

The Structural Dynamics of Soluble and Membrane Proteins Explored through Molecular Simulations

Dissertation
zur Erlangung des akademischen Grades
Doktor rerum naturalium (Dr. rer. nat.)

von
Samira Yazdi
geboren am 27. Mai 1981 in Mashhad, Iran

genehmigt durch die Fakultät für Verfahrens- und Systemtechnik
der Otto-von-Guericke-Universität Magdeburg

Promotionskommission:
Dr. rer. nat. Matthias Stein
Prof. Dr. rer. nat. Michael Naumann
Prof. Dr. rer. nat. habil. Helmut Weiß

eingereicht am: 3 Mai 2016
Promotionskolloquium am: 27 September 2016

To mom and dad

Contents

Abstract	i
Kurzfassung	ii
Abbreviations and Acronyms	iii
A. Introduction	2
B. Protein Structure	6
B.1 Primary structure	6
B.2 Secondary structure.....	7
B.2.1 The α , 3_{10} , and π helices.....	8
B.2.2 The β sheet and β turn	9
B.2.3 Loops.....	9
B.3 Tertiary and quaternary structures.....	10
C. Repeat Proteins.....	12
C.1 Ankyrin repeats.....	12
C.2 NF- κ Bs and their ankyrin repeat inhibitors	13
C.2.1 NF- κ B proteins	13
C.2.2 The I κ B family.....	14
C.3 The canonical NF- κ B activation pathway	14
C.4 Medical relevance of malfunctioning in the NF- κ B/I κ B α complex.....	15
C.5 Phosphorylation meets the degron recognition motif.....	15
C.6 I κ B α structural map.....	16
C.7 Free I κ B α	16
C.8 The SRD in I κ B α	16
D. Voltage-gated Ion Channels.....	19
D.1 The biological membrane	19
D.2 The membrane potential.....	20
D.3 Voltage-gated potassium channels.....	21
D.4 The voltage sensor and voltage sensing mechanism	21
D.5 Modulation of voltage sensor in the <i>Shaker</i>	22
D.6 Disease and dysfunctional K $_v$ channels.....	23

D.7 PUFA mediated modulation	23
E. Methodology.....	26
E.1 Molecular modeling	26
E.1.1 Comparative modeling and threading	27
E.2 Molecular dynamics simulations	28
E.2.1 Theory	28
E.2.2 Force fields	30
E.3 Free energy calculations.....	31
E.3.1 Sampling along a reaction coordinate	32
E.3.2 Accelerated weight histogram	32
F. Study I.....	35
F.1 Structural modeling of the N-terminal signal-receiving domain of I κ B α	35
F.2 Methods.....	36
F.2.1 Structural modeling	36
F.2.2 System assembly and protocol for MD simulations.....	37
F.3 Results and Discussion	37
F.3.1 Structural elements in the signal-receiving domain (SRD).....	37
F.3.2 Structural refinement by molecular dynamics simulations	40
F.3.3 Conformational change induced in I κ B α in its bound form to NF- κ B.....	45
F.3.4 Free I κ B α vs bound I κ B α	48
G. Study II.....	53
G.1 Double phosphorylation-induced structural changes in the signal-receiving domain of I κ B α in complex with NF- κ B	53
G.2 Methods	55
G.2.1 Structural model of I κ B α /NF- κ B.....	55
G.2.2 Molecular simulation setup	55
G.3 Results.....	56
G.3.1 Local divergence in structural stability promoted by double-phosphorylation.....	56
G.3.2 Double-phosphorylation induces an extended N-terminal conformation of SRD	58
G.3.3 Variation in solvent exposure in Ser32 and Ser36	59
G.3.4 Double-phosphorylation stabilizes region by novel hydrogen bond interactions.....	61
G.3.5 Electrostatic effects	62
G.4 Discussion	63

