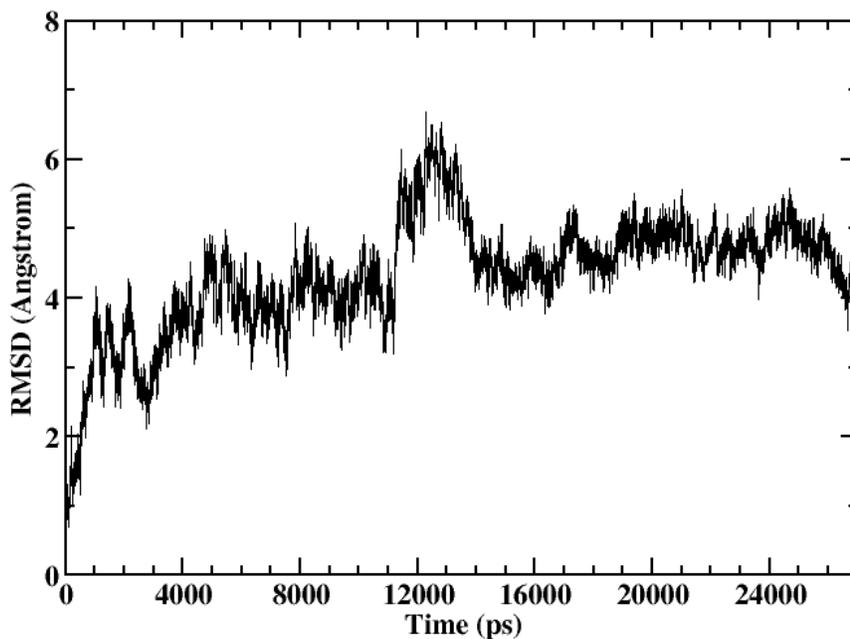


Fig 4.9: Residues 2-41 of setup 1, the entire N-terminus RMSD plot of α carbon atoms.

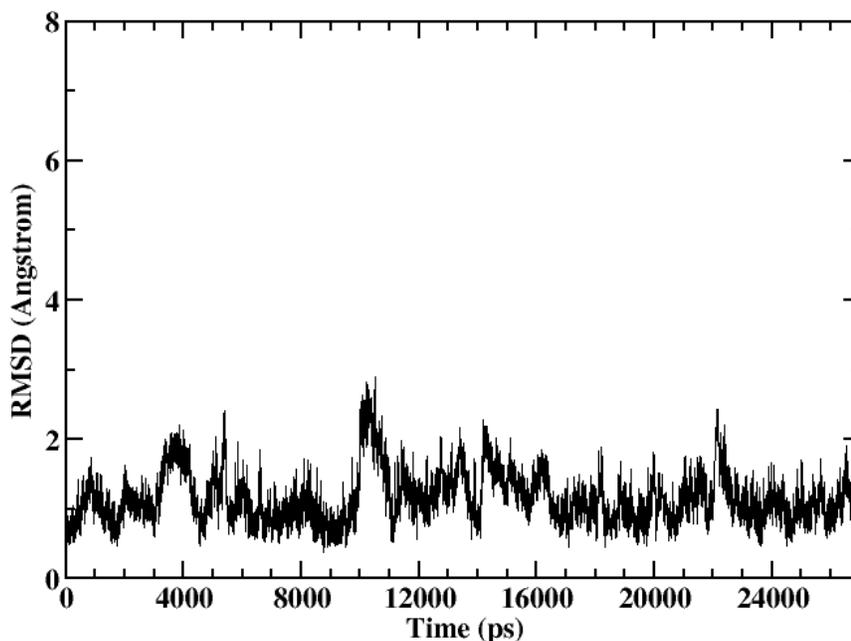


Fig 4.10: Residues 2-26, the helical regions of the N-terminus, RMSD plot of α carbon atoms of setup 1.

(V) RMSD Plot for Full-Length Annexin I in setup 2

The RMSD plot for the full-length protein in setup 2 is illustrated in Figure 4.9. There is less overall conformational change in the MD simulation for setup 2, due to the fact that the phospholipid layers initially are closer than in setup 1, resulting in greater steric hindrance of the protein and less space for it to move. The N-terminus remained parallel to phospholipid surface throughout this simulation, and was not observed to interact with the core domain. The plot stabilizes at 1850-16550 ps, with a range of 1.718-3.831 Å and a mean value of 2.572 Å. For the entire simulation, the range of RMSD values is 1.526Å-4.442 Å, with a mean value of 2.966 Å. There is one large jump in the plot at 16551 ps. The jump at 16551 ps results from a conformational change in residues 27-41, the unstructured coil connecting the N-terminal helices to the core

domain. Lacking any secondary structure, and therefore hydrogen bonds to other residues, this unstructured coil can undergo conformational changes with more ease than the rest of the highly alpha-helical protein. This is verified by comparing the RMSD plots of all 41 residues of the N-terminus, depicted in Figure 4.10, to only residues 2-26, that is the helical region, depicted in Figure 4.11. The values of this plot range from 0.583-2.774 Å, with a mean value of 1.61 Å. When residues 27-41 are excluded from the calculation, the plot appears very stable, yet when those residues are included, the jump at 16551 ps is very pronounced.

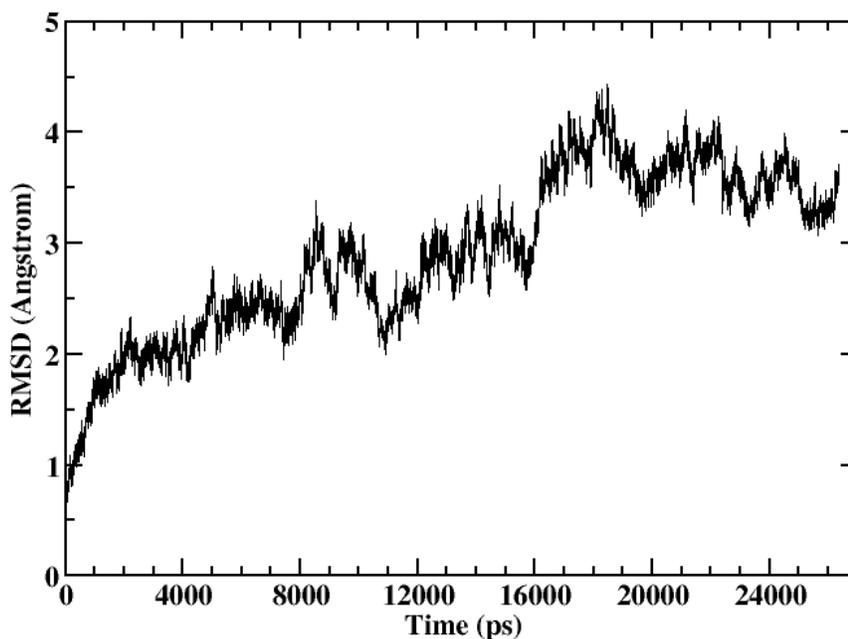


Fig 4.11: RMSD plot of α carbon atoms for the full-length protein in setup 2.

(VI) RMSD Plot of the Core Domain in setup 2

The RMSD of the core domain for setup 2 is illustrated in Figure 4.12. The plot appears very stable throughout the simulation and experiences very small fluctuations.

The core domain has been shown to be very stable, resistant to proteolysis and highly conserved in biochemical assays, and this plot verifies those findings. The values of this plot range between 0.854-1.860 Å, with a mean value of 1.333 Å.

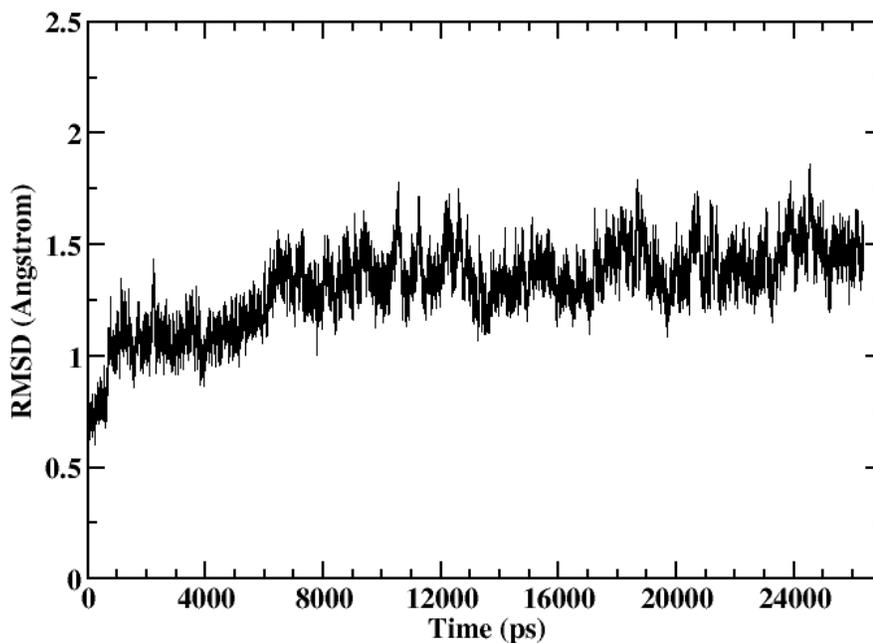


Fig 4.12: RMSD plot of α carbon atoms for the core domain in setup 2.

(VII) RMSD Plot by Core Repeats in setup 2

Figure 4.13 illustrates the RMSD plots of the four repeats of annexin I over the course of setup 2. Repeat III in this simulation did not show the relatively high degree of deviation that was displayed in setup 1 (Figure 4.6). This seems logical since the N-terminus was not observed to interact with repeat III in setup 2 as it did in setup 1.

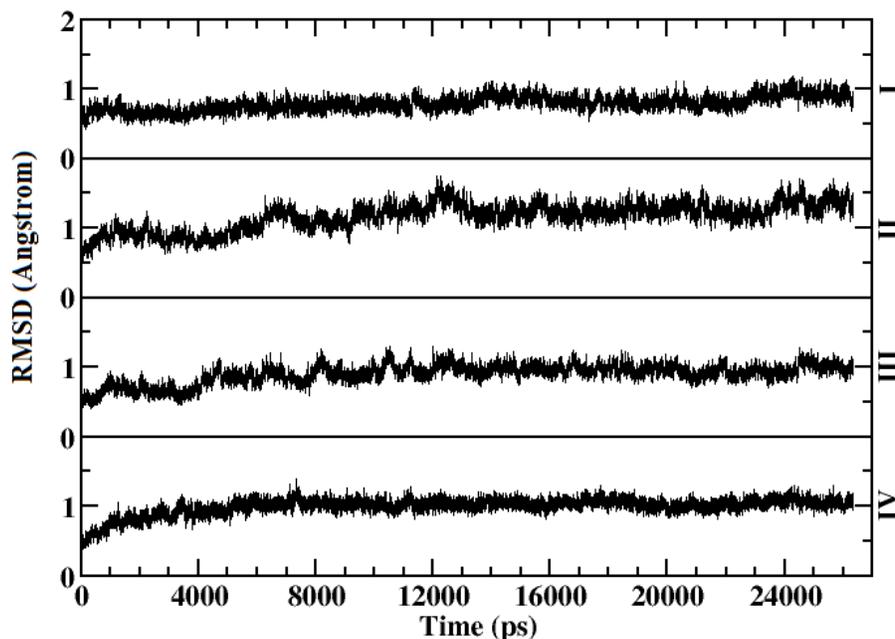


Fig. 4.13: RMSD plot of α carbon atoms of annexin repeats I-IV in setup 2.

(VIII) RMSD Plot of N-terminus in setup 2

Figure 4.12 depicts the RMSD of alpha carbon atoms of all 41 residues of the N-terminus during setup 2. The plot is consistent with the full-length RMSD, because around 16300 ps a rapid jump is seen in the plot. The unstructured coil region, residues 27-41 of the N-terminus, is presumed responsible for the rapid leap in value, since the helical regions of the N-terminus, residue 2-26, show in Figure 4.13, appear to have stable RMSD values. The protein is more sterically hindered in setup 2 as well, as the phospholipid layers are approximately 4 Å closer together, leaving less space for the protein to undergo significant conformational changes. It would be reasonable, therefore, for the unstructured coil region, lacking any hydrogen bonds which play a role in stabilizing the helix conformation, to be the region of greatest flexibility.

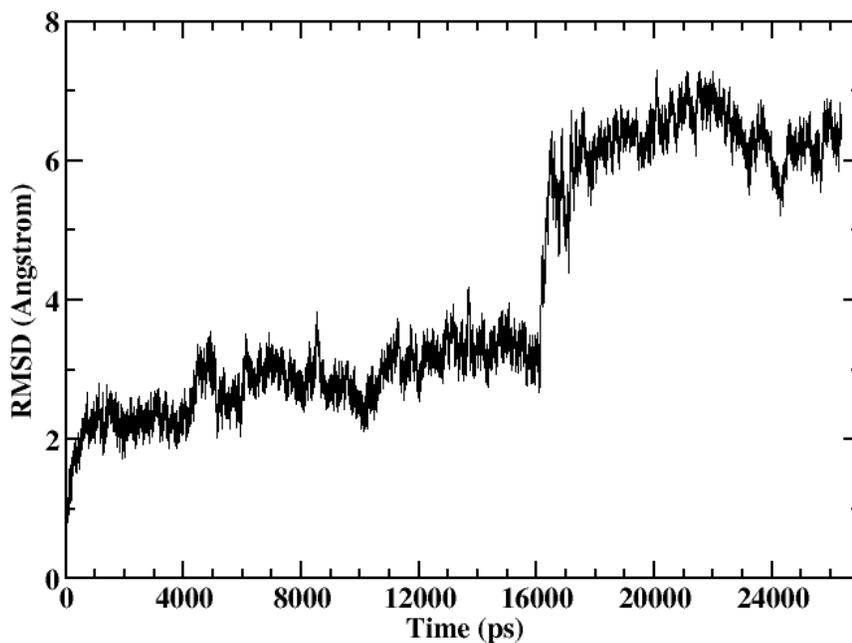


Fig 4.14: RMSD plot of α carbon atoms of the 41 residues of the N-terminus in setup 2.

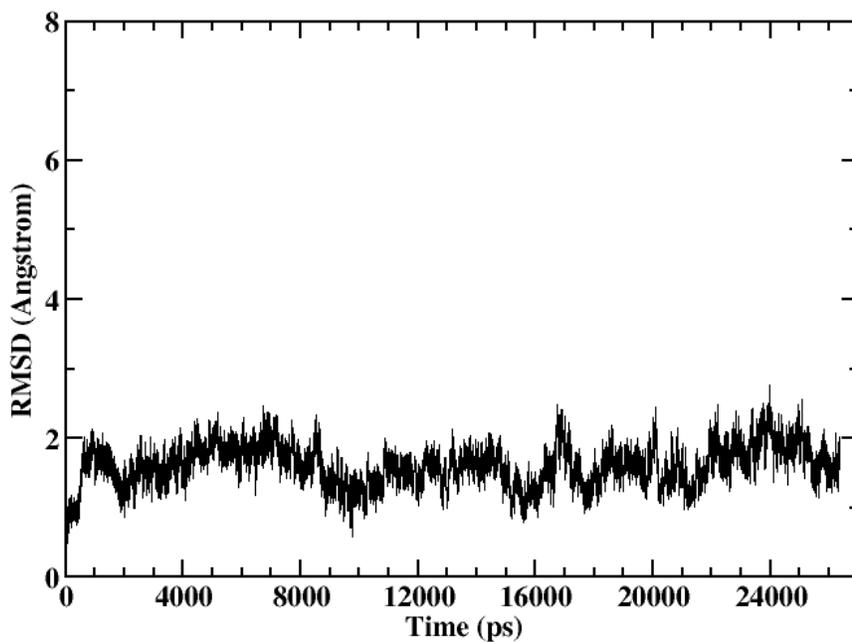


Figure 4.15: RMSD plot of α carbon atoms of the residues 2-26, the helical region of the N-terminus in setup 2.

4.5: Conclusions Drawn from RMSD Plots

Setup 1 appears to give a more accurate picture of how annexin I interacts with phospholipid bilayers to induce membrane aggregation and fusion, since the protein is not as sterically hindered as it is in setup 2. The rises and falls in the RMSD plot of the full length protein in setup 1 (Figure 4.3) are for the most part caused by conformational changes in the entire N-terminus, whereas the unstructured coil of residues 27-41 appear to be mostly responsible for the deviations in setup 2.

A significant difference was observed in the RMSD plots of repeat III of the core domain between setup 1 and setup 2. Setup 1 shows a higher deviation in this repeat due to non-bonded interactions with the N-terminus, whereas these interactions are not half as strong in setup 2.

4.6: RMSF

The root mean square positional fluctuations of the protein backbone residues, including the alpha-carbon, the carbonyl carbon and the nitrogen, were analyzed using the PTRAJ program in AMBER. RMSF plots were created using the protein backbone atoms (C α , N, C) as well as the sidechain heavy atoms for the entire simulations and are represented in Figures 4.14, 4.15, 4.16, and 4.17. RMSF values were compared for the two systems to understand how the protein fluctuates on average when in close proximity to phospholipids layers.

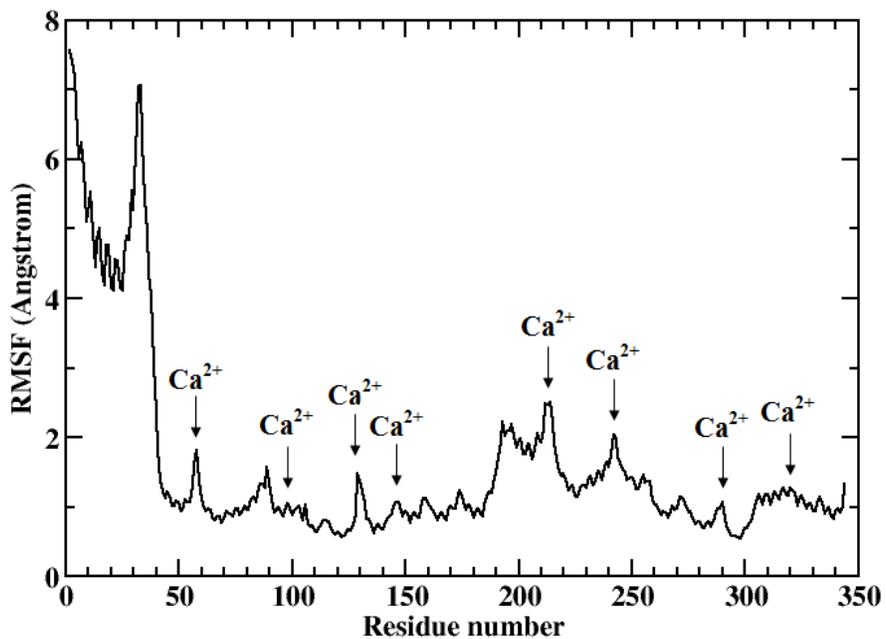


Fig 4.16: RMSF of backbone heavy atoms per residue in setup 1. Calcium binding residues are indicated with arrows.

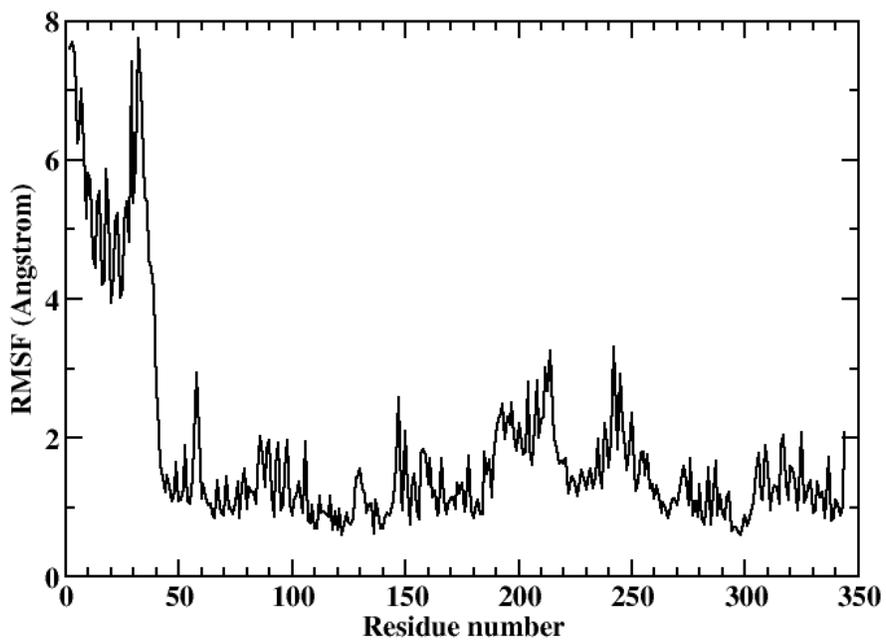


Fig 4.17: RMSF of sidechain heavy atoms per residue in setup 1.

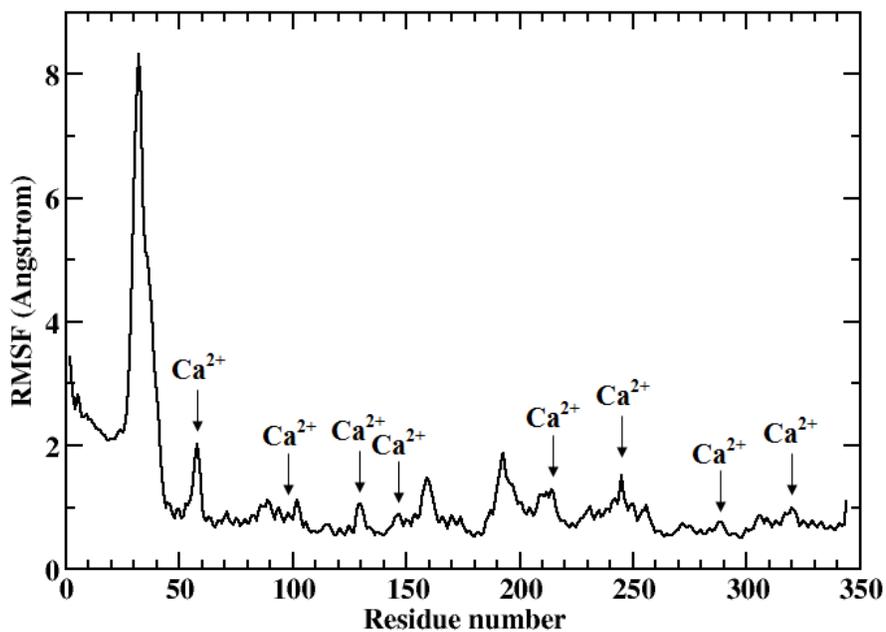


Fig 4.18: RMSF of backbone heavy atoms per residue in setup 2. Calcium binding residues are indicated with arrows.

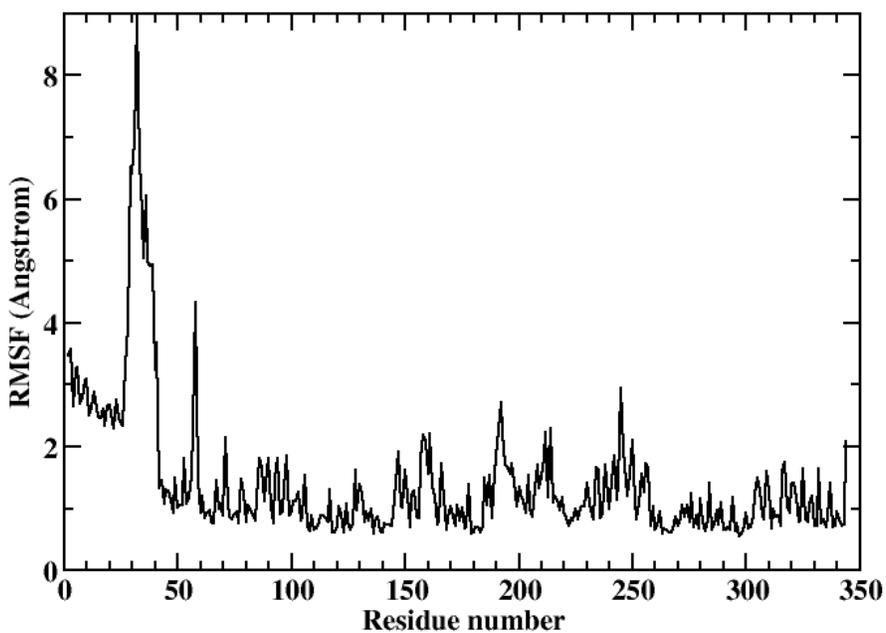


Fig 4.19: RMSF of sidechain heavy atoms per residue in setup 2.

4.7: Observations Made from RMSF Plots

The RMSF curves showed similar patterns for setup 1 and setup 2. The RMSF curves from both setups show that within the core domain of the protein, the calcium binding residues on average display a higher fluctuation than non-coordinating residues. Also consistent between both setups is that repeat III of the core domain, comprised of residues 196-273, displays a higher fluctuation than repeats I, II and IV. This is consistent with the structural data reported by Rosengarth *et al.* in that upon calcium binding, the N-terminus is ejected from a hydrophobic pocket in repeat III and is replaced by the D helix.⁶ Repeat III is therefore expected to adopt a conformation necessary for calcium and phospholipid binding and would be expected to display higher fluctuation values.

Another similarity between the RMSF plots of the two setups is that the N-terminal region displays much higher fluctuation values than the core domain. Particularly, residues 27-41, the unstructured coil, displays the highest fluctuations. These residues, lacking the hydrogen bonds necessary in stabilizing the helical conformation, would be expected to fluctuate more so than the helical residues. A difference between setup 1 and setup 2, however, is observed at the helical regions of the N-terminus, residues 2-26. Residues 2-26 have significantly higher fluctuation values in setup1 than in setup 2, consistent with the observation that those residues interact with repeat III in setup 1. Since setup 2 is more sterically hindered than setup 1, lower fluctuation values of the N-terminal helices is consistent with simulation conditions. Moreover, the N-terminal helices underwent very little conformational changes in setup 2. In fact, the only part of the protein able to fluctuate in setup 2 was the

unstructured coil connecting the N-terminus and core domain. Because there was greater steric hinderance in setup 2, the N-terminus was unable to interact with repeat III, which in turn led to lower fluctuation values of repeat III residues in setup 2 than setup 1. The mean fluctuation value for repeat III residues in setup 1 is 1.515 Å, whereas the mean value for this repeat in setup 2 is 0.911 Å.

Residues A-2, M-3 and V-4 display the highest fluctuation of backbone atoms during setup 1, with values of 7.573, 7.406 and 7.196 Å respectively. Residues Gly31, Ser32 and Ala33, located in the unstructured coil, display the highest fluctuation of backbone atoms during setup 2, with values of 8.329, 7.522 and 7.432 Å respectively.

4.8: B Factors

The B factor of the alpha carbon atoms over the course of the simulation was also calculated and compared with the B factor taken from X-ray crystallography data. By specifying the keyword “bfactor” in the ‘atomicfluct’ input file, the data is output as B-factors rather than atomic positional fluctuations (which simply means the results are multiplied by a factor of $(8/3)\pi^2$). B factors reflect the fluctuation of a given alpha carbon atom about its average position, which gives valuable information regarding the dynamics of a protein. Figures 4.20 and 4.21 illustrate the B factors of setup 1 and setup 2, respectively. As shown, the calculated B factors from the simulations align well with the B factor values taken from X-ray data. This helps to validate that the simulation set-up and the force field ff03 used during the simulation are rational choices giving fairly consistent data with experimental results

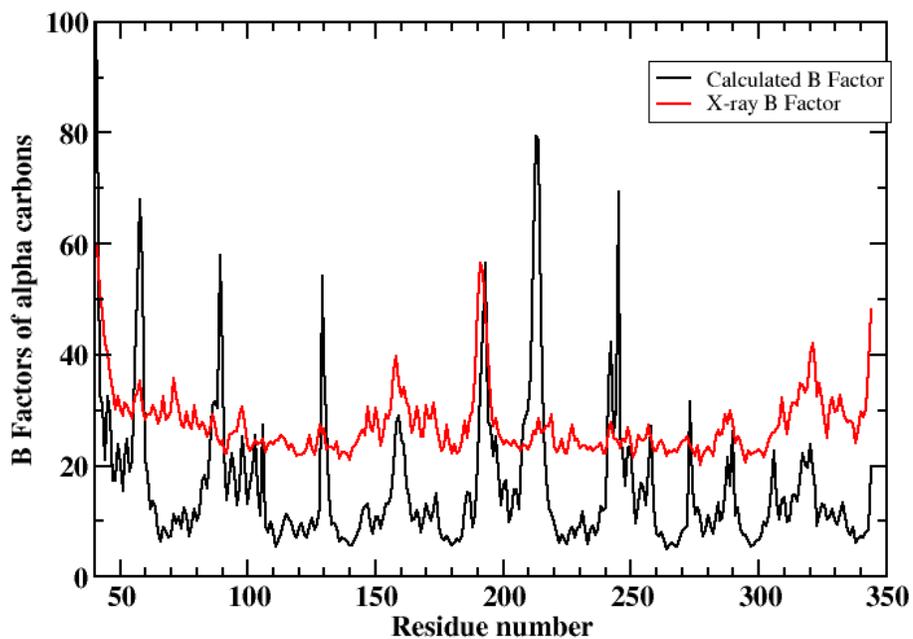


Fig 4.20: Calculated (black) and X-ray (red) B factors of α carbon atoms in setup 1.

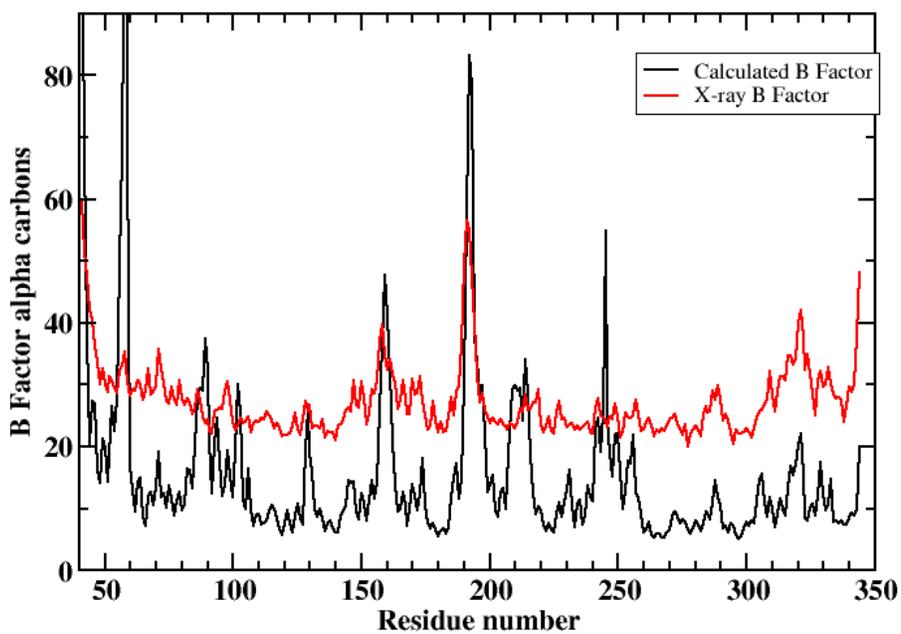


Fig 4.21: Calculated (black) and X-ray (red) B factors of α carbon atoms in setup 2.

4.9: MM-PBSA

Non-bonded interactions between residues in the protein and phospholipids were also analyzed. This helped to identify which residues had a greater impact in the overall conformational changes the protein underwent. This was accomplished using MM-PBSA module in the AMBER suite of programs. This method combines molecular mechanics and continuum solvent approaches to estimate binding energies. The initial sets of structures were collected from MD simulation and then used in the AMBER program to calculate the total binding energy of the system as shown in the equation:

$$\Delta G_{\text{bind}} = \hat{G}_{\text{complex}} - [\hat{G}_{\text{protein}} + \hat{G}_{\text{ligand}}]$$

Specifically, we analyzed:

- Amino terminal residues – bottom layer phospholipid interactions
- Amino terminal residues – core domain residues

Van der Waals interactions were truncated at 12 Å and were found to have little influence in the non-bonded interactions. The calculations indicated that the primary non-bonded interactions were comprised of electrostatic forces. The electrostatic contribution to the solvation free energy is calculated using the Poisson-Boltzmann method in the pbsa program of AMBER. Four MM-PBSA calculations were performed on setup 1, and three calculations were performed on setup 2. Each calculation was performed over a particular period of time based on the results from the RMSD plots. Periods in which the RMSD plots changed significantly were chosen to be analyzed by MM-PBSA calculations so that a correlation could be made between conformational changes in the protein backbone and the influence of the N-terminus's non-bonded interactions on those conformational changes. The time intervals that were chosen to be

analyzed are circled in Figures 4.22 and 4.23. Figures 4.24-4.25 show MM-PBSA calculations from 10240-11150 ps during which the RMSD plot (Figure 4.3) first stabilizes around 3 Å in setup 1. Figures 4.26-4.27 show MM-PBSA calculations in which the RMSD plot of setup 1 (Figure 4.3) rises to stabilize around 4 Å, from 11770-13330 ps. Figures 4.28-4.29 show MM-PBSA calculations in which the RMSD plot of setup 1 (Figure 4.3) rapidly falls from 4 Å to 2.5 Å over the period of 14540-15575 ps. Figures 4.30-4.31 show MM-PBSA calculations in which the RMSD plot of setup 1 (Figure 4.3) finally stabilizes at around 3 Å during the period of 17015-17640 ps. Figures 4.32-4.33 show MM-PBSA calculations in which the RMSD plot of setup 2 (Figure 4.11) falls from 3 Å to 2 Å during the period of 10692-12014 ps. Figures 4.34-4.35 show MM-PBSA calculations in which the RMSD plot of setup 2 (Figure 4.11) stabilizes around 3 Å over the period of 13672-14771 ps. Figures 4.36-4.37 show MM-PBSA calculations in which the RMSD plot of setup 2 (Figure 4.11) rises from 3.5 Å to 4.5 Å during the period of 16922-18469 ps. From each of these time periods, ten snapshots were generated at regular intervals, and the average structure from these snapshots was used in the calculation. Each residue in the N-terminus has a single value from the summation of all non-bonded interactions it has with either all 225 bottom layer phospholipids, or all 300 residues of the core domain. The charged residues, which include the terminal alanine, lysines, aspartates or glutamates, were the only residues to show significant non-bonded energies with either the core domain or the phospholipid layer. Electrostatic energies were also significantly greater than van der Waals energies. Included under each of the MM-PBSA Figures are Tables (4.2-4.8) which list specific

energies between the charged N-terminal residues and either the core domain or the phospholipid layer.

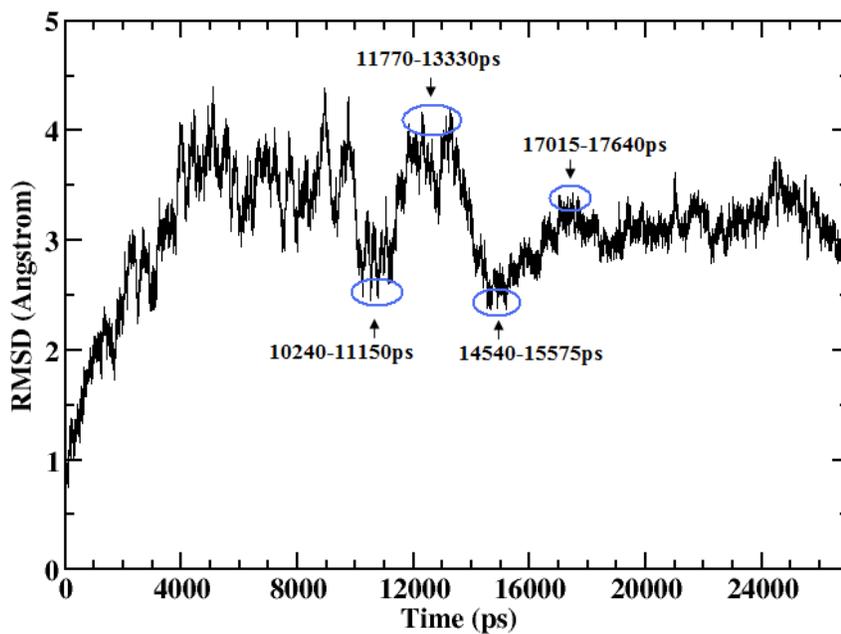


Fig 4.22: Regions from the RMSD plot of setup 1 (Fig 4.3) in which MM-PBSA calculations were performed are circled in blue.

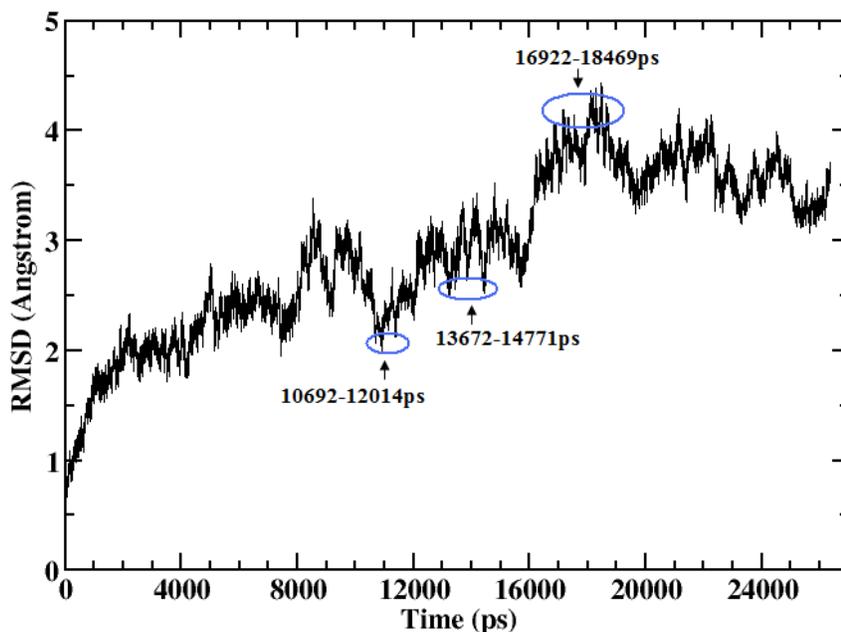


Fig 4.23: Regions from the RMSD plot of setup 2 (Fig 4.11) in which MM-PBSA calculations were performed are circled in blue.

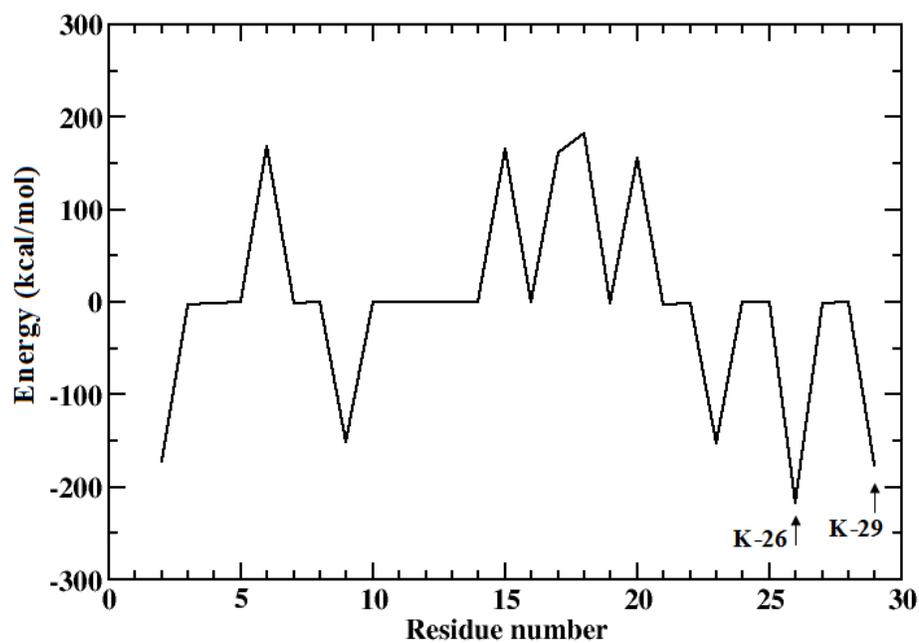


Fig 4.24: MM-PBSA calculation of non-bonded energy between each N-terminal residue and the bottom phospholipid layer during 10240-11150 ps in setup 1. Lysines 26 and 29 are indicated.

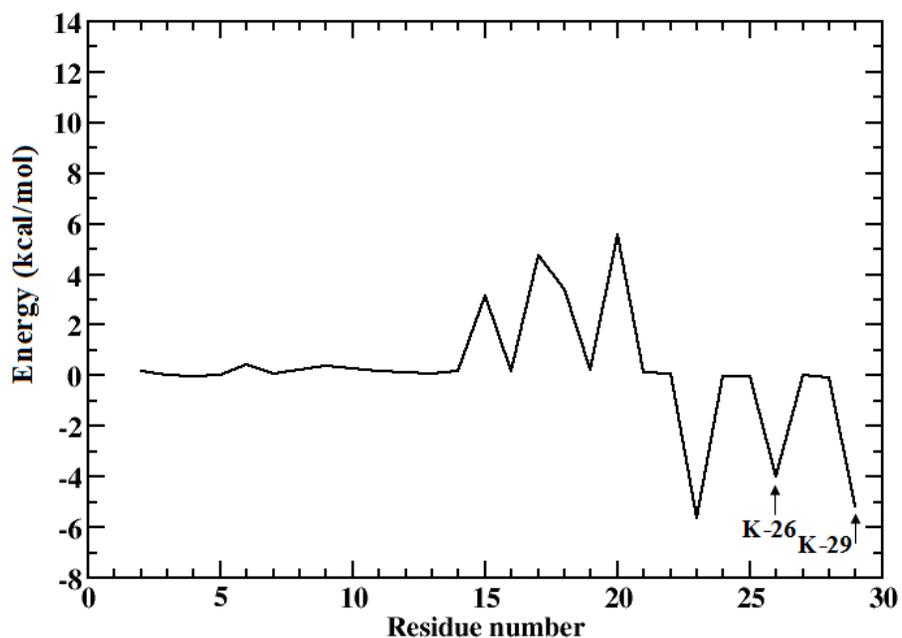


Fig 4.25: MM-PBSA calculation of non-bonded energy between each N-terminal residue and the core domain during 10240-11150 ps in setup 1. Lysines 26 and 29 are indicated.