H. Study III	67
H.1 The Molecular Basis of Polyunsaturated Fatty Acid Interactions with the <i>Shaker</i> Voltage-Gated Potassium Channel.....	67
H.2 Methods	70
H.2.1 <i>Shaker</i> channel system setup in open state.....	70
H.2.2. PUFA and SFA system setup in open state simulations	70
H.2.3 Modeling of <i>Shaker</i> channel in closed state	71
H.2.4 In silico mutagenesis of the VSD.....	72
H.2.5 Small-scale PUFA and SFA systems.....	72
H.3 Results	72
H.3.1 Saturation levels in the carbon tail affect the structural dynamics	72
H.3.2 PUFA interactions with the K_v <i>Shaker</i> channel in the open state	74
H.3.3 PUFA interactions with the K_v <i>Shaker</i> channel in the closed state	77
H.3.4 SFA interactions with the K_v <i>Shaker</i> channel in the open and closed state	78
H.3.5 Electrostatic interactions characterize the PUFA interaction sites	80
H.4 Discussion	82
I. Study IV	85
I.1 The Regulatory Role of Polyunsaturated Fatty Acids on the Free-Energy Landscape of K_v <i>Shaker</i> Channel Deactivation	85
I.2 Methods.....	87
I.2.1 <i>Shaker</i> channel in PUFA and PUFA-free systems.....	87
I.2.2 Accelerated weight histogram calculations.....	88
I.3 Results	88
I.3.1 Sampling approach	88
I.3.2 PUFA affects the free-energy landscape of deactivation	89
I.3.3 PUFA stabilizes the open state configuration	90
I.3.4 3_{10} -helix integrity.....	91
I.4 Discussion	93
Concluding Remarks and Future Perspectives	95
Bibliography	96
Acknowledgements	115
List of Publications	116

Abstract

From a deeper knowledge of the dynamical nature of biological macromolecules we get an improved understanding of their life sustaining roles. Computational methods are of absolute necessity in bringing this about when experimental procedures face their limitations. Throughout this thesis, several computational techniques have been applied to study the complex nature of protein complexes either in a solvent or a membrane environment. The two classes of molecules presented in this thesis are the water-soluble proteins represented by repeat proteins (studies I & II) and membrane proteins illustrated by voltage-gated ion channels (studies III & IV). The insight gained into the structural characteristics and patterns of interactions of these proteins is invaluable in drug design.

STUDIES I & II. I κ B α , the transcription factor NF- κ B inhibitor, is an ankyrin repeat protein that retains NF- κ B in the cytoplasm. Recent protein crystal structures of I κ B α in complex with NF- κ B have not revealed structural details about the N-terminal signal receiving domain (SRD), which harbors the sites of post-translational modifications at sites Ser32 and Ser36. By combining secondary structure annotation and domain threading followed by molecular dynamics simulation, I have showed that the SRD possesses well-defined secondary structure elements: 3 additional stable α -helices supplementing the six ankyrin repeats (ARs) present in crystallized I κ B α . Moreover, differences in structural topology and dynamics were observed by comparing the structures of free and bound I κ B α . I also investigated the effect of post-translational mono- and double-phosphorylation of the serine residues of the SRD. Mono-phosphorylation at either Ser32 or Ser36 was not sufficient to induce significant structural changes in the secondary structure of the SRD of I κ B α . Double-phosphorylation yielded a reduced distance between the C α atoms of these serine residues, inducing an extended conformation, which renders it accessible by the E3 ligase.