Table 4.2: Setup 1 MM-PBSA results of charged residues from 10240-11150 ps.

Residue Number	Interaction Energy with Core Domain (kcal/mol)	Interaction Energy with Bottom Phospholipid Layer (kcal/mol)
A2	0.18	-172.11
E6	0.43	168.61
K9	0.36	-151.79
D15	3.17	164.75
E17	4.73	160.89
E18	3.39	182.41
E20	5.56	155.67
K23	-5.61	-151.99
K26	-3.97	-217.11
K29	-5.17	-176.35

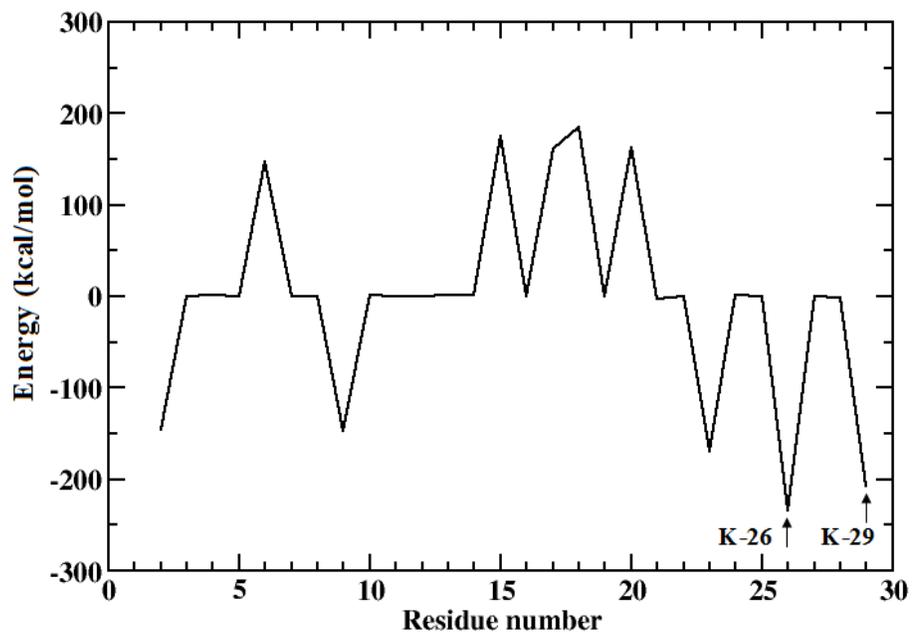


Fig 4.26: MM-PBSA calculation of non-bonded energy between each N-terminal residue and the bottom phospholipid layer during 11770-13330 ps in setup 1. Lysines 26 and 29 are indicated.

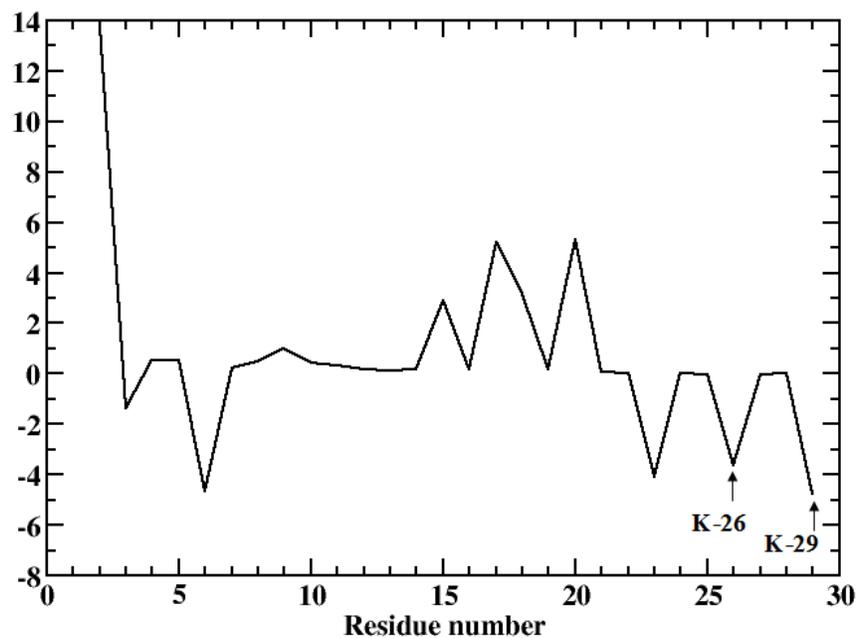


Fig 4.27: MM-PBSA calculation of non-bonded energy between each N-terminal residue and the core domain during 11770-13330 ps in setup 1. Lysines 26 and 29 are indicated.

Table 4.3: Setup 1 MM-PBSA results of charged residues from 11770-13330 ps.

Residue Number	Interaction Energy with Core Domain (kcal/mol)	Interaction Energy with Bottom Phospholipid Layer (kcal/mol)
A2	13.78	-145.38
E6	-4.65	146.97
K9	1.01	-147.45
D15	2.89	175.34
E17	5.19	161.29
E18	3.14	184.72
E20	5.33	162.23
K23	-4.11	-169.59
K26	-3.64	-233.34
K29	-4.74	-207.18

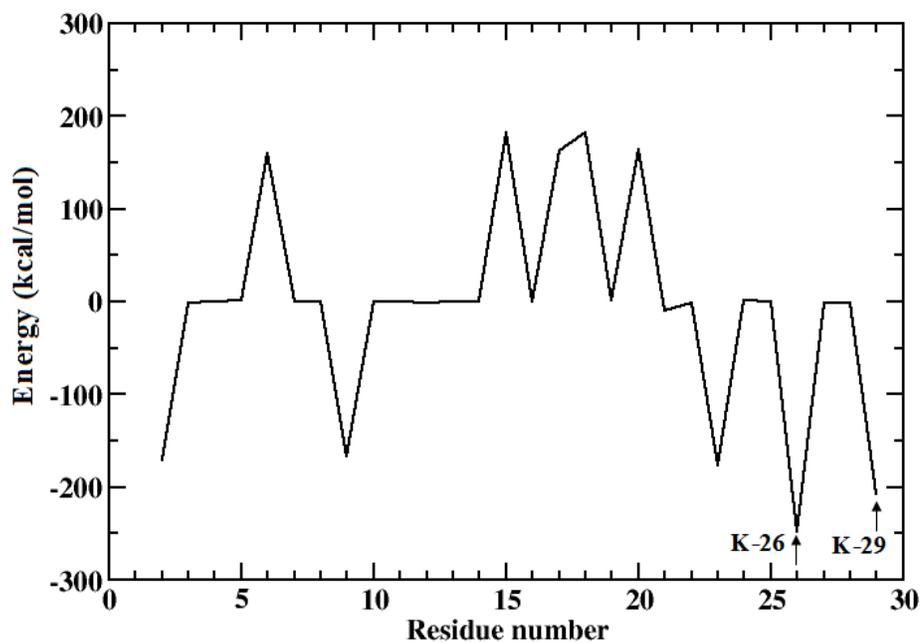


Fig 4.28: MM-PBSA calculation of non-bonded energy between each N-terminal residue and the bottom phospholipid layer during 14540-15575 ps in setup 1. Lysines 26 and 29 are indicated.

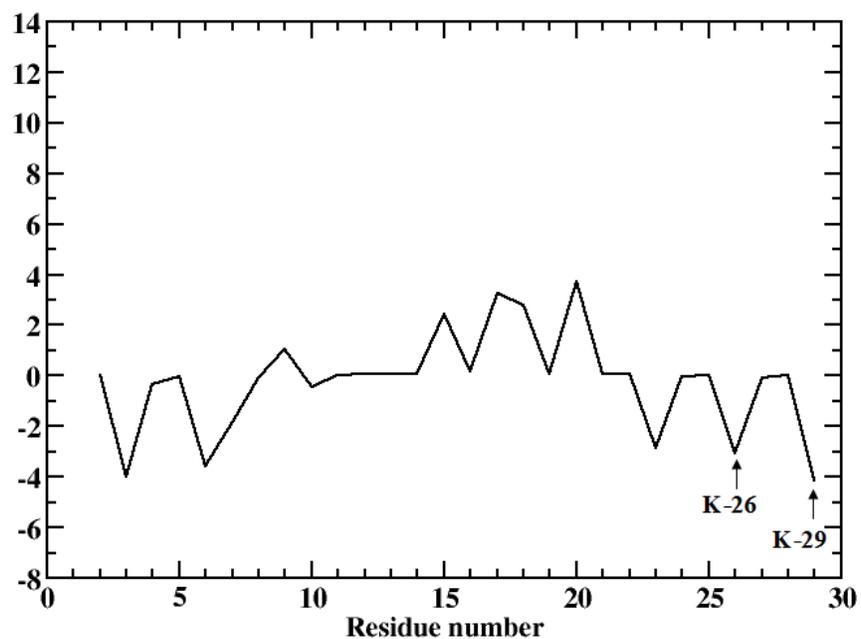


Fig 4.29: MM-PBSA calculation of non-bonded energy between each N-terminal residue and the core domain during 14540-15575 ps in setup 1. Lysines 26 and 29 are indicated.

Table 4.4: Setup 1 MM-PBSA results of charged residues from 14540-15575 ps.

Residue Number	Interaction Energy with Core Domain (kcal/mol)	Interaction Energy with Bottom Phospholipid Layer (kcal/mol)
A2	0.04	-170.68
E6	-3.56	159.41
K9	1.07	-166.24
D15	2.43	181.68
E17	3.26	161.79
E18	2.78	182.31
E20	3.73	164.49
K23	-2.86	-176.11
K26	-3.06	-248.51
K29	-4.12	-206.62

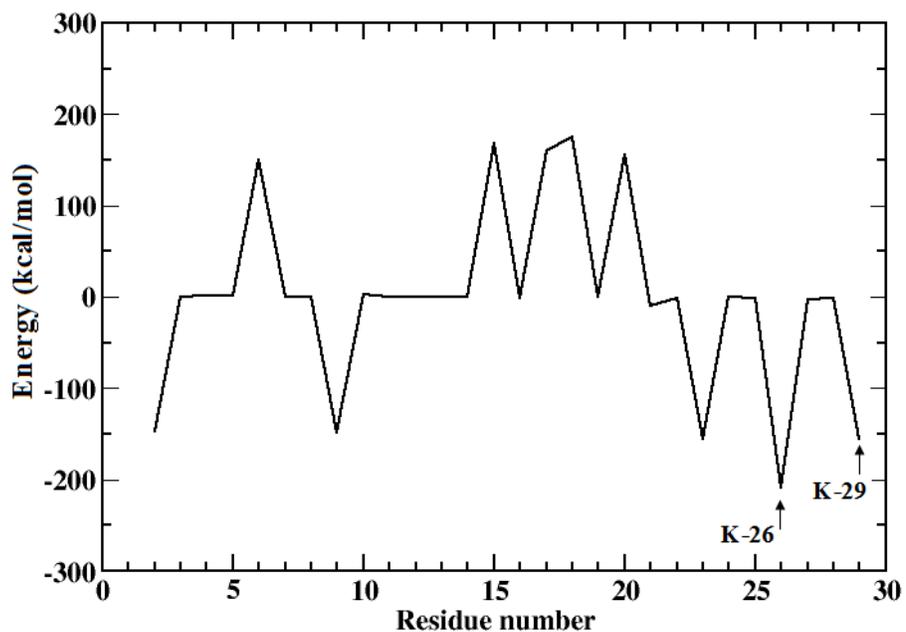


Fig 4.30: MM-PBSA calculation of non-bonded energy between each N-terminal residue and the bottom phospholipid layer during 17015-17640 ps in setup 1. Lysines 26 and 29 are indicated.

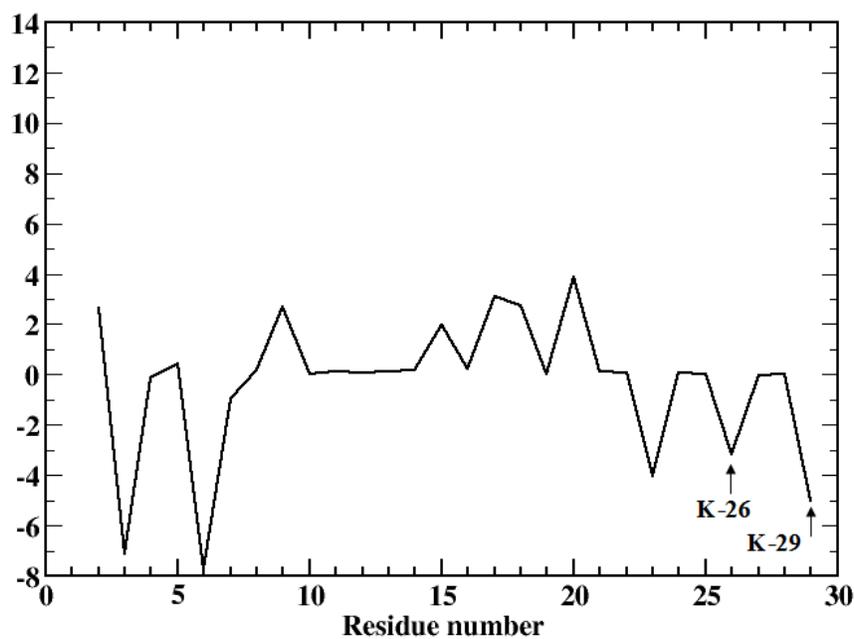


Fig 4.31: MM-PBSA calculation of non-bonded energy between each N-terminal residue and the core domain during 17015-17640 ps in setup 1. Lysines 26 and 29 are indicated.

Table 4.5: Setup 1 MM-PBSA results of charged residues from 17015-17640 ps.

Residue Number	Interaction Energy with Core Domain (kcal/mol)	Interaction Energy with Bottom Phospholipid Layer (kcal/mol)
A2	2.66	-147.58
E6	-7.75	149.84
K9	6.28	-148.31
D15	1.99	168.46
E17	3.1	159.89
E18	2.75	174.70
E20	3.88	155.60
K23	-4.03	-155.57
K26	-3.18	-208.97
K29	-5.03	-155.45

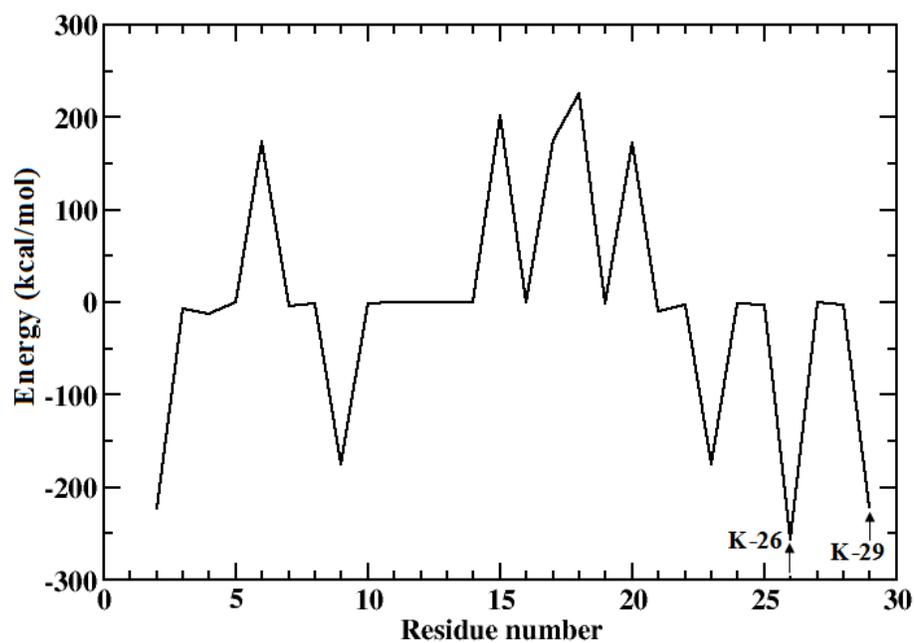


Fig 4.32: MM-PBSA calculation of non-bonded energy between each N-terminal residue and the bottom phospholipid layer during 10692-12014 ps in setup 2. Lysines 26 and 29 are indicated.

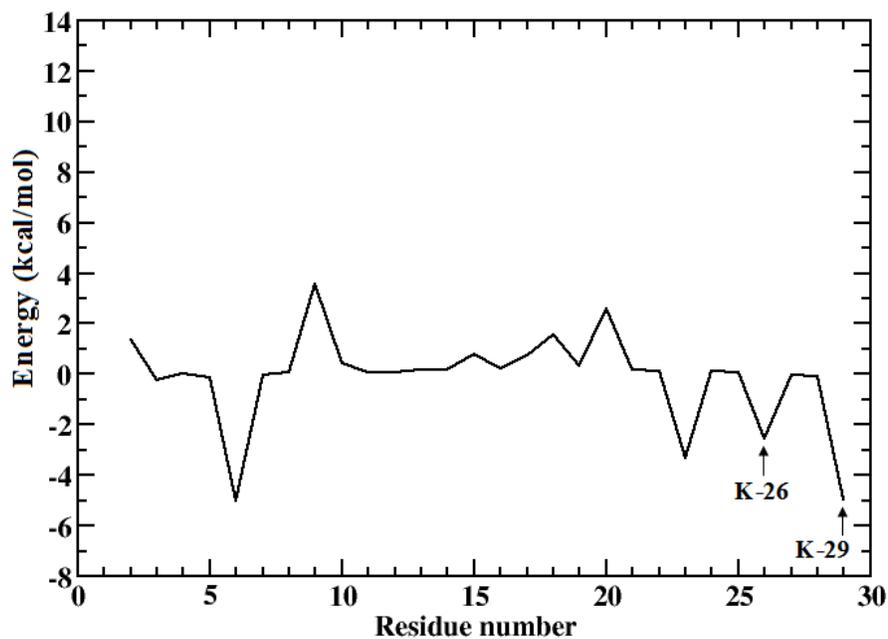


Fig 4.33: MM-PBSA calculation of non-bonded energy between each N-terminal residue and the core domain during 10692-12014 ps in setup 2. Lysines 26 and 29 are indicated.

Table 4.6: Setup 2 MM-PBSA results of charged residues from 10962-12014 ps.

Residue Number	Interaction Energy with Core Domain (kcal/mol)	Interaction Energy with Bottom Phospholipid Layer (kcal/mol)
A2	1.26	-222.39
E6	-5.65	173.08
K9	3.97	-174.62
D15	2.03	201.61
E17	3.70	175.24
E18	2.81	225.75
E20	5.34	172.12
K23	-4.83	-175.05
K26	-3.52	-256.87
K29	-5.34	-221.84

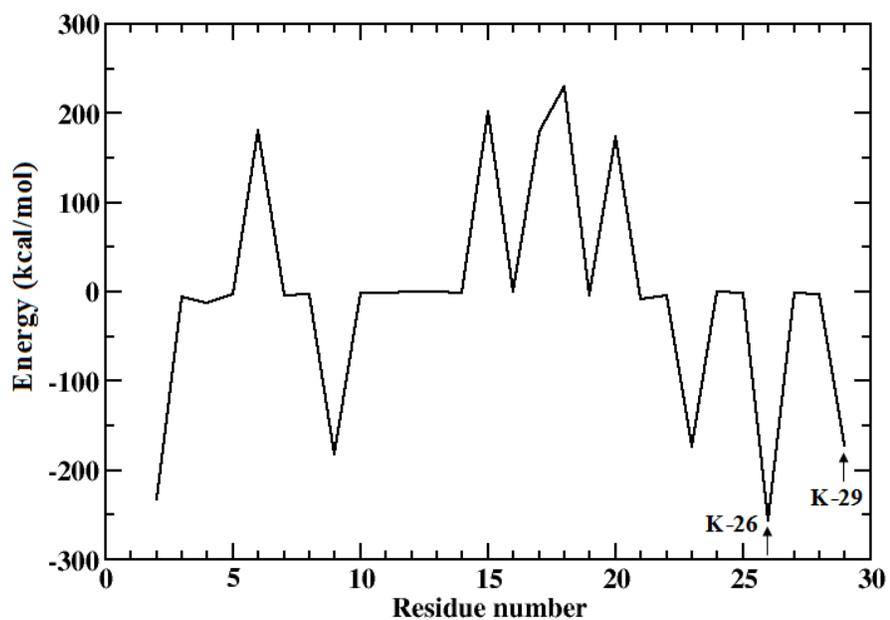


Fig 4.34: MM-PBSA calculation of non-bonded energy between each N-terminal residue and the bottom phospholipid layer during 13672-14771 ps in setup 2. Lysines 26 and 29 are indicated.

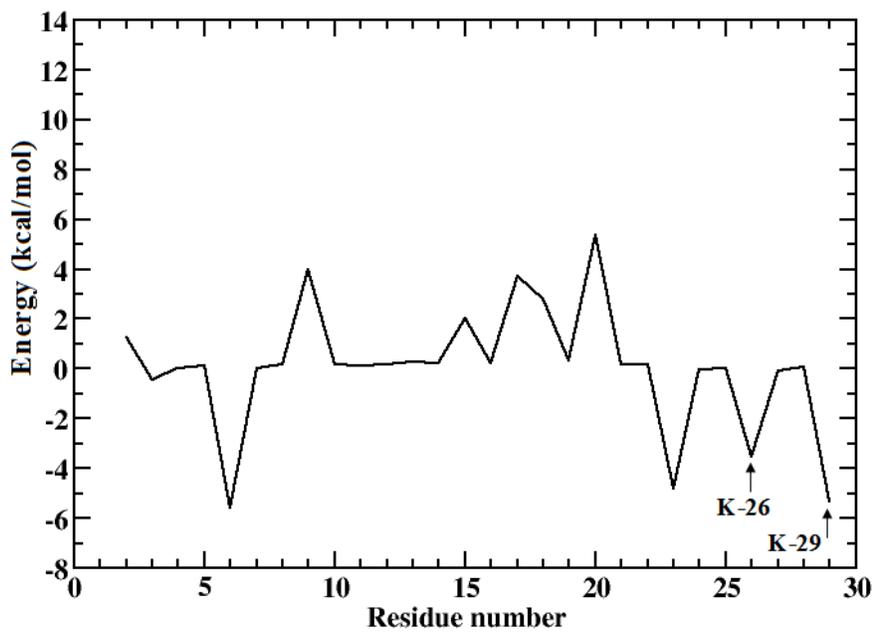


Fig 4.35: MM-PBSA calculation of non-bonded energy between each N-terminal residue and the core domain during 13672-14771 ps in setup 2. Lysines 26 and 29 are indicated.

Table 4.7: Setup 2 MM-PBSA results of charged residues from 13672-14771 ps.

Residue Number	Interaction Energy with Core Domain (kcal/mol)	Interaction Energy with Bottom Phospholipid Layer (kcal/mol)
A2	1.36	-232.22
E6	-5.01	181.19
K9	3.58	-182.49
D15	0.78	201.63
E17	0.75	179.79
E18	1.54	229.66
E20	2.60	173.84
K23	-3.31	-173.72
K26	-2.56	-255.69
K29	-4.97	-171.92

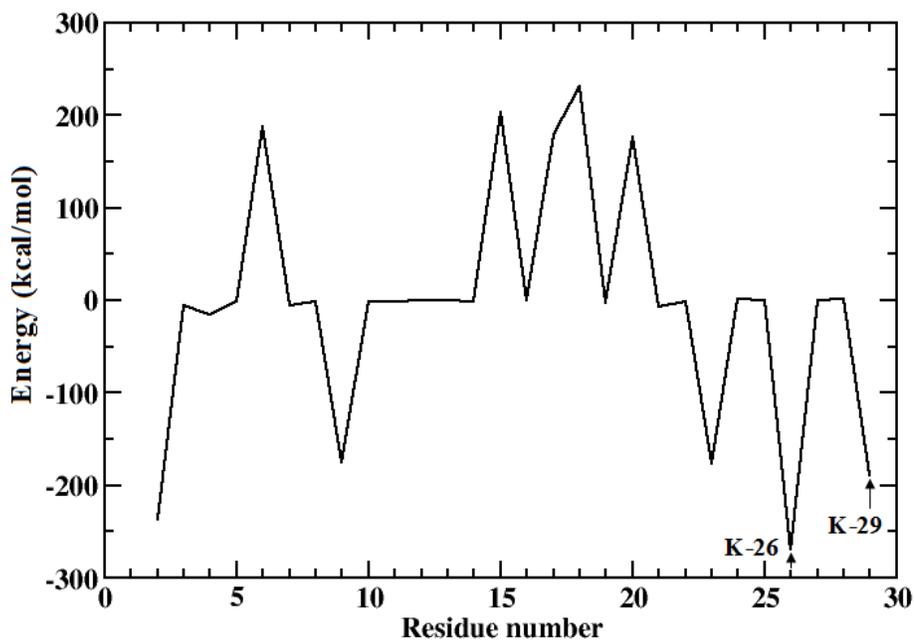


Fig 4.36: MM-PBSA calculation of non-bonded energy between each N-terminal residue and the bottom phospholipid layer during 16922-18469 ps in setup 2. Lysines 26 and 29 are indicated.

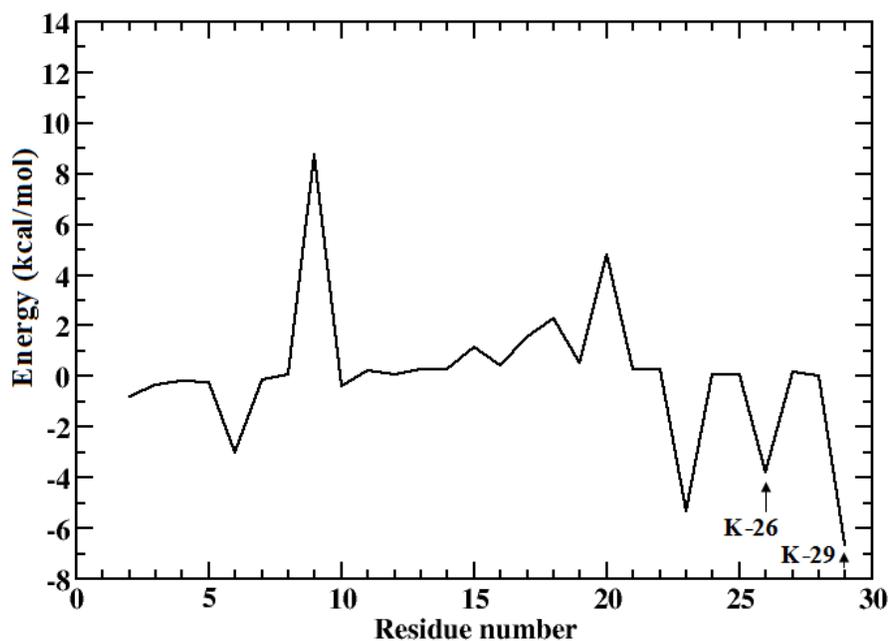


Fig 4.37: MM-PBSA calculation of non-bonded energy between each N-terminal residue and the core domain during 16922-18469 ps in setup 2. Lysines 26 and 29 are indicated.

Table 4.8: Setup 2 MM-PBSA results of charged residues from 16922-18469 ps.

Residue Number	Interaction Energy with Core Domain (kcal/mol)	Interaction Energy with Bottom Phospholipid Layer (kcal/mol)
A2	-0.78	-237.19
E6	-3.01	187.15
K9	8.77	-175.25
D15	1.16	202.81
E17	1.58	179.73
E18	2.28	231.63
E20	4.78	176.74
K23	-5.34	-176.11
K26	-3.76	-269.95
K29	-6.68	-188.46

4.10: Observations Made from MM-PBSA Plots

MM-PBSA calculations were applied to study the amino terminal interactions with the negatively charged phospholipid layer. As Bitto and Cho eloquently demonstrated by site-specific mutations, K-26 and K-29 are essential in annexin I for its membrane aggregation activity.¹⁴ They found that annexin I with residues 1-24 truncated showed the same aggregation activity as wild-type annexin, whereas truncated residues 1-29 showed lower residual activity. These two lysines both showed a very strong electrostatic attraction for the bottom layer phospholipids in our calculations. Furthermore, after viewing a snapshot of setup 1 using Pymol, the side chains of these residues underwent a conformational change in which they were oriented in very close proximity of the phosphoryl groups of the phospholipids. The conformational change in the side-chains of K-26 and K-29 imply that the ϵ -amino groups of these residues act as an electrostatic anchor for the peripheral binding of the N-terminal α -helix to a negatively charged phospholipid bilayer. This could help provide a clearer understanding of how the amino terminus acts as a second binding site for phospholipid bilayers and helps to promote membrane aggregation.

From the first time interval analyzed in setup 1, 10240-11150 ps, the terminal alanine residue shows the strongest attraction to the phospholipid layer of any other interval (-172.11kcal/mol), which is consistent with the RMSD plot in that the N-terminus is closest to the phospholipid layer at this time. The second interval in setup 1, 11770-13330 ps, shows the terminal alanine to have the weakest attraction to the phospholipid layer (-145.38 kcal/mol) and at this time the first helix of the N-terminus is farthest away from the phospholipid layer than at any other point. Overall, it can be

stated that the terminal alanine has very similar interaction energies with the core domain and with the phospholipid layer during intervals 1 and 3, calculations over 10240-11150 ps and 14540-15575 ps, which both have RMSD values of approximately 2.5 Å. Also, the terminal alanine has very similar interaction energies with the core domain and with the phospholipid layer during intervals 2 and 4, that is calculations over 11770-13330 ps and 17015-17640 ps, which have RMSD values of 4 and 3.5 Å respectively. Therefore, the statement can be made that when the protein backbone has a high RMSD value due to fluctuation of the N-terminus, Ala2 has a low interaction energy with the phospholipid layer since it has moved away from the phospholipid layer. Lysines 26 and 29 give fairly consistent values for all four time intervals in setup 1 with the phospholipid layer, with lysine 26 always having the greater value since it is in closer proximity to phospholipid layer.

The terminal alanine residue also consistently displayed a very strong attraction for the bottom layer of phospholipids for setup 2. Furthermore, the only significant interaction that residue had with the core domain was during 11770-13350 ps in setup 1, when it was in closest proximity to the core, and experienced a repulsive force. It is readily apparent that the negatively charged phospholipid layer contributed significantly higher non-bonded energies to the N-terminal residues than did the core domain, as can be seen from the values listed in Tables 4.2-4.8. This seems logical since there are 75 DOPG molecules in the phospholipid layer, each with a total charge of negative one, whereas the entire core domain has a net charge of zero. Overall, the plots displayed similar results, with the charged residues contributing the most to the non-bonded interactions of the N-terminus.

4.11: Inter Atomic Distances

Distance changes between calcium ions and their oxygen ligands were analyzed over the course of the simulation. Calcium ions are known to bind to oxygen atoms preferentially over any other element, and their coordination numbers range from 6 to 8. Based on data taken from x-ray crystallography experiments, the average distance between a calcium ion and its surrounding ligands in the crystal structure of full-length annexin A1 is 2.47 Å, with a range of 2.24-3.03 Å and a standard deviation of 0.1813.⁸ Stabilization of the inter-atomic distance at approximately 2.5 Å is an indication that an oxygen atom is coordinated with a calcium ion. Prior to plotting calcium-oxygen distance changes, the 'ambmask' program in AMBER was used to identify all oxygen atoms within 6 Å of each calcium ion. The 'ambmask' program acts as a filter which takes amber topology and coordinate files and applies a selection string to identify specific atoms. Distance analysis was calculated for all oxygen ligands that entered a calcium coordination sphere, including atoms from water molecules, phospholipid molecules and protein residues. The aim was to monitor a calcium ion binding to a phospholipid molecule, and also to observe ligand exchange. The following Figures 4.38-4.45 display the distance between a specified ion and its ligand as a function of time, with the y-axis measured in angstroms and labeled with the oxygen ligand.

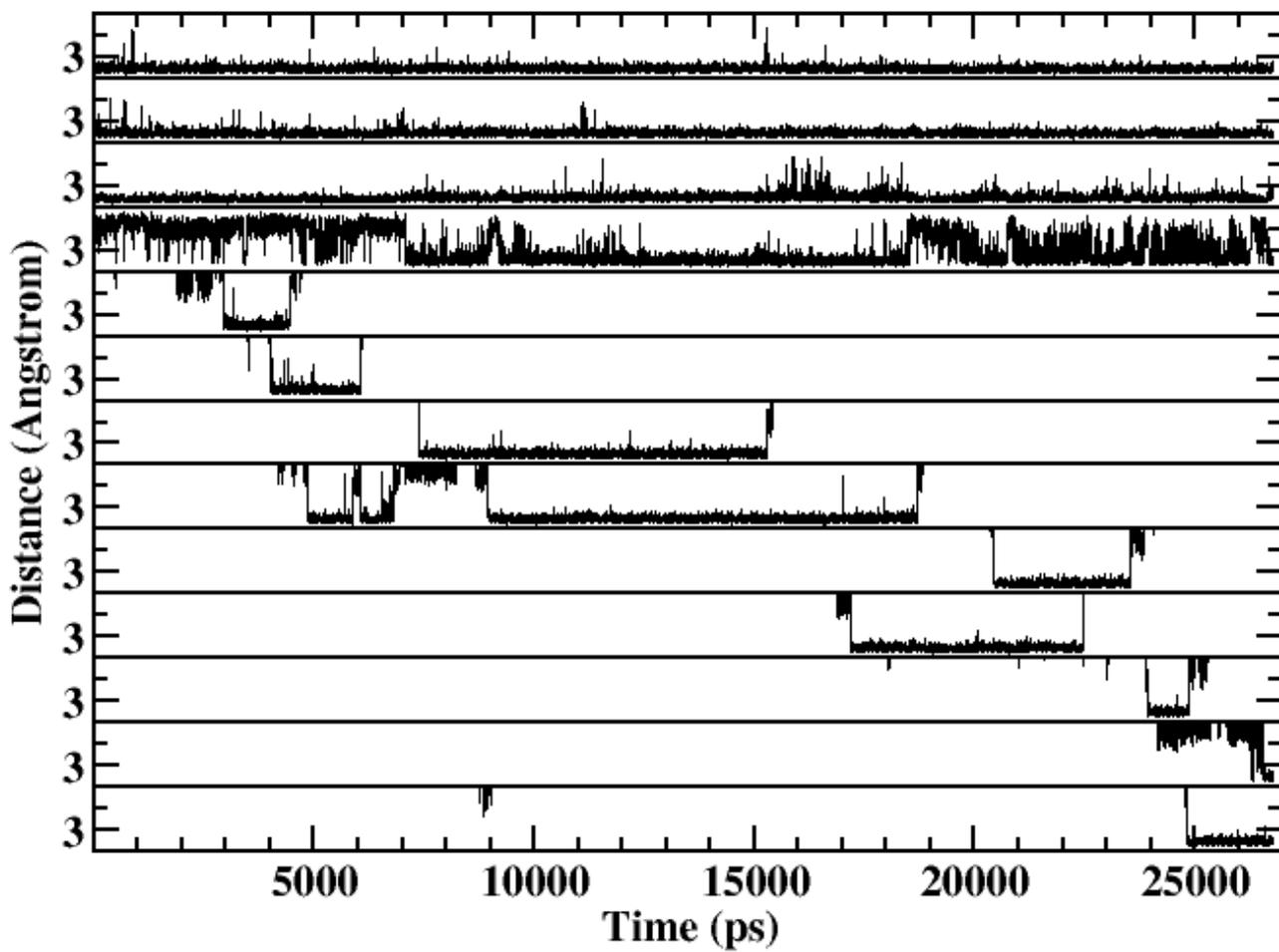


Fig 4.38: Inter-atomic distance plots between calcium 1 and its oxygen ligands from the protein and water molecules in setup 1. Row 1 is OE1-62, row 2 is OE2-62, row 3 is OD2-334, row 4 is OD1-334, row 5 is WAT 78207, row 6 is WAT 91242, row 7 is WAT 70413, row 8 is WAT 76149, row 9 is WAT 75360, row 10 is WAT 93156, row 11 is WAT 84846, row 12 is WAT 70803 and row 13 is WAT 82752.

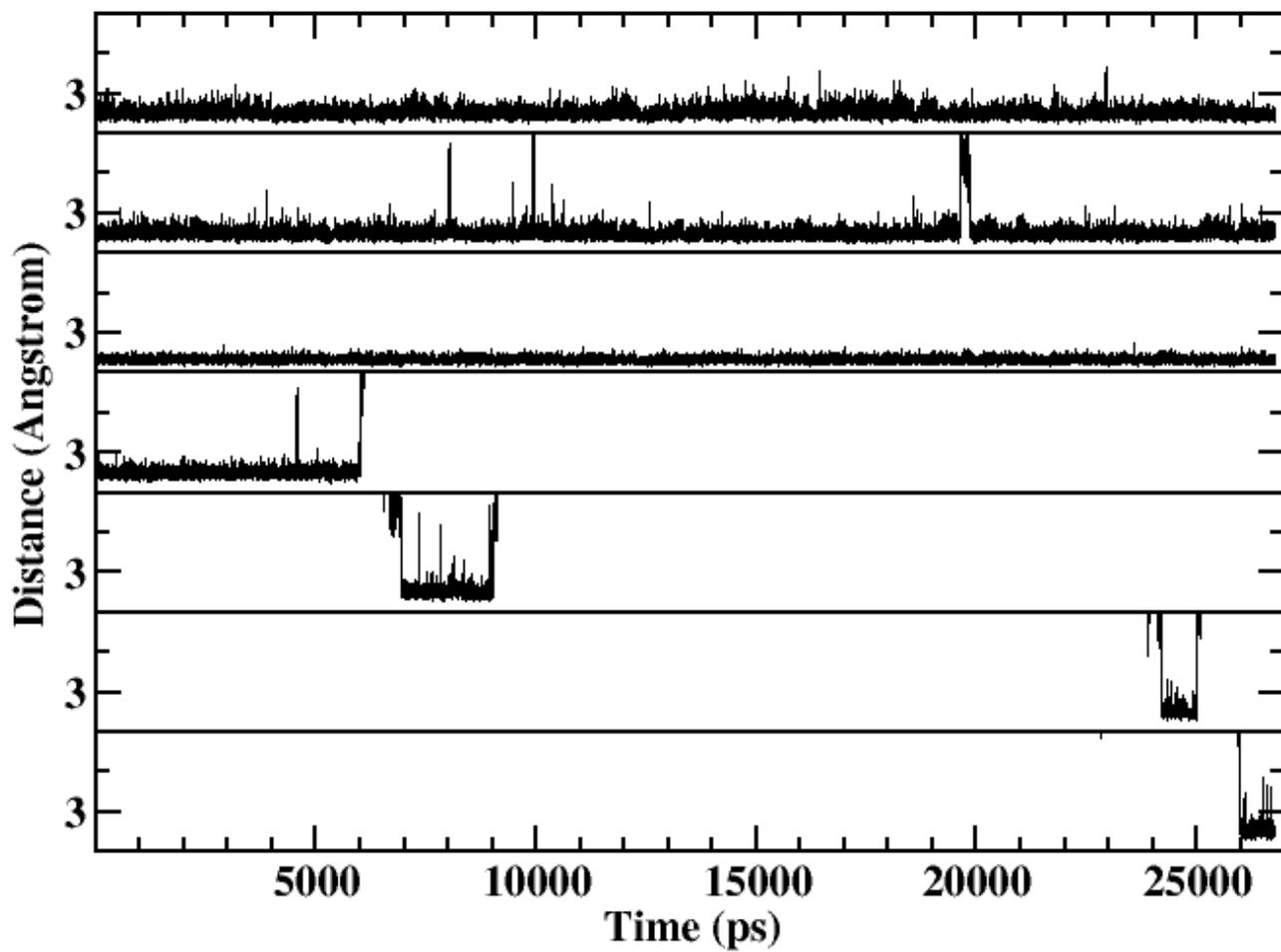


Fig 4.39: Inter-atomic distance plots between calcium 2 and its oxygen ligands from the protein and water molecules in setup 1. Row 1 is K-97, row 2 is L-100, row 3 is OE2-105, row 4 is WAT 89787, row 5 is WAT 79059, row 6 is WAT 88977 and row 7 is WAT 93267.