STUDIES III & IV. Polyunsaturated fatty acids (PUFA) have been reported to influence the gating mechanism by electrostatic interactions to the voltage-sensor domain (VSD) of voltage-gated potassium (K_V) channels, however the exact PUFA-protein interactions have remained elusive. I have reported on the interactions between the *Shaker* K_V channel in open and closed states and a PUFA-enriched lipid bilayer using molecular dynamics simulations. A putative PUFA binding site was determined in the open state of the channel located at the protein-lipid interface in the vicinity of the extracellular halves of the S₃ and S₄ helices of the VSD. In particular, the lipophilic PUFA tail covered a wide range of non-specific hydrophobic interactions in the hydrophobic central core of the protein-lipid interface, while the carboxylic head group displayed more specific interactions to polar/charged residues at the extracellular regions of the S₃ and S₄ helices, encompassing the S₃-S₄ linker. Moreover, by studying the interactions between saturated fatty acids (SFA) and the *Shaker* K_V channel, the study confirmed an increased conformational flexibility in the polyunsaturated carbon tails compared to saturated carbon chains, which may explain the specificity of PUFA action on channel proteins. I also computed the energetics of the free-energy landscape of the *Shaker* K_V channel in lipid bilayers free from as well as enriched with PUFAs. By choosing a reaction coordinate along the vertical translation of S₄ towards its down state in the deactivation pathway, I investigated the free energy differences in passing the first energy barrier of deactivation, i.e. going from the open (O) state to the closed-1 (C₁) state, in a K_V channel that appears to be affected by PUFAs.

Kurzfassung

Durch das tiefere Verständnis der dynamischen Natur biologischer Makromoleküle gewinnen wir eine verbesserte Einsicht in ihre lebenswichtigen Rollen und Funktionen. Computergestützte Ansätze sind hierbei eine absolute Notwendigkeit, wenn experimentelle Methoden angesichts ihrer zeitlichen und räumlichen Begrenzungen nicht ausreichen. In dieser vorgelegten Arbeit wurden verschiedene computergestützte Techniken angewendet, um die komplexe Art von verschiedenen Proteinkomplexen in Lösung und in einer Membranumgebung zu untersuchen. Die zwei Klassen von Molekülen in dieser Arbeit sind wasserlösliche Proteine am Beispiel der Repeatproteine (Studien I und II) und membrangebundene Proteine am Beispiel des spannungsgesteuerten Ionenkanals (Studien III und IV). Die gewonnenen Einsichten in die strukturellen Besonderheiten und Wechselwirkungsmuster dieser Proteine ist wertvoll für mögliche Medikamentenentwicklungen.

STUDIEN I und II. $\text{I}\kappa\text{B}\alpha$, der Inhibitor des Transkriptionsfaktors NF- κB , ist ein Ankyrin-Repeat-Protein, welches NF- κB im Zytoplasma hält. Die neuesten Proteinkristallstrukturen von $\text{I}\kappa\text{B}\alpha$ im Komplex mit NF- κB haben nicht die strukturellen Details der N-terminalen signalempfangenden Domäne (SRD) aufklären können, die die Orte der posttranslationalen Modifikationen an den Positionen Ser32 und Ser36 enthält. Durch eine Kombination von Sekundärstrukturannotationen und Proteindomänenthreading, gefolgt von Molekulardynamiksimulationen, habe ich gezeigt, dass die SRD wohl definierte sekundärstrukturelle Elemente enthält: 3 weitere stabile α -Helices, zusätzlich zu den sechs Ankyrin Repeats (AR) aus dem kristallisierten $\text{I}\kappa\text{B}\alpha$. Des Weiteren wurden strukturelle Topologien und die Dynamiken von Strukturen des freien und gebundenen $\text{I}\kappa\text{B}\alpha$ untersucht. Ich habe auch den Effekt der posttranslationalen einfachen und zweifachen Phosphorylierung der Serinaminosäuren in der SRD untersucht. Die einfache Phosphorylierung entweder am Ser32 oder Ser36 war nicht ausreichend, um signifikante strukturelle Veränderungen in der Sekundärstruktur der SRD von $\text{I}\kappa\text{B}\alpha$ zu induzieren. Die zweifache Phosphorylierung führte zu einem verringerten Abstand zwischen den Ca-Atomen dieser Serinaminosäurereste, induzierte eine gestreckte Konformation des Proteinrestes und machte diesen so der E3 Ligase zugänglich.