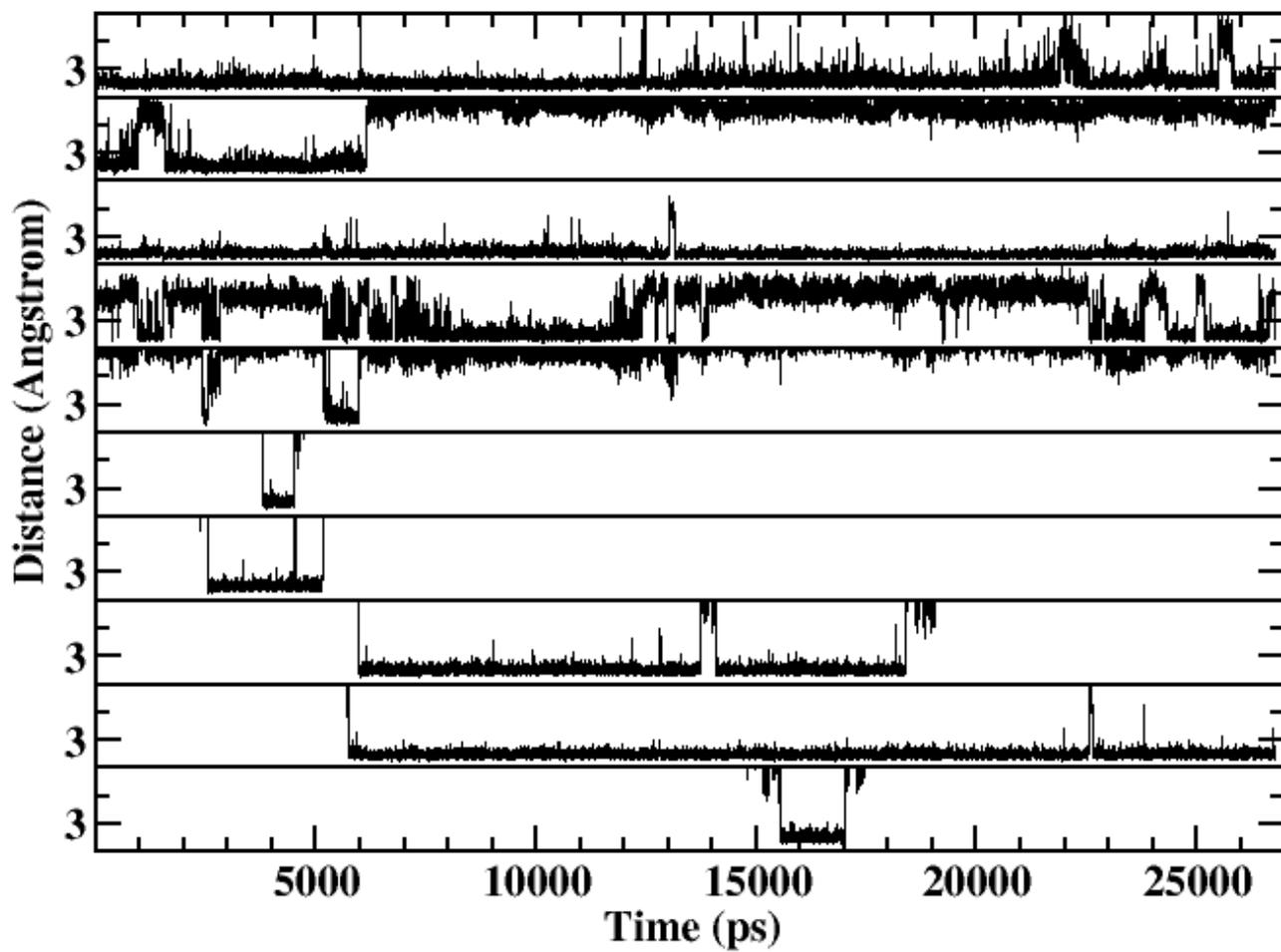


Fig 4.40: Inter-atomic distance plots between calcium 3 and its oxygen ligands from the protein and water molecules in setup 1. Row 1 is M-127, row 2 is G-129, row 3 is OD1-171, row 4 is OD2-171, row 5 is G-131, row 6 is WAT 96807, row 7 is WAT 98733, row 8 is WAT 76650, row 9 is WAT 87936 and row 10 is WAT 93240.

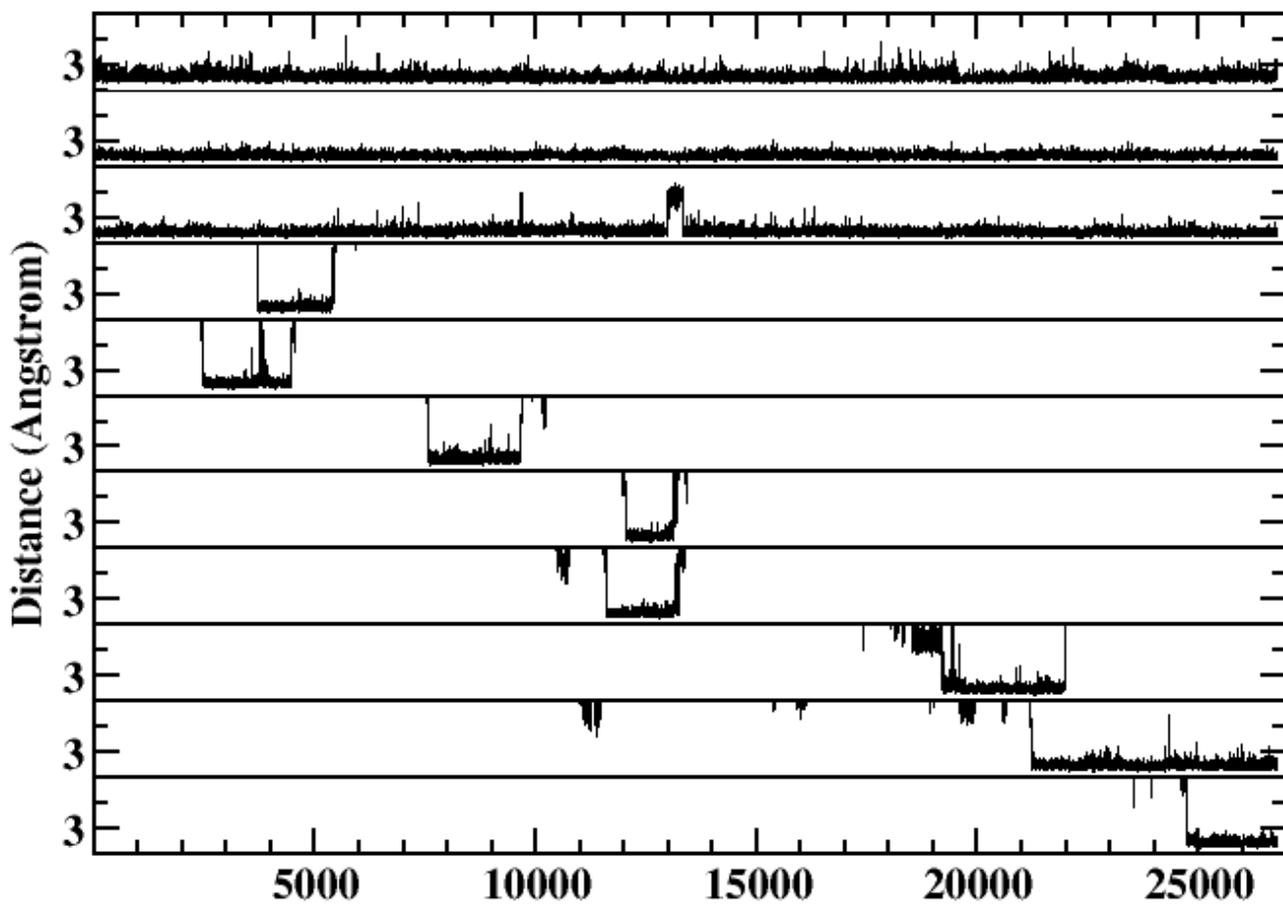


Fig 4.41: Inter-atomic distance plots between calcium 4 and its oxygen ligands from the protein and water molecules in setup 1. Row 1 is T-132, row 2 is OE1-134, row 3 is OE2-134, row 4 is WAT 84567, row 5 is WAT93339, row 6 is WAT86652, row 7 is WAT77907, row 8 is WAT98988, row 9 is WAT780192, row 10 is WAT70737 and row 11 is WAT87141.

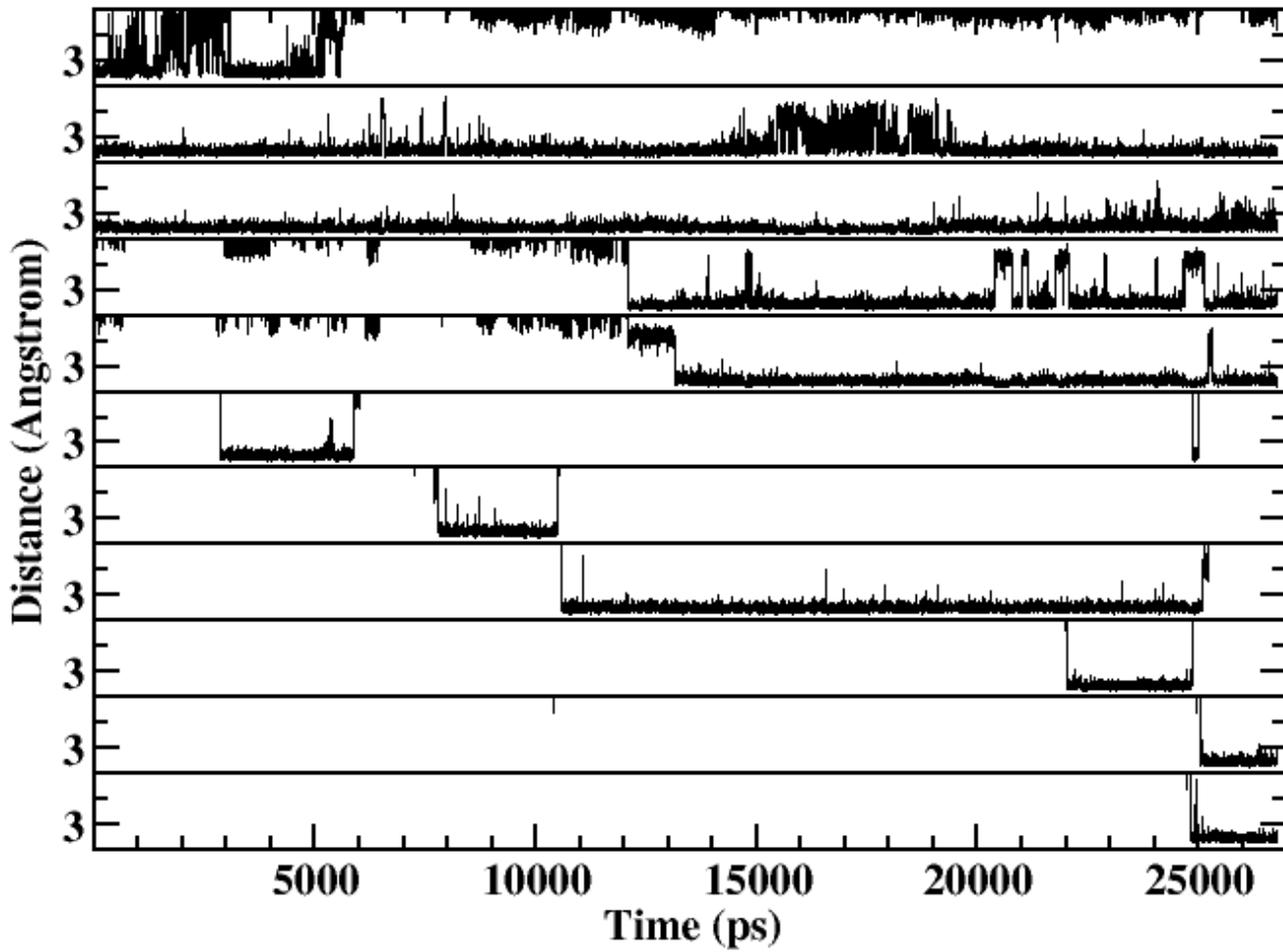


Fig 4.42: Inter-atomic distance plots between calcium 5 and its oxygen ligands from the protein and water molecules in setup 1. Row 1 is G-210, row 2 is OE1-225, row 3 is OE2-225, row 4 is OE2-211, row 5 is OE1-211, row 6 is WAT 87555, row 7 is WAT 68457, row 8 is WAT 67467, row 9 is WAT 82200, row 10 is WAT 72945 and row 11 is WAT 77917.

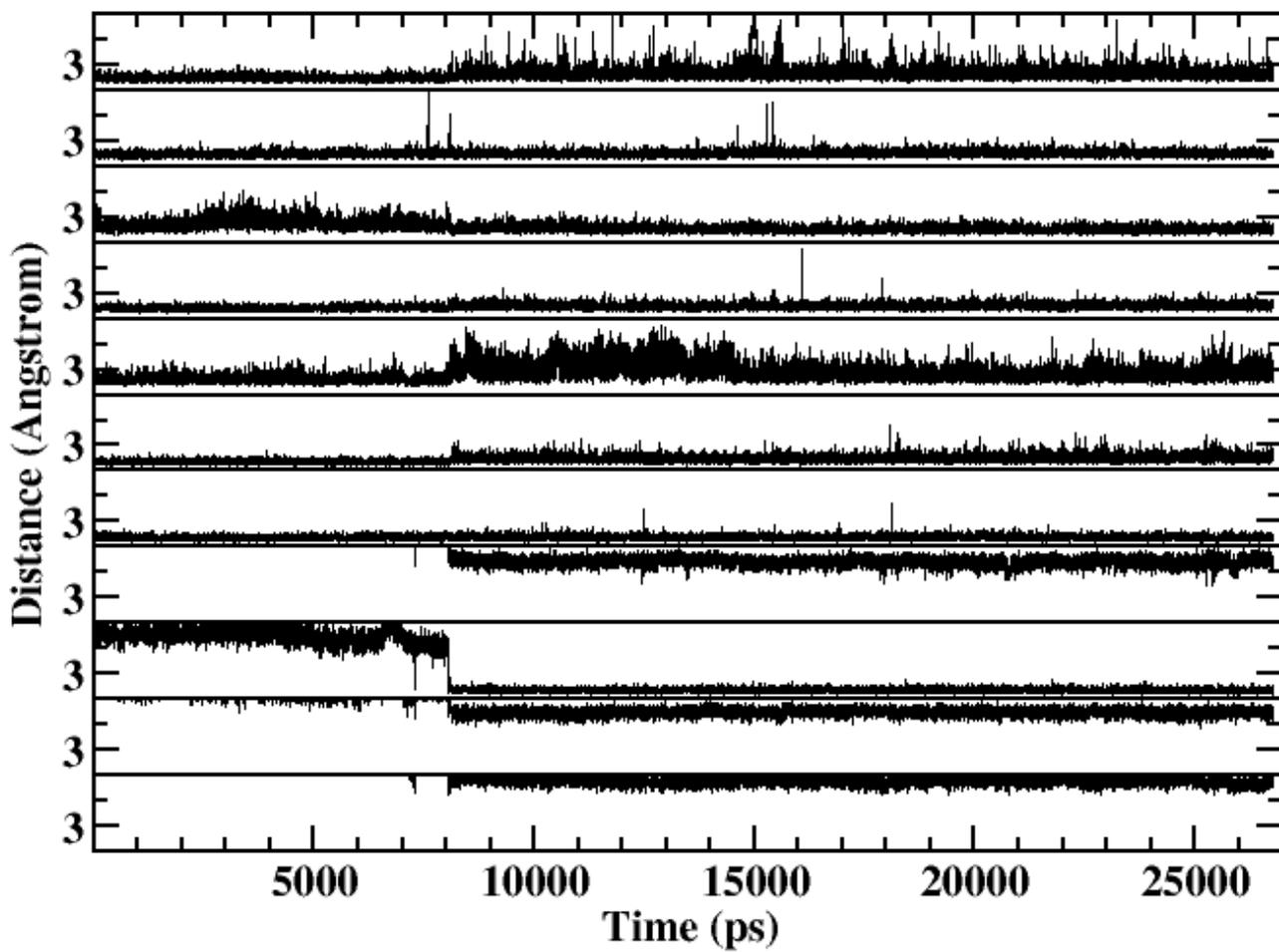


Fig 4.43: Inter-atomic distance plots between calcium 6 and its oxygen ligands from the protein and membrane molecules in setup 1. Row 1 is A-286, row 2 is M-286, row 3 is G-288, row 4 is G-290, row 5 is T-291, row 6 is OE1-330, row 7 is OE2-330, row 8 is O2-DOPG124, row 9 is O3-DOPG124, row 10 is O4-DOPG124 and row 11 is O5-DOPG124.

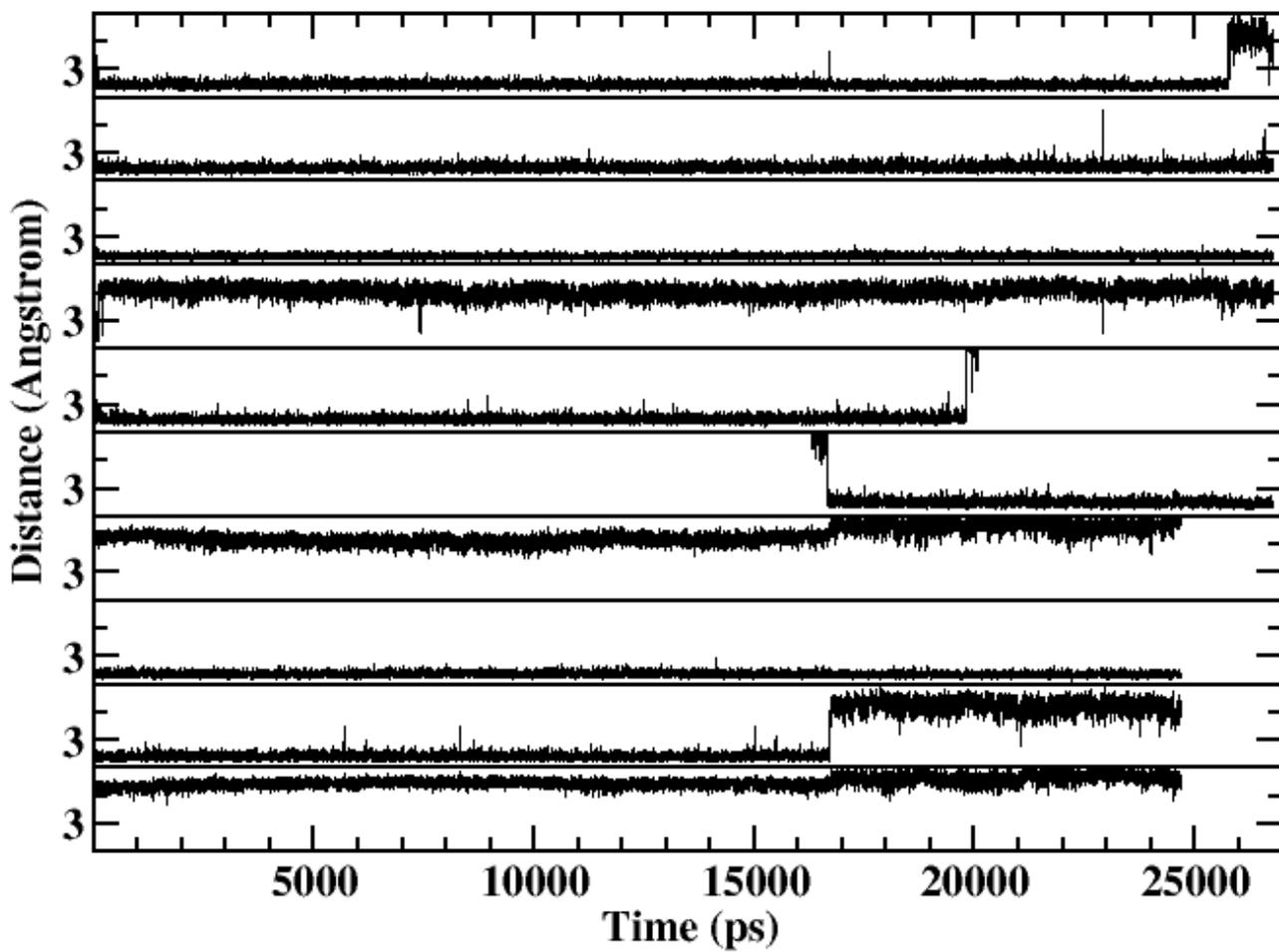


Fig 4.44: Inter-atomic distance plots between calcium 7 and its oxygen ligands from the protein, membrane and water molecules in setup 1. Row 1 is L-328, row 2 is T-331, row 3 is OE1-336, row 4 is OE2-336, row 5 is WAT 77856, row 6 is WAT 81036, row 7 is O-DOPC118, row 8 is O1-DOPC118, row 9 is O2-DOPC118 and row 10 is O3-DOPC118.

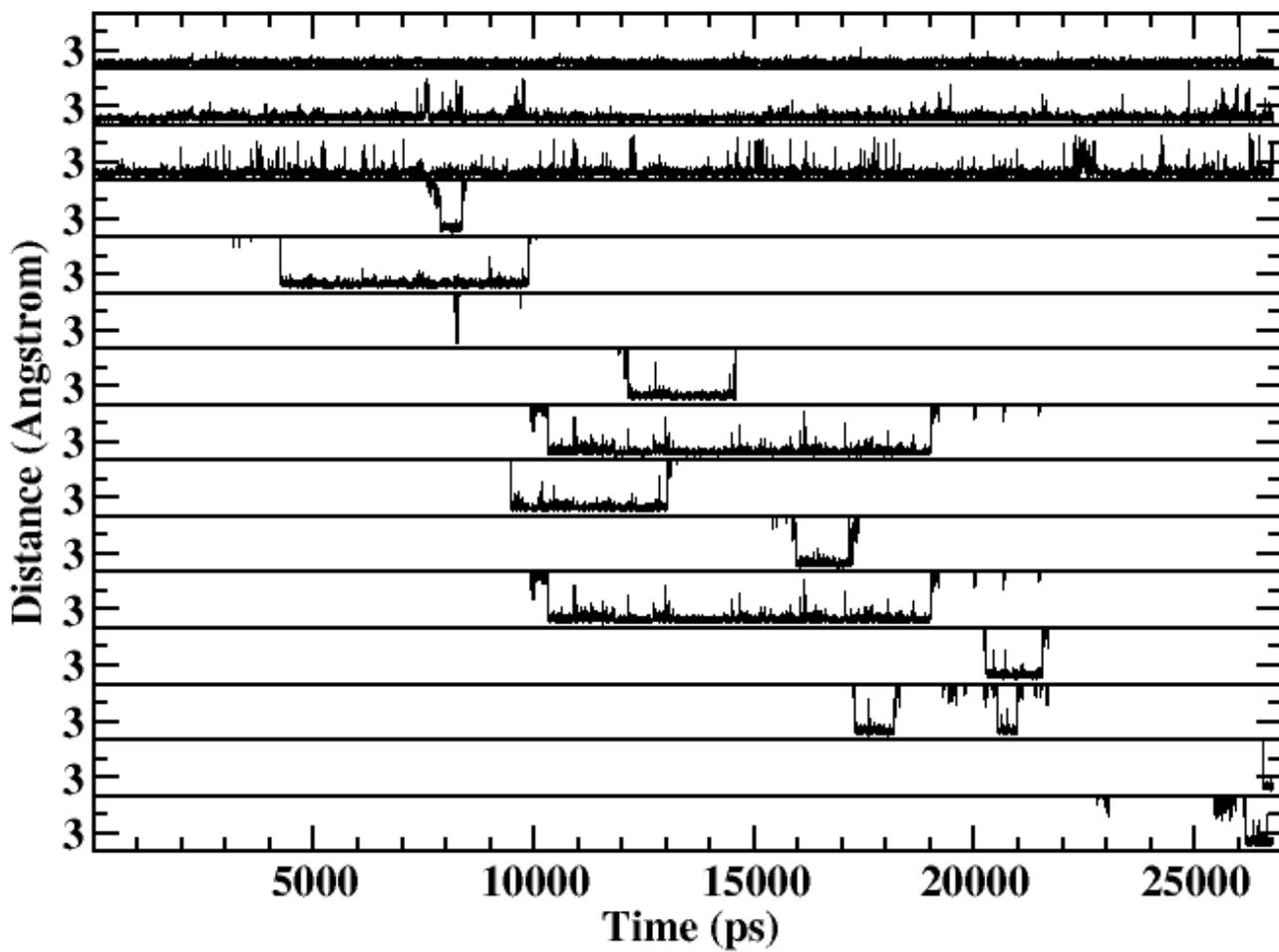


Fig 4.45: Inter-atomic distance plots between calcium 8 and its oxygen ligands from the protein and water molecules in setup 1. Row 1 is OD1-253, row 2 is OE1-261, row 3 is OE2-261, row 4 is WAT 69636, row 5 is WAT 89070, row 6 is WAT 95700, row 7 is 74769, row 8 is WAT 98490, row 9 is WAT 98697, row 10 is WAT 91281, row 11 is WAT 98490, row 12 is WAT 69300, row 13 is WAT 80868, row 14 is WAT 73047 and row 15 is WAT 87525.

4.12: Observations Made in Inter Atomic Distance Plots

Several ligand exchanges were apparent after plotting the distance changes between calcium ions and their oxygen ligands, and one calcium ion was observed to coordinate with a phospholipid over the course of setup 1. The most significant observation in setup 1 was calcium 6 becoming coordinated with O3 of DOPG124 at 8049 ps, O3 being one of the oxygen atoms with a partial negative charge on the phosphate group. At the same moment in the simulation, the hydroxyl oxygen atom on T-291 leaves the coordination sphere of calcium 6. Calcium 7 in setup 1 was observed to have coordinated with O2 of DOPC118, also an oxygen atom with a partial negative charge on the phosphate group, to rotate outside of the coordination sphere at 16717 ps, while the other oxygen atom stayed within the coordination sphere, and an oxygen from water 81036 entered the coordination sphere at 16660 ps, a difference of 57 ps for the ligand exchange. In all eight calcium ion plots, water ligands enter and exit the coordination sphere so that calcium's coordination number is kept within the favorable range of 6-8.

4.13: Visualization of MD Simulation Trajectories

The following cartoons display the conformation changes of annexin I over regular intervals. The cartoons colored green are from setup 1 and are shown in Figure 4.46 and the cartoons colored cyan are from setup 2 and are shown in Figure 4.47, and the light-blue spheres represent the calcium ions.

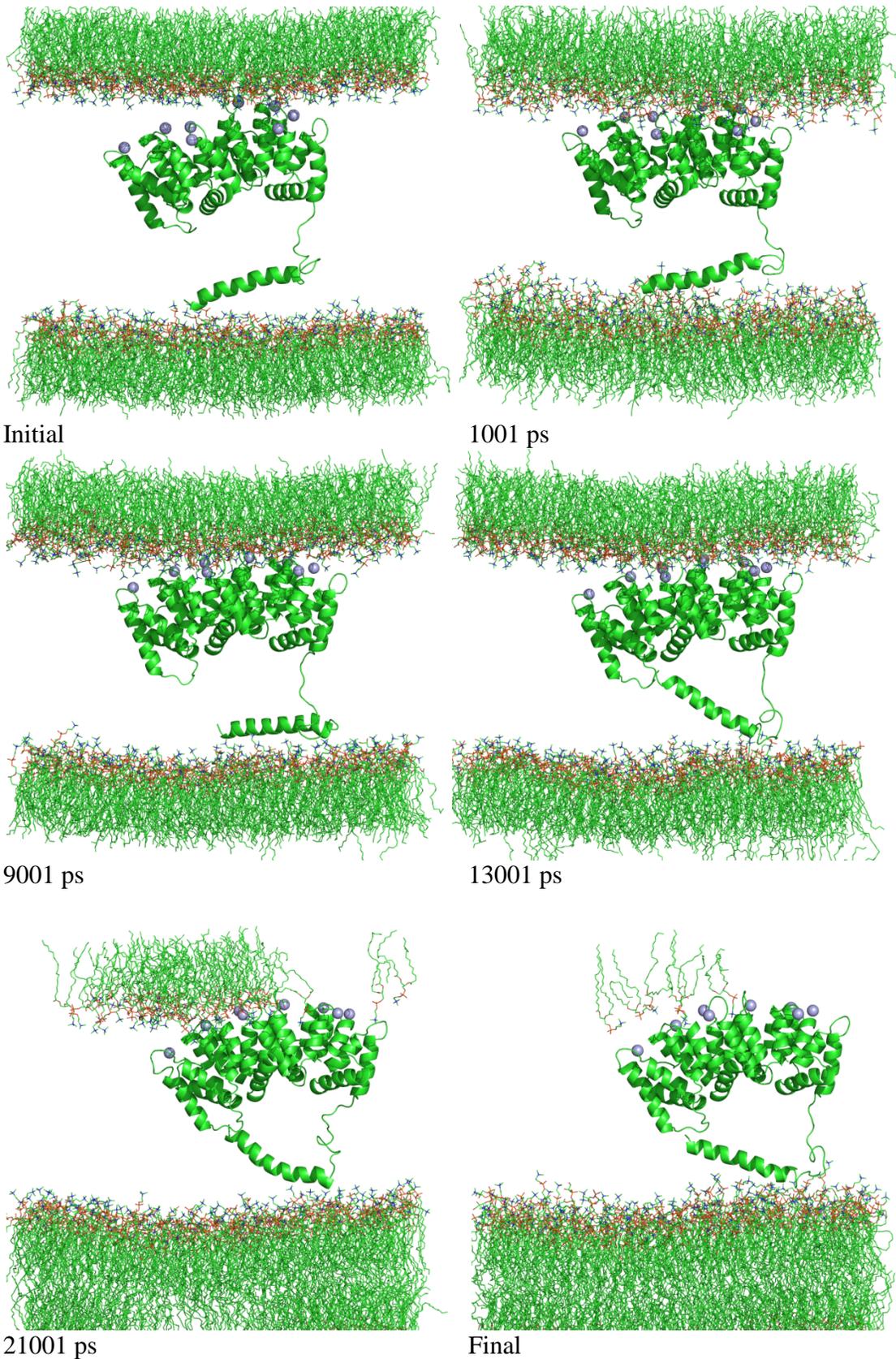


Fig 4.46: Snapshots taken from the trajectory of setup 1 over regular intervals.

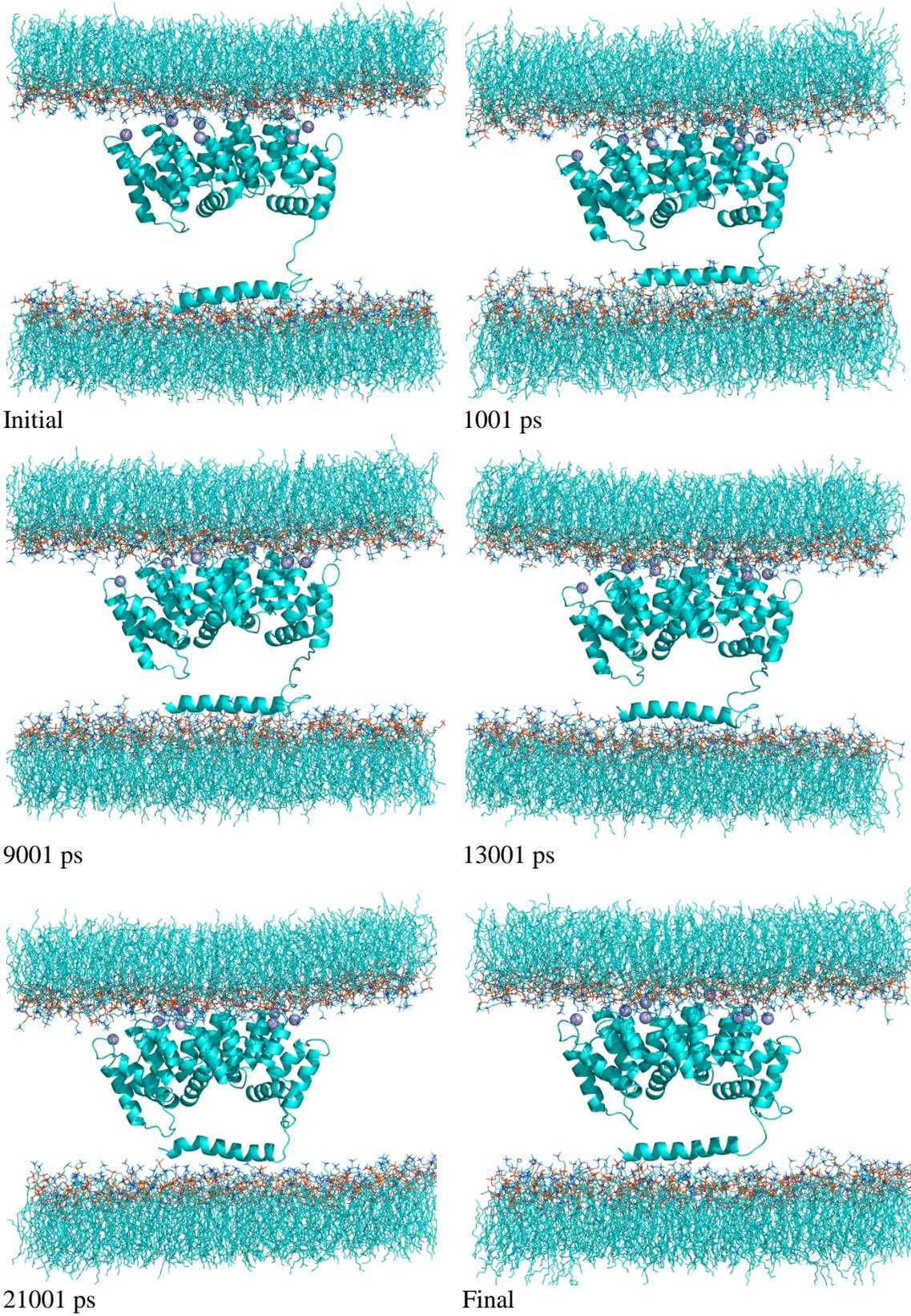


Fig 4.47: Snapshots taken from the trajectory of setup 2 over regular interval

4.14: Observations Made from the Visualization of MD Simulation Trajectories

The above cartoons help to strengthen the statements made about annexin I in the previous sections. Specifically, for setup 1, snapshots taken at 10001 and 10245 clearly display the kink in the second turn of the N-terminal helix, as it is oriented perpendicular to the phospholipid surface. Furthermore, snapshots 11412-13328 display the rise of the N-terminus to a 45 degree angle with respect to the phospholipid surface during setup 1. It is also observed that annexin I does not undergo much conformational change in setup 2, save for the unstructured coil region. The N-terminus does retain its secondary structure throughout both simulations.

CHAPTER 5 - MUTATION STUDIES

5.1: Background

Francoise Porte *et al.* reported results of mutation studies on the N-terminus of human annexin I in 1996.⁸ It has been postulated by researchers that phosphorylation may modulate annexin I functions, e.g., its action on membrane aggregation. Wang and Creutz demonstrated that a S27D mutant of human annexin I increases the calcium concentration required to aggregate chromaffin granules.³⁹ Human annexin I is phosphorylated to equal extent by protein kinase C on Thr24, Ser27 and Ser28. The researchers investigated the effects of mutating Ser27 to glutamic acid, thereby mimicking phosphorylation of Ser27, by analyzing the extent of liposome aggregation, liposome binding and self-association. Two mutants were prepared in this study: the S27E mutant and a control mutant, S27A, which prevents phosphorylation. Wild-type annexin I was also studied.

Wild-type annexin I and the S27A mutant displayed the same calcium dependence for phospholipid vesicle aggregation, while the S27E mutant showed a higher calcium concentration requirement and lower maximal extent of aggregation. In contrast, the wild-type and two mutants all required identical calcium concentrations for liposome binding, the first event in membrane association, and self-association to its dimer formation. These results suggest that modifications in the conformation of the N-terminus alters properties of the protein and that phosphorylation modulates the functions of the protein.

Furthermore, the researchers found that the S27E mutant had an increased sensitivity to trypsin proteolysis. The protein was cleaved to a 32 kDa fragment, most likely corresponding to truncation in the N-terminus at residue Lys26. The researchers suggested that the addition of a negative charge in the N-terminal tail may induce conformational changes and flexibility in that region.

5.2: Methods

Molecular dynamics simulations were implemented to study and compare the conformational changes of wild-type porcine annexin I and the S28E and S28A mutant proteins. The human and porcine annexin I proteins share an 80% homology in the N-terminal region, however they do differ at residue 27. The human primary sequence for residues 26-29 is: K-S-S-K. The porcine primary sequence for residues 26-29 is: K-G-S-K. Therefore, for my study, Ser28 was mutated rather than Ser27.

The biopolymer module in Insight II was used to create the site-specific mutations on porcine annexin I. The proteins shared an identical spatial conformation, apart from the site-specific mutation, in that calcium was bound to the core domains and the

N-terminus was in an exposed position. Six simulations were performed: the wild-type, S28A and S28E full length 341 residue proteins; and the wild-type, S28A and S28E peptides, containing only the first 42 residues. All systems were neutralized with chloride counter-ions and explicitly solvated in a TIP3P water box using the xleap program in AMBER 9.

All computations were performed on a 128-processor SGI Altix. All calculations were carried out using AMBER's force field (ff03) in the AMBER 9 software package. The SANDER module carried out the energy minimizations and molecular dynamics computations. Periodic boundary conditions were applied to the system with a non-bonded cutoff of 12 Å to truncate VDW interactions. The particle-mesh Ewald method was used to treat long range electrostatic interactions, with a cubic B-spline interpolation. The SHAKE algorithm was used to constrain all bonds involving hydrogen atoms with a geometric tolerance of 0.00001 Å.

The system was energy minimized twice before the MD simulation. A restrained minimization was first performed on the solvent and the counter ions while keeping the protein fixed. Next, the entire system was minimized. The minimized system was then warmed for 20 ps to 300 K from an initial temperature of 10 K. Constant pressure dynamics were then performed on the warmed system for 100 ps. A time step of 2 fs was used for the molecular dynamics simulations. Finally, constant volume dynamics (NVT) was performed on the system. Each full-length protein simulation is summarized in Table 5.1, and each peptide simulation is summarized in Table 5.2.

Table 5.1: Dimensions of each system set-up, including total number of atoms used per full-length simulation and timescale of each setup.

	Wild-type	S28A mutant	S28E mutant
Box dimension (Å)	111.689 x 104.558 x 82.048	111.689 x 104.558 x 82.048	111.689 x 104.558 x 82.048
Number of atoms	80159	80172	80160
Simulation time (ps)	9605.0	9386.8	9607.2

Table 5.2: Dimensions of each system set-up, including total number of atoms used per peptide simulation and timescale of each setup.

	Wild-type	S28A mutant	S28E mutant
Box dimension (Å)	75.567 x 71.710 x 58.211	75.567 x 71.710 x 58.211	75.567 x 71.710 x 58.211
Number of atoms	28900	28902	28905
Simulation time (ps)	14956.6	15015.2	15007.2

5.3: Results

RMSD and calculations were performed on the α carbon atoms of the full-length protein and also the core domains from the trajectories of the three full-length systems. RMSF calculations were performed for each α carbon atom from each of the three systems. The following Figures 5.1-5.8 depict the results of those calculations.

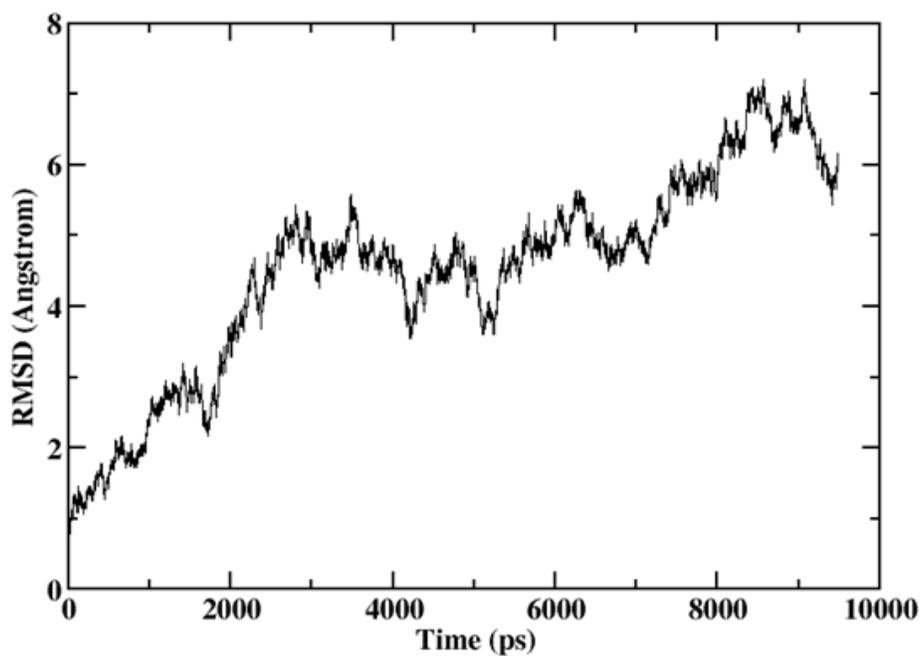


Fig 5.1: RMSD plot of alpha carbon atoms for the full-length wild-type protein.

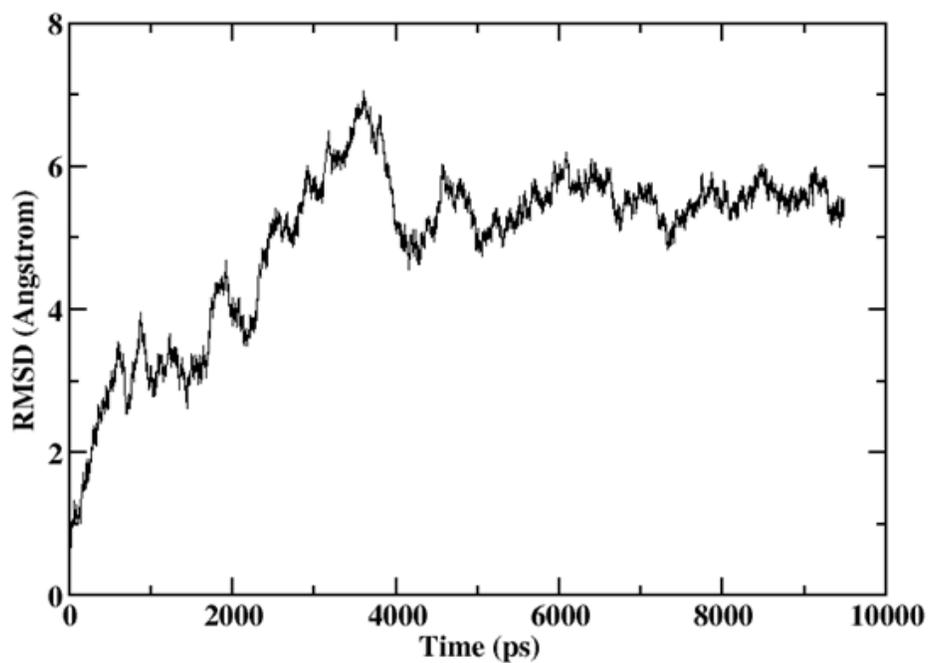


Fig 5.2: RMSD plot of alpha carbon atoms for the full-length S28A mutant protein.