STUDIEN III UND IV. Mehrfachungesättigte Fettsäuren ('polyunsaturated fatty acids', PUFA) wurden in der Literatur beschrieben, die Schaltvorrichtung des spannungsgesteuerten Kalium (K_V) Kanals durch elektrostatische Wechselwirkungen mit der Spannungssensordomäne (VSD) zu aktivieren. Die genaue PUFA-Protein Wechselwirkungen konnten bisher nicht aufgeklärt werden. Ich habe die Wechselwirkungen zwischen dem *Shaker* K_V -Kanal im offenen und geschlossenen Zustand und einer PUFA-angereicherten Lipiddoppelschicht mit Molekulardynamiksimulationen untersucht. Eine mögliche PUFA-Bindungsstelle wurde im offenen Zustand des Kanals lokalisiert an der Protein-Lipid-Grenzfläche in der Nähe der extrazellulären Hälften der S_3 - und S_4 -Helices der VSD. Der lipophile PUFA-Schwanz zeigte eine große Anzahl von nicht-spezifischen hydrophoben Interaktionen im zentralen hydrophoben Kern der Protein-Lipid-Grenzfläche. Die Carboxylat-Kopfgruppe dagegen wies spezifischere Wechselwirkungen mit polaren/geladenen Resten an den extrazellulären Regionen der S_3 - und S_4 -Helices auf, den S_3 - S_4 Linker umfassend. Durch eine Untersuchung der Wechselwirkungen zwischen gesättigten Fettsäuren (SFA) und dem *Shaker* K_V -Kanal konnte darüber hinaus eine zunehmende konformationelle Flexibilität der ungesättigten im Vergleich zu den gesättigten Kohlenstoffketten bestätigt werden. Dies kann die Spezifität der Wirkung von ungesättigten Fettsäuren (PUFA) auf Kanalproteine erklären. Ich habe ebenfalls die relativen Energien der freien Energielandschaft des *Shaker* K_V -Kanals in Lipiddoppelschichten in Abwesenheit und angereichert mit PUFA-Molekülen berechnet. Durch die Wahl einer Reaktionskoordinate entlang der vertikalen Translation von S_4 vom offenen zum geschlossenen Zustand entlang des Deaktivierungspfades konnten die Unterschiede in freien Energien bei der Überquerung der ersten Barriere der Deaktivierung, d.h. vom offenen (O) zum geschlossenen-1 (closed-1', C1) Zustand in einem K_V -Kanal berechnet werden, der betroffen durch PUFAs ist.

Abbreviations and Acronyms

AR	ankyrin repeat
ARD	ankyrin repeat domain
AWH	accelerated weight histogram
C-terminal	carboxyl-terminal
DHA	docosahexaenoic acid
K _v	voltage-gated potassium channel
MD	molecular dynamics
NLS	nuclear localization sequence
N-terminal	amino-terminal
NMR	nuclear magnetic resonance
PD	pore domain
PDB	protein data bank
PEST	proline glutamate serine threonine
PIP2	phosphatidylinositol-4,5-bisphosphate
POPC	1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine
PUFA	polyunsaturated fatty acid
RMSD	root mean square displacement
RMSF	root mean square fluctuation
SCF	Skp1-Cullin-F-box
SFA	saturated fatty acid
SRD	signal receiving domain
SSE	secondary structure element
TM	transmembrane
VSD	voltage-sensor domain

One and three letter codes for amino acids:

Ala (A)	Alanine	Leu (L)	Leucine
Arg (R)	Arginine	Lys (K)	Lysine
Asn (N)	Asparagine	Met (M)	Methionine
Asp (D)	Aspartic acid	Phe (F)	Phenylalanine
Cys (C)	Cysteine	Pro (P)	Proline
Gln (Q)	Glutamine	Ser (S)	Serine
Glu (E)	Glutamic acid	Thr (T)	Threonine
Gly (G)	Glycine	Trp (W)	Tryptophan
His (H)	Histidine	Tyr (Y)	Tyrosine
Ile (I)	Isoleucine	Val (V)	Valine

Chapter A

Introduction

Living organisms are composed of cells, basic structural and functional units of life that carry out vital functions such as maintenance, recycling, disposing of waste, adapting to its environment and replication. Cells can exist as uni-cellular organisms such as many species of bacteria and protozoa, or they can be part of multi-cellular complex units.