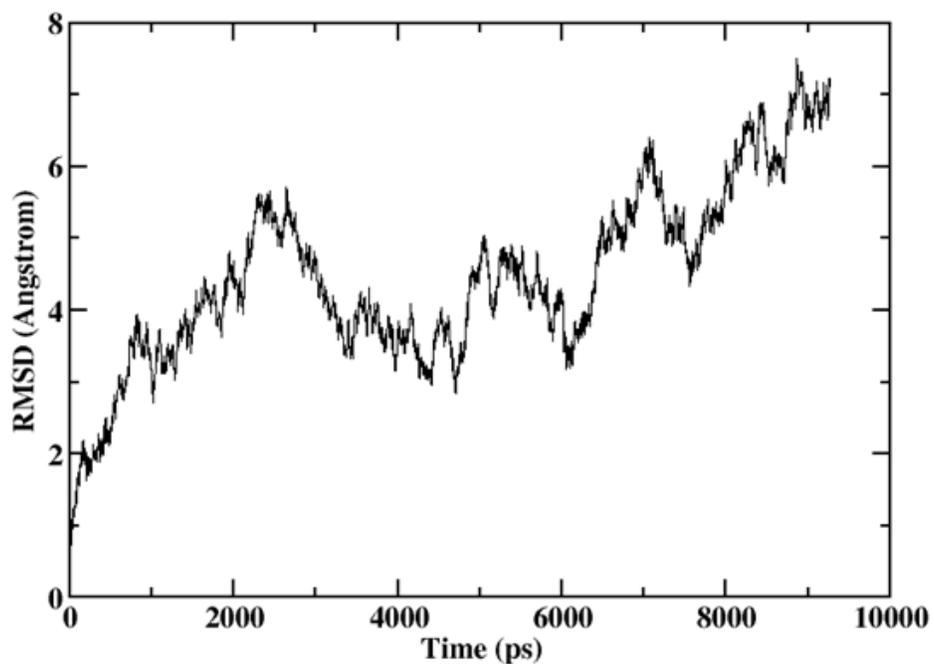


Fig 5.3: RMSD plot of alpha carbon atoms for the full-length S28E mutant protein.

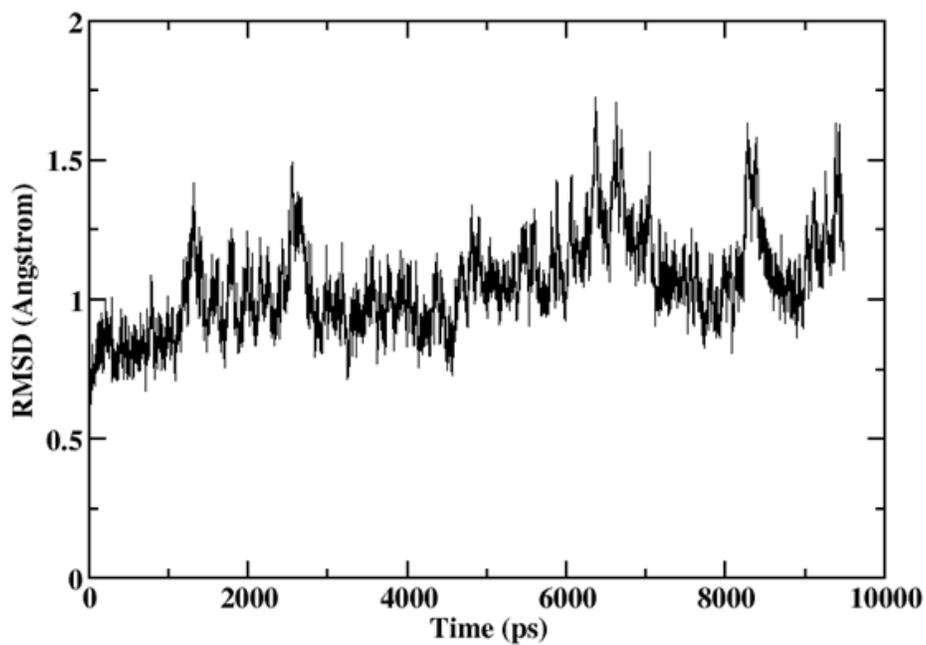


Fig 5.4: RMSD plot of alpha carbon atoms for the core domain of the wild-type protein.

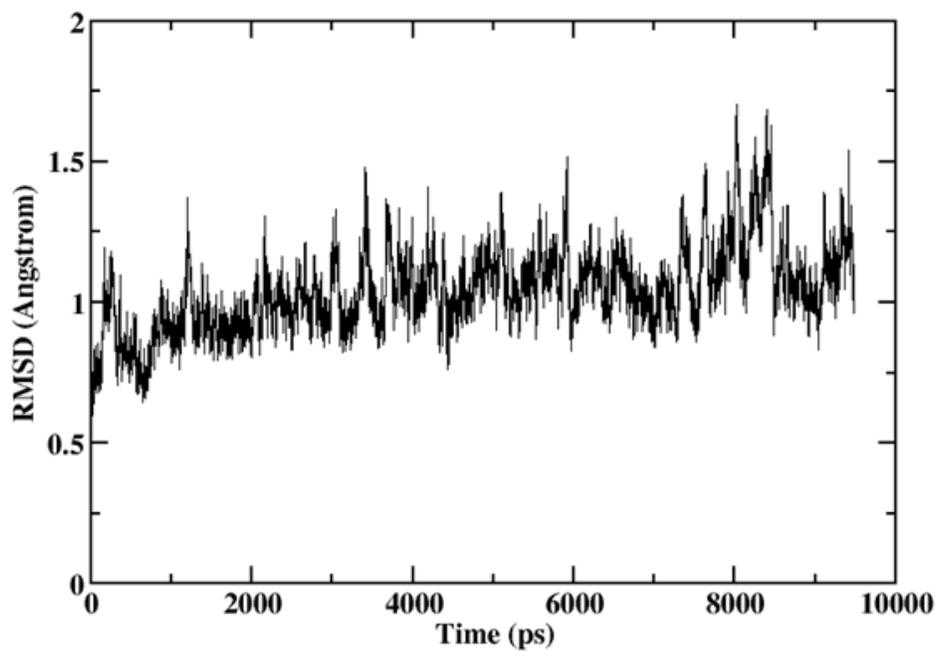


Fig 5.5: RMSD plot of alpha carbon atoms for the core domain of the S28A mutant protein.

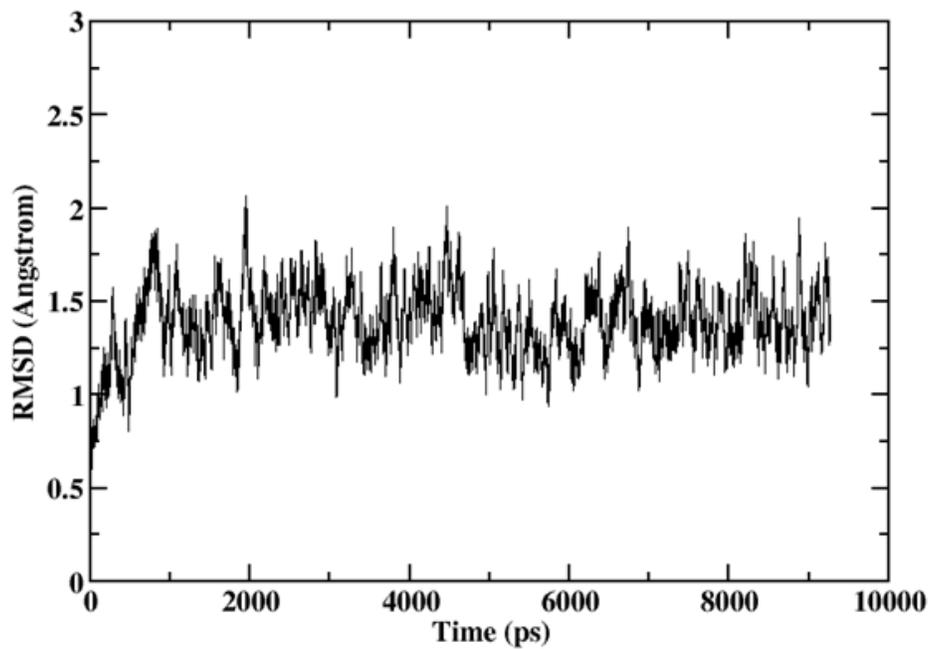


Fig 5.6: RMSD plot of alpha carbon atoms for the core domain of the S28E mutant protein.

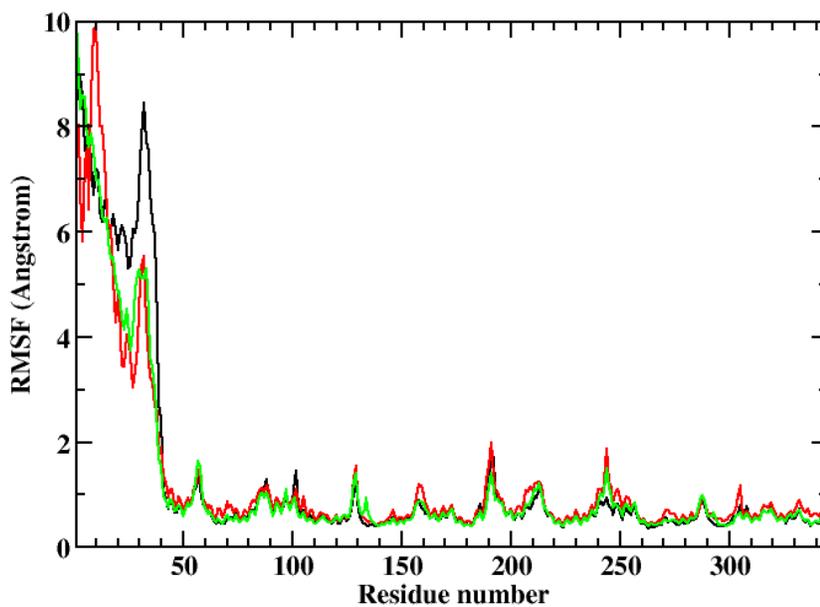


Fig 5.7: RMSF overlay plot of alpha carbon atoms per residue in the wild-type and both mutant proteins. The wild-type plot is colored black, the **S28A** plot is colored red and the **S28E** plot is colored green.

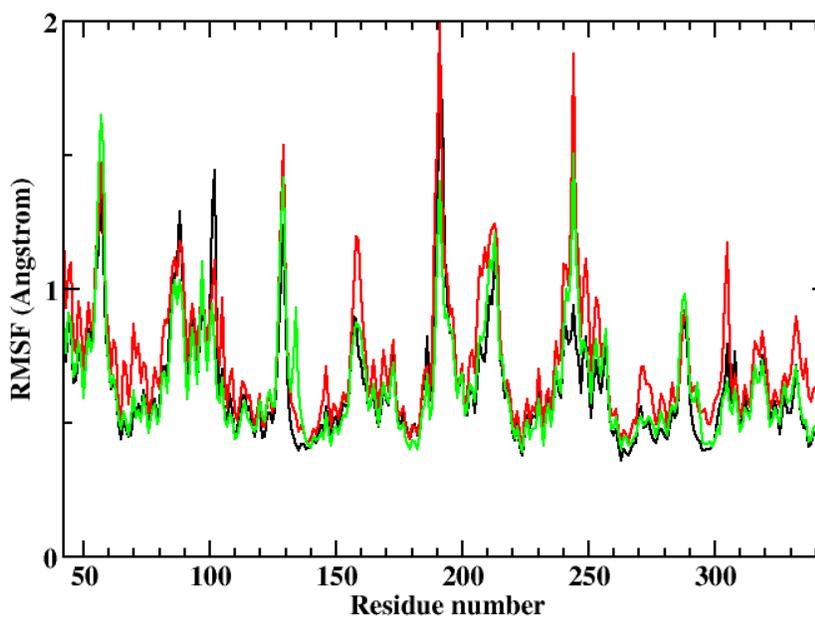
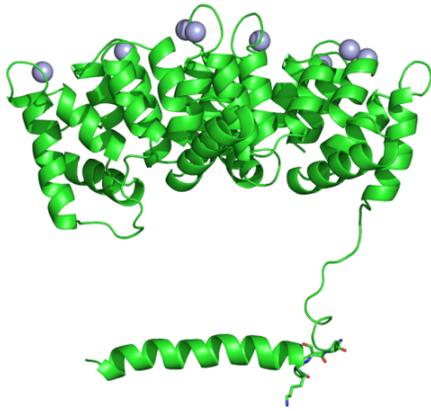


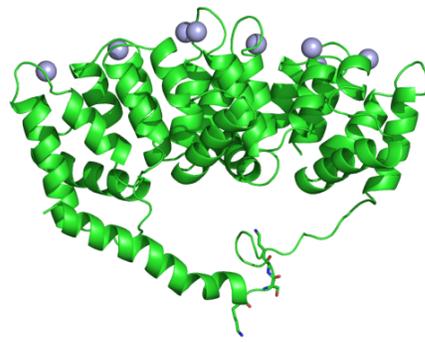
Fig 5.8: RMSF overlay plot of alpha carbon atoms per residue in the wild-type and both mutant full-length proteins. Enlarged view of the core domain residues. The wild-type plot is colored black, the **S28A** plot is colored red and the **S28E** plot is colored green.

5.4: Visualization of MD Simulation Trajectories

The following cartoons display the conformation changes of annexin I from the initial and final snapshots of the full-length and peptide simulations. Figure 5.9 displays the wild-type and mutant full-length protein's initial and final snapshot from their trajectories. Figure 5.10 displays the wild-type and mutant peptide's initial and final snapshot from their trajectories. The cartoons are color-coded, so that the green cartoon represents the wild-type protein, the cyan cartoon represents the S28A mutant, and the magenta cartoon represents the S28E mutant.



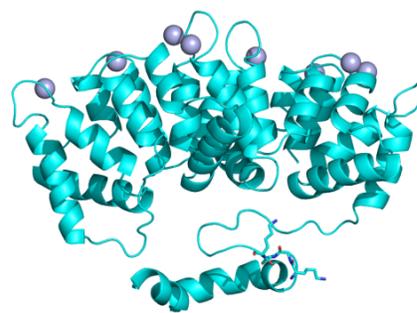
Wild-type Initial Conformation



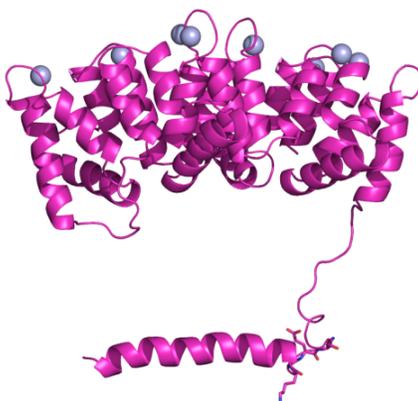
Wild-type Final Conformation



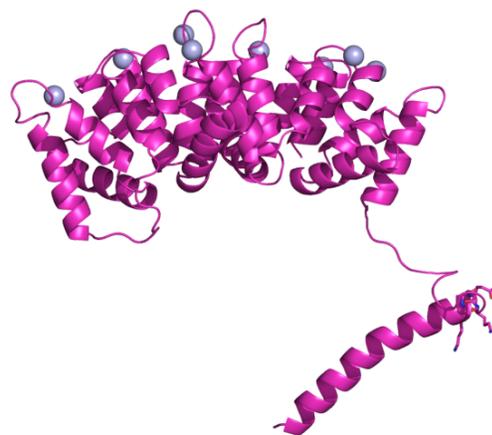
S28A Mutant Initial Conformation



S28A Mutant Final Conformation

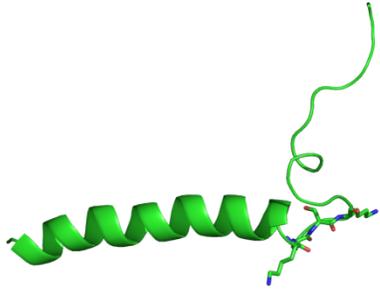


S28E Mutant Initial Conformation

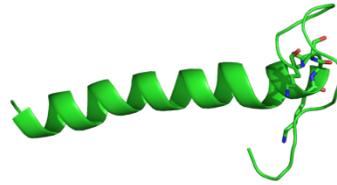


S28E Mutant Final Conformation

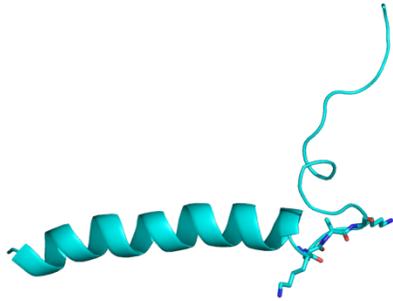
Fig 5.9: Initial and final snapshots from the full-length protein simulations.



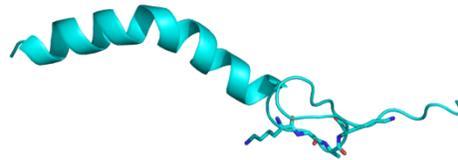
Wild-type Peptide Initial Conformation



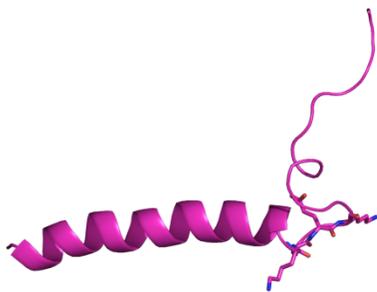
Wild-type Peptide Final Conformation



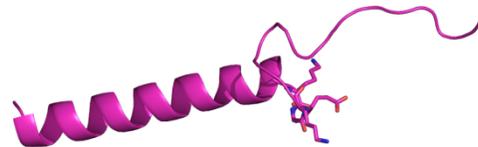
S28A Mutant Peptide Initial Conformation



S28A Mutant Peptide Final Conformation



S28E Mutant Peptide Initial Conformation



S28E Mutant Peptide Final Conformation

Fig 5.10: Initial and final snapshots from the full-length protein simulations.

5.5: Implications of Mutation Studies

The most significant difference between the three protein's values on the RMSF plot is located in the N-terminal region, residues 2-41. The wild-type protein displays the highest values for residues 26-29. Also, there are slight variations in the core domain residues between the three proteins. Although statistically insignificant, the S28E mutant has ever so slightly lower values at the protein ligand residues of calcium's 3, 4, 5 and 8, which are located in domains II and III, respectively.

From the snapshots of the initial and final conformations of the proteins, we can see that there are obvious similarities between the wild-type protein and the S28A mutant. In their final conformation, the N-termini of both proteins are in closer proximity with the core domain than it was initially. By contrast, the N-terminus of the S28E mutant has been pushed further away from the core domain. Also visible from the snapshots, the N-termini of the wild-type protein and S28A mutant are beginning to lose their secondary structure in their final conformation. The S28A mutant contains an unstructured region between residues 8-10, whereas residues 19-22 on the wild-type protein appear to be unwinding from its initial helical shape. To re-enforce the visualization of the loss of helical structure between residues 8-10 of the S28A mutant, the RMSF plot of the N-terminus (Figure 5.8) clearly shows residues 8-13 having significantly higher values than the other two proteins. The S28E mutant, on the other hand, does not lose its helical structure in the N-terminal domain at all.

The snapshots from the peptide trajectories are consistent with the snapshots from the full-length protein trajectories. The S28A mutant peptide has a kink in its helix at residue 12, the same area where the full-length mutant was losing its secondary structure.

The snapshots of the mutant proteins are also in agreement with the results obtained from the MD simulations containing the phospholipid layers and the results from Porte *et al.* In setup 1 of my simulations, the N-terminus eventually oriented its first few residues of the alpha helix in very close proximity with repeat III of the core domain, while K-26 and K-29 acted as an anchor to the phospholipid layer at the other end of the helix. In the wild-type and S28A protein simulations, the N-termini also oriented their helices to within close proximity of repeat III. The S28E mutant did not orient its N-terminus towards repeat III, in fact, it moved in the opposite direction. Porte *et al.* reported that both the wild-type and S27A protein aggregated phospholipid vesicles to an equal extent, and the S27E protein required a much higher calcium concentration for aggregation. Therefore, based on the results of the simulations, it would seem that phosphorylation of either S27 or S28 stabilizes the structure of the N-terminus by electrostatically neutralizing the side-chains of either K26 or K29 and thereby interfering with the N-terminus interacting with the core domain. A more stable protein would not be as energetically driven to bind to a second membrane than a less stable protein would. It is clear that phosphorylation induces conformational changes in annexin I, and in turn those conformational changes result in altered protein functions.

Further analysis needs to be done on the simulations of annexin mutants. The inter-atomic distance between the mutant S28E and both lysines 26 and 29 needs to be

calculated to verify if the residues do move within closer proximity of one another driven by an electrostatic attraction.

CHAPTER 6 - DISCUSSION

Annexin I is a calcium-dependent phospholipid binding protein and is active in membrane aggregation and fusion processes. Members of the annexin family share a conserved core domain comprised of four homologous repeats, and also a variable N-terminal domain. The precise molecular mechanism of membrane aggregation remains to be elucidated, but the importance of the N-terminal domain of annexin I in this process has been described in biochemical studies. The purpose of this study was to investigate the conformational changes that occur in calcium bound annexin I when in close proximity with negatively charged phospholipid layers by performing MD simulation techniques on two 'setups', differing in the initial distance between the centers-of-mass of the lipid headgroups. Setup 1, which had the greater distance between the phospholipid layers, exhibited not only a definite calcium induced binding to the top phospholipid layer through a ligand exchange process, but also a substantial conformational change of the N-terminus which may provide a mechanism for a second membrane binding event.

6.1: Residue Flexibility

Anja Rosengarth *et al.* proposed that calcium binding to the convex face of the core domain of annexin I triggers a series of events in which the N-terminus is ultimately ejected from repeat III of the core domain and the D helix folds back into the proper helical conformation. In this proposed active conformation, the N-terminus would be free to move around *via* the flexible linker formed by residues 27-41. Electron density

studies indicated that hydrophobic residues of the N-terminus (Met3, Val4, Phe7) would favorably be packed into a hydrophobic pocket formed by residues Phe221, Leu225, Phe237 and Val268 of repeat III. This idea is supported by our analysis of the average RMSF values, which are highest in the core domain at repeat III, giving further evidence to the structural role played by this region of the protein during membrane binding.

Furthermore, several calcium coordinating residues displayed relatively high magnitudes of fluctuation compared with non-coordinating residues as shown on the RMSF Figures. Anja Rosengarth *et al.* reported the calcium coordinating residues of annexin I. The results of a study conducted by D. Cregut *et al.*, in which molecular dynamics simulations were performed on annexin V, also found that calcium coordinating residues displayed an increased flexibility.

6.2: Anchor Residues

According to Eduard Bitto and Wonhwa Cho, K-26 and K-29 play an essential role in the membrane aggregation activity of annexin I.¹⁴ Bitto and Cho systematically assessed the contribution of the amino terminus to membrane aggregation by first truncating the entire N-terminus and then measuring the effect of incremental addition of amino terminal residues on vesicle aggregation activity. They reported that annexin I Δ^{1-41} showed no detectable aggregating activity under normal assay conditions, that Δ^{1-29} lowered the activity of the core, whereas Δ^{1-24} fully restored the wild-type activity. The analysis of MM-PBSA data from simulation trajectories confirms their statement. Throughout both simulations, those two lysine residues displayed the strongest electrostatic attractions particularly to the phospholipid layer. The positively charged side-chains of these residues, especially K-26, were observed to orient themselves in

close proximity (~ 3.5 Å) with the polar headgroups of the phospholipids, as shown in Figure 6.1. The side-chains of these residues appeared to act as an electrostatic anchor to the surface of the phospholipids throughout the simulations. K-26 and K-29 are absolutely conserved to all annexins I from different species, so this mechanism would apply to all annexins I. In the present study, K-26 and K-29 both orient their positively charged side chains down towards the negatively charged phospholipid layer, possibly acting as an electrostatic anchor so that the protein may adopt a more favorable conformation in its environment. K-26 consistently displayed the highest electrostatic attraction to the bottom phospholipid layer throughout both simulations, reaching a maximum energy of -248.51 kcal/mol between 14540-15575 ps. Over the course of setup 1, the RMSD values of the full-length protein stabilized at around 17ns, and this was attributed to the first N-terminal helix reaching an equilibrium distance to repeat III of the core domain. From the final snapshot of the setup 1 trajectory, K-29 is located at the vertex of the angle between the helical part of the N-terminus and the disordered part, and figuratively acted as a hinge in which the first residues of the N-terminal helix rotated upwards toward repeat III while K-29 remained fixed on the phospholipid layer. These lysines are reported to be the electrostatic anchor because they are located in the unstructured coil region of the N-terminus, and as indicated on the RMSF plots, this region has the highest flexibility than any other region in the protein. Therefore, it is their flexibility and their charge that permit K-26 and K-29 to anchor to an anionic membrane. Based on the simulations it appears that the secondary membrane binding site is mediated by membrane-annexin interactions.

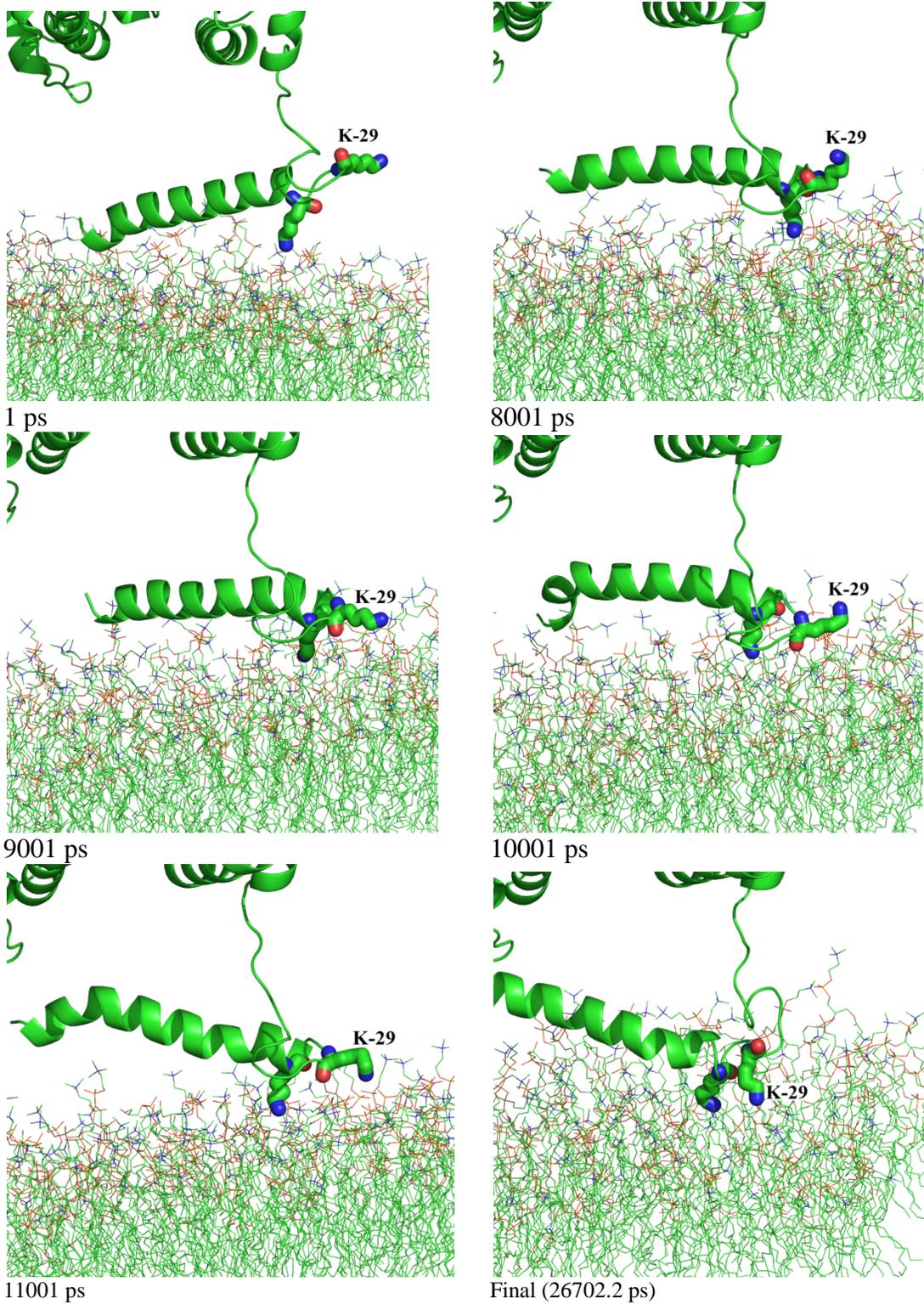


Fig 6.1: Snapshots taken from the trajectory of setup 1 showing orientation of the side-chains of K-26 and K-29. It was observed the K-29 rotated from its initial position to be in closer proximity with the anionic phospholipid layer.

6.3: Calcium Bridging Interactions

Calcium ions bound to sites in the core domain act as bridges connecting the protein with anionic lipid headgroups. Several ligand exchanges within the coordination spheres of calcium ions were observed over the course of the simulation. Of particular interest was calcium 6, located in repeat IV, adopting an oxygen atom from a DOPG molecule into its coordination sphere in exchange for a hydroxyl oxygen atom from T-291 during setup 1. Calcium 6 also loses the hydroxyl oxygen atom from T-291 in setup 2 for a water oxygen atom. This ligand exchange process would seem to be applicable to all ions in annexin I and appears to be how the protein initially binds to a membrane surface, so that the ions coordination number is kept within the favorable range of 6-8.

6.4: Phosphorylation Implications

Porcine annexin I has three potential phosphorylation sites (Tyr21, Thr24 and Ser28), and the *in vitro* phosphorylation of these sites reduces the membrane aggregation activity and increases the required calcium concentration for aggregation by annexin I. This would suggest that this region is involved in membrane aggregation and possibly that it regulates the physiological properties of annexin I through phosphorylation. In an effort to study the conformational effects of phosphorylation at the molecular level, site specific mutations were made at Ser28, after which MD simulations were performed. The results from the trajectories of the wild-type, S28A and S28E mutants were then compared. Porte *et al.* reported that the wild-type and S27A mutants showed the same calcium dependence for vesicle aggregation, while the S27E mutant showed a higher calcium dependence and lower extent of aggregation. Based on the present study, the

wild-type protein and the S28A mutant both underwent similar conformational changes in that the N-terminus ultimately moved closer to the core domain and displayed a partial loss of its helical structure. The S28E mutant retained the secondary structure of its N-terminus throughout the simulation and ultimately moved away from the core domain.

6.5: Stabilization of the N-terminus

Previous studies conducted by Sesham *et al.* demonstrated that the N-terminal domain lost all secondary structure in the absence of phospholipids and solvent accessible after 6 ns. They suggested that the loss of structure could be an initial step in the truncation process, which helps to explain the failed attempts to crystallize the full-length protein in the presence of calcium. This was not the case in the present study. The N-terminus retained its alpha helical secondary structure over a 27 ns trajectory. This implies that electrostatic interactions between the protein residues and the phospholipids play a predominant role in stabilizing the structure N-terminus. The N-terminus also retained its secondary structure during the S28E mutation simulations, whereas it lost its helical structure during the wild-type and S28A simulations. This would suggest that phosphorylation stabilizes the N-terminal helices, and this stabilization is the reason why the S28E mutant shows lower aggregation activity in biochemical assays.

This study is unique in that there have been no reports detailing a molecular dynamics simulation of annexin I positioned between two phospholipid layers to date. Several of our findings were in agreement with previous conclusions regarding annexin I. There are still several modifications which can be applied to the experiment to understand more about the mechanism of fusion the protein induces, such as site specific mutations to further identify residues of interest as well as the possibility of using an

annexin I dimer positioned between the two phospholipid layers, although the latter would be very computationally expensive.

Questions still remain regarding the exact mechanism of aggregation induced by annexin I, such as

1. What is the effect of phosphorylation of S-27 and S-28, two residues between the speculated critical residues K-26 and K-29?
2. Do calcium ions in specific repeats of the the core domain favorably bind to phospholipids over other repeats?
3. What effect would cholesterol, an essential component of mammalian cell membrane, have on the dynamics of annexin I if it were inserted in the phospholipid layer?

Overall, further research is required to answer these questions and to identify the exact location of the interaction site for membrane aggregation. It seems clear though that the mechanism of annexin I induced membrane aggregation is quite complex. Analysis of this system in the future will require a combination of biochemical and MD simulations to fully understand this phenomenon.

6.6: Conclusions

Annexin A1, a 37 kDalton protein previously known as lipocortin 1, has been shown to aggregate neutrophil and chromaffin cell granules and artificial membrane vesicles in the presence of high levels of Ca^{2+} . Molecular dynamics simulations were performed to understand the behavior at the molecular level of annexinA1 when it is in close proximity with two phospholipid layers. MD simulations were performed using the AMBER 9 software package.

The ptraj module of AMBER was used to analyze RMSD, RMSF, B-factor, and inter-atomic distance values over the course of the simulations. The MM-PBSA program was utilized to calculate non-bonded interactions between protein-protein residues and protein-lipid molecules. Visualization of the trajectories was depicted using PYMOL. The results obtained from the analysis were used to draw conclusions on the dynamics of the protein when in close proximity with negatively charged phospholipid layers.

The core domain is relatively stable when compared with the N-terminal domain. Repeat III of the core domain displayed higher RMSF fluctuation than other domains in setup 1. Calcium coordinating residues showed a higher fluctuation than non-coordinating residues.

The calculated B factors of the protein are in close agreement with the X-ray B factors. Also, the residues coordinated with calcium ions appear to have a larger fluctuation over the course of the simulation than those not coordinated. This is in agreement with previous results.

Based on MM-PBSA analysis, the N-terminus has significant electrostatic interactions with the negatively charged phospholipid layer. K-26 and K-29 appear to have a strong attraction to the phospholipid layer, and their side-chains are in close proximity with the phosphoryl groups. This could shed insight into how the N-terminus binds to a second membrane bilayer.

Based on distance analysis, calcium 6 in repeat IV becomes coordinated with an oxygen atom of DOPG124. This calcium loses coordination with a protein residue (T356) as it gains coordination with the phospholipid. This calcium loses coordination

with a protein residue (T291) as it simultaneously gains coordination with the phospholipid.

Based on mutation studies, it appears that phosphorylation of the N-terminus does indeed cause conformational changes to annexin I. These conformational changes in turn can alter the properties of the protein. The wild-type and S28A mutant showed similar conformational changes, in which the N-terminus came within closer proximity to the core domain and was beginning to lose its secondary structure. The S28E mutant retained its secondary structure and its N-terminus moved away from the core domain.

To conclude, based on the simulations performed for this study, annexin I initially binds to an anionic phospholipid layer initially *via* a calcium ‘bridge’ on the convex face of the protein. The exposed N-terminus conforms so that its amino terminus is in close proximity with repeat III of the core domain and the carboxy terminus of its helix is in close proximity with the phospholipid layer. Lysines 26 and 29 appear to help anchor the N-terminus to the phospholipid layer as it undergoes a conformational change.

In an effort to better understand the exact mechanism of membrane aggregation and fusion caused by annexin I, computational and biochemical studies will need to be conducted.

REFERENCES

1. Volker Gerke *et al.* (2002). Annexins: From Structure to Function. *Physiol Rev*, 82, 331-371.
2. Jacob Rand. (2000). The annexinopathies: a new category of diseases. *Biochimica et Biophysica Acta*, 1498, 169-173.
3. L. Parente and E. Solito. (2004). Annexin 1: more than an antiphospholipase protein. *Inflamm. Res.*, 53, 125-132.
4. Matthew Hayes and Stephen Moss. (2004). Annexins and disease. *Biochemical and Biophysical Research Communications*, 332, 1166-1170.
5. Olivier Lambert *et al.* (1997). Structural Analysis of Junctions formed between lipid membranes and several Annexins by cryo-electron microscopy. *J. Mol. Biol*, 272, 42-55.
6. Anja Rosengarth *et al.* (2001). X-ray structure of full length Annexin I and implications for membrane aggregation. *J. Mol. Biol*, 306, 489-498.
7. Maxim Dorovkov *et al.* (2004). Phosphorylation of Annexin I by TRPM7 Channel-Kinase. *Biol. Chem*, 279, 50643-50646.
8. Francoise Porte *et al.* (1996). Change in the N-terminal domain conformation of annexin I that correlates with liposome aggregation is impaired by Ser-27 to Glu mutation that mimics phosphorylation. *BBA*, 1293, 177-184.
9. Joachim Seeman *et al.*(1997). Structural requirements for annexin I – S100C complex formation. *Biochemistry J.*, 319, 123-129.
10. Stephane Rety *et al.*(2000). Structural basis of the Ca²⁺ dependent association between S100C (S100A11) and its target, the N-terminal part of annexin-I. *Structure*, 8 (2), 175-183.
11. Joachim Seemann. (1997). Annexin I targets S100C to early endosomes. *FEBS Letters*, 413 (1), 185-190.
12. J. Seemann *et al.* (1996). The association of annexin I with early endosomes is regulated by Ca²⁺ and requires an intact N-terminal domain. *Mol. Biol. Cell*, 7, 1359-1374.

13. H.A.M. Andree *et al.* (1993). Aggregation of phospholipid vesicles by a chimeric protein with the N-terminal of annexin I and the core of annexin V. *Biochemistry*, *32*, 4634-4640.
14. Eduard Bitto & Wonhwa Cho. (1999). Structural Determinant of the Vesicle Aggregation Activity of Annexin I. *Biochemistry*, *38*, 14094-14100.
15. Volker Gerke, Carl Creutz & Stephen Moss. (2005). Annexins: Linking Ca²⁺ Signalling to Membrane Dynamics. *Nature Reviews*, *6*, 449-461.
16. Manal Swairjo & Barbara Seaton. (1994). Annexin Structure and Membrane Interactions: A Molecular Perspective. *Annu. Rev. Biophys. Biomol. Struct.*, *23*, 193-213.
17. Anja Rosengarth & Hartmut Luecke. (2003). A Calcium-driven Conformational Switch of the N-terminal and Core Domains of Annexin A1. *J. Mol. Biol.*, *326*, 1317-1325.
18. J. A. McCammon. (1976). Molecular dynamics study of the bovine pancreatic trypsin inhibitor. *Models for Protein Dynamics*, pp. 137.
19. J. A. McCammon *et al.* (1977). Dynamics of folded proteins. *Nature*, *267*, pp. 325.
20. W. F. van Gunsteren & H. C. Berendsen. (1977). Algorithms for macromolecular dynamics and constraint dynamics. *Mol. Phys.*, *34*, 1311-1322.
21. Y. Duan, L. Wang & P. A. Kollman. (1998). The early stage of folding of villin headpiece subdomain observed in a 200-nanosecond solvated molecular dynamics simulation. *Proc. Natl Acad. Sci. USA.*, *95*, 9897-9902.
22. J. M. Haile. (1997). *Molecular Dynamics Simulation: Elementary Methods*. USA, Wiley.
23. G. Moraitakis, A. Purkiss & J. Goodfellow. (2003). Simulated dynamics and biological macromolecules. *Rep. Prog. Phys.*, *66*, 383-406.
24. W. Jorgensen *et al.* (1983). Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, *79*, 926-935.
25. H. Berendsen *et al.* (1981). Interaction models for waters in relation to protein hydration. *Intermolecular forces*, pp. 331-342.
26. L. Verlet. (1967) Computer 'experiments' on classical fluids. ii. Equilibrium correlation functions. *Phys. Rev.*, *165*, 201-204.

27. T. Darden, D. York & L. Pedersen. (1993). Particle mesh Ewald: an $N\text{-log}(N)$ method for Ewald sums in large systems. *J. Chem. Phys*, 98, 10089-10092.
28. U. Essmann *et al.* (1995). A smooth particle mesh Ewald potential. *J. Chem Phys*, 103, 8577-8592.
29. G.V. Musat *et al.* (1997). Structure of human annexin I: Comparison of homology modeling and crystallographic experiment. *Biochimie*, 79, 691-703.
30. D. Cregut *et al.* (1998). Hinge-bending in annexins: molecular dynamics and essential dynamics of apo-annexin V and of calcium bound annexin V and I. *Protein Engineering*, 11, 891-900.
31. Tru Huynh *et al.* (1999). Molecular dynamics simulations of the isolated domain 1 of annexin I. *Theoretical Chemistry Accounts: Theory, Computation and Modeling (Theoretica Chimica Acta)*, 101, 82-86.
32. Jana Sopkova-de Oliveira Santos *et al.* (2000). Pathway for Large-Scale Conformational Change in Annexin V. *Biochemistry*, 39, 14065-14074.
33. Jana Sopkova-de Oliveira Santos *et al.* (2001). Annexin A5 D226K structure and dynamics: identification of a molecular switch for the large-scale conformational change of domain III. *FEBS Letters*, 493, 122-128.
34. Tru Huynh *et al.* (2002). Protein Unfolding Transitions in an Intrinsically Unstable Annexin Domain: Molecular Dynamics Simulation and Comparison with Nuclear Magnetic Resonance Data. *Biophys J*, 83, 681-698.
35. Pierre Montaville *et al.* (2002). A New Consensus Sequence for Phosphatidylserine Recognition by Annexins. *The Journal of Biological Chemistry*, 277, 24684-2469.
36. Franci Merzel *et al.* (2006). New Force Field for Calcium Binding Sites in Annexin-Membrane Complexes. *J Comput Chem*, 27, 446-452.
37. R.D. Shesham, L.J. Bartolotti and Y. Li. (2008). Molecular dynamics simulation studies on Ca^{2+} -induced conformation changes of annexin I. *Protein Engineering*, 2, 115-120.
38. D.A. Case *et al.* (2004), AMBER 8, University of California, San Francisco.
39. Wei Wang and Carl E. Creutz. (1992). Regulation of the Chromaffin Granule Aggregating Activity of Annexin I by Phosphorylation. *Biochemistry*, 31, 9934-9939.