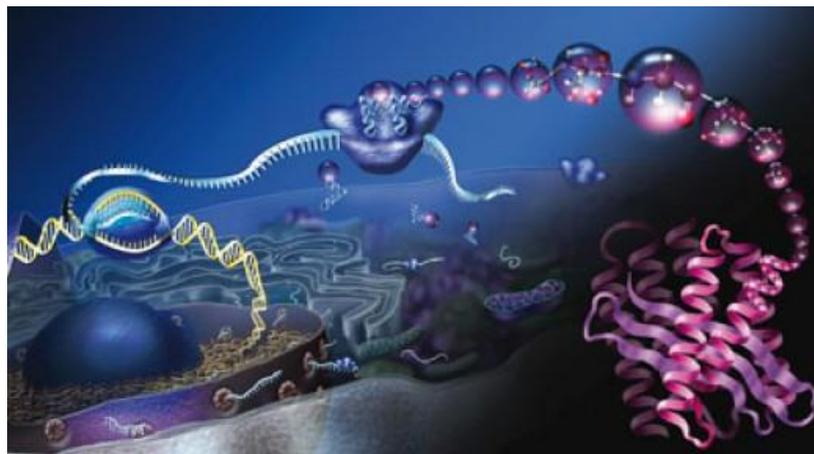
Cells are shaped by a structure known as the cell membrane, also referred to as the plasma membrane, separating the contents of the cell from the surroundings. The plasma membrane is constituted of phospholipids, fat-based molecules, which prevent hydrophilic substances from entering or leaving the cell. Embedded within the membrane are proteins that serve a range of functions. The gatekeepers decide on which substances that are allowed to diffuse across the membrane; markers identify a cell as friendly and part of the organism or as a foreign molecule; fasteners bind cells together and enable them to function as one unit; and the communicators have the purpose of sending and receiving signals from other cells and the environment.

Within the interior walls of the cells an aqueous solution called the cytoplasm retains all the cellular machinery and structural elements of the cell. The largest organic intra-cellular components in the cytoplasm include carbohydrates, lipids, nucleic acids, and proteins. Carbohydrates expressed in simple and complex forms, are the starches and sugars in the cell. The simple carbohydrates supply the cell's immediate energy demands and the complex carbohydrates act as intracellular energy storage. Lipids are the components of the plasma membrane in addition to many intracellular membranes. They partake in conveying signals both within the cell and from the outside to the inner cell. A cell's genetic code is expressed through nucleic acids of which there are two major classes: deoxyribonucleic acid (DNA) and ribonucleic acid (RNA). DNA contains the information, which is necessary to build and maintain the cell, while RNA is involved with the expression

of the information stored in DNA. An organism's genome is the complete set of genes that is made up of DNA and contained within the nucleus. In bacteria and archaea, the nucleus is not separated from the cytoplasm, whereas in eukaryotes, the nucleus contains nuclear material enclosed within a double membrane, the nuclear envelope. The information programmed into the nucleotide sequence of a cell's nucleic acids determines the structure of every protein and essentially of every molecule and cellular component. The flow of genetic information in cells from DNA encoding RNA and RNA encoding protein is known as the central dogma of life (Figure A.1) [1].

Figure A.1 The central dogma of life: from DNA, to RNA, to protein. Figure adapted from [2].

Proteins serve a variety of functions in the cell, both catalytic and structural and are made from chains of smaller molecules called amino acids. The malfunctioning of proteins leads to a great many illnesses, ergo understanding their structure and thus function is central to deciphering biology and life. In 1958, John Kendrew and colleagues showed for the first time the three-dimensional structure of a protein, the low-resolution structure of myoglobin, a monomeric heme protein [3]. This was followed two years later, in 1960, by



the 5.5 Å resolution structure of haemoglobin, a tetrameric heme protein, solved by Max Perutz and co-workers which was the first indication of protein families [4]. This marked only the beginning of a fascinating newborn world of protein science where atomic-level resolution structures came to give way for a new understanding of protein structure and their intricate function.

The aim of this thesis has been to study proteins, their structure and interactions, and concludes the results and findings of my projects during my PhD studies. I have studied both soluble and membrane proteins employing computational techniques such as molecular modeling, molecular dynamics simulations, and free-energy calculations. In this thesis I will present the obtained results on four different studies. In study I, to better understand the structural aspects of the NF- κ B activation pathway, I have modeled the three-dimensional structure of the N-terminal region of the NF- κ B inhibitor, I κ B α . Study II has been the focus of posttranslational modification, in particular phosphorylation, and its structural effects on I κ B α . In study III, I change focus over to membrane proteins and investigate the molecular details of the interactions between the *Shaker* voltage-gated

potassium channel and polyunsaturated fatty acids. Finally, study IV deals with mapping the free-energy landscape of polyunsaturated fatty acid modulation on the *Shaker* voltage-gated potassium channel.

Chapter B

Protein Structure

B.1 Primary structure

The primary structure of proteins is the sequence of amino acids, which forms a chain and determines the fold of a protein. Each amino acid is specified by a nucleotide sequence. Nucleotides are the building blocks of DNA that contain information for the synthesis of RNA or protein. There are 4 major bases in DNA: adenine (A), guanine (G), cytosine (C), and thymine (T) and these are linked covalently together via a phosphodiester linkage. A triplet of nucleotides, named a codon, codes for a specific amino acid. The three letter code gives $4^3 = 64$ combinations of amino acids with most of the 20 natural amino acids having more than one codon.

Every amino acid has a hydrogen, carboxyl group, and an amino group bound to a central $C\alpha$ in addition to a side chain group. This side chain group varies in structure, size, and electric charge between the different amino acids. Classification of amino acids based on the properties of the side chain results in five major classes: nonpolar aliphatic, aromatic, polar uncharged, positively charged, and negatively charged. Amino acids come together to form proteins by joining covalently through an amide linkage, named a peptide bond. The formation of the peptide bond occurs through a condensation reaction in which the carboxyl group of one amino acid reacts with the amino group of another resulting in the production of a water molecule.

Several amino acids join together to form a peptide, and once incorporated into a peptide chain, amino acids are referred to as residues. A peptide chain has two free ends: an amino-terminal (N-terminal) end which is the amino acid residue at the end with a free amino group, and the carboxyl-terminal (C-terminal) end which is the residue at the other

end with a free carboxyl group. Protein chains built up of less than 50 amino acids are referred to as peptides, and when longer either polypeptide or 'protein' [1].

B.2 Secondary structure

The secondary structure of a protein describes stable arrangements of amino acid residues resulting in recurring structural patterns. Three dihedral angles, ϕ (phi), ψ (psi), and ω (omega), known as torsion angles, define protein formation by rotation about each bond in the protein backbone.

ϕ involves the C-N-C $_{\alpha}$ -C bonds, with rotation occurring about the N-C $_{\alpha}$ bond; ψ involves the N-C $_{\alpha}$ -C-N bonds, with rotation occurring about the C $_{\alpha}$ -C bond; and ω involves the C $_{\alpha}$ -C-N-C $_{\alpha}$ bonds, with rotation occurring about the C-N bond. The ϕ and ψ angles are defined as $\pm 180^{\circ}$, but the peptide bond ω is not free to rotate. In theory, the ϕ and ψ angles can take any value in the -180° and $+180^{\circ}$ range, but because of steric interference between the atoms in the protein backbone and amino acid side chains there is a restriction on the allowed angles.

The Ramachandran plot is a two-dimensional map of the allowed ϕ and ψ angles in a protein, based on calculations using known van der Waals radii and dihedral angles (Figure B.1) [5,6]. When the torsion angles ϕ and ψ remain the same throughout a segment they shape a regular secondary structure. Two of the main secondary structure folds proposed by Pauling *et al.* are the α helix and β sheet [7-9]. Other commonly found secondary structure elements include the 3_{10} -helix, π -helix, β -turn, loops and coils. In addition to being the foundation of tertiary structure, secondary structure folds play other roles such as allowing efficient packing of atoms, contributing to protein stability, and forming biologically active structures.

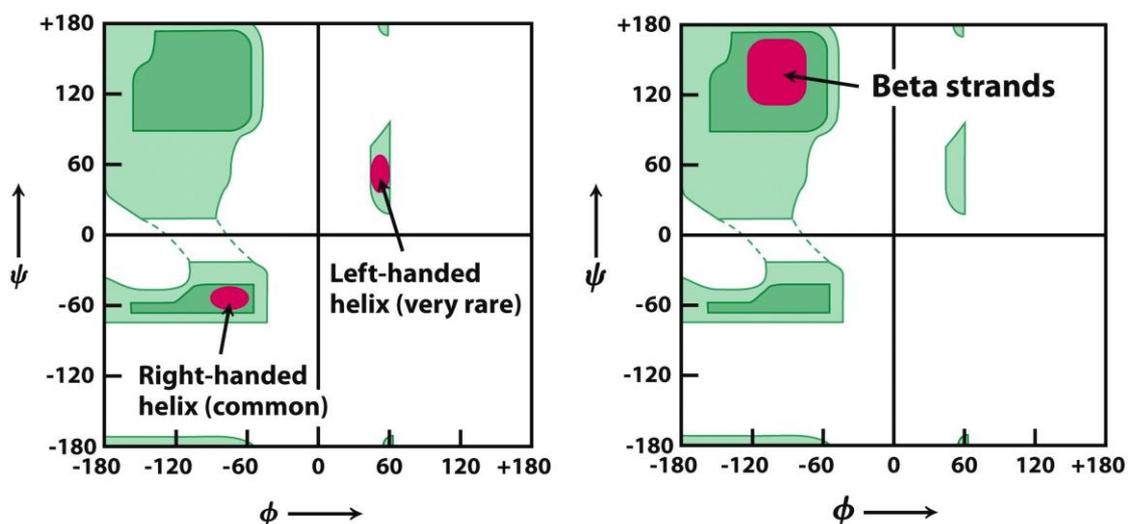


Figure B.1 Ramachandran plot. The shaded areas in light and dark green represent the allowed and favored regions for backbone dihedral angles ψ against ϕ . The pink regions correspond to regions allowing α -helical (left) and β -sheet conformations (right). Figure adapted from [10].

B.2.1 The α , 3_{10} , and π helices

The α -helix structure is the simplest arrangement a polypeptide chain can adopt given its rigid peptide bonds. The helical structure can be between 5 to 40 amino acid residues, corresponding to backbone ϕ and ψ angles of -57° and -47° , respectively and each helical turn including 3.6 amino acid residues [1]. It happens that these structures deviate from the dihedral angles giving rise to bends or kinks in the helical structure.

The α -helix is a right-handed helix that is stabilized by an extensive network of hydrogen bonds between the peptide bonds, where each successive turn of the α helix is held to adjacent turns by three or four hydrogen bonds (Figure B.2). In specific, each bond is formed between a backbone carbonyl group (C=O) and a backbone amide group (N-H) of the fourth amino acid located down the amino-terminal side of the peptide. Hydrogen bonds are not the only main stabilizing factor in the helix, nonpolar and van der Waals interactions between the residue side chains in particular the C_α atom also contribute to the stability of the helix [11,12]. Since the helical conformation is the most compact out of all secondary elements, amino acid residues that have smaller and more linear side chains are more preferred since there is less chance for steric interference. The most commonly occurring amino acids in α -helices are Ala, uncharged Glu, Leu, Arg, charged Met, and charged Lys.

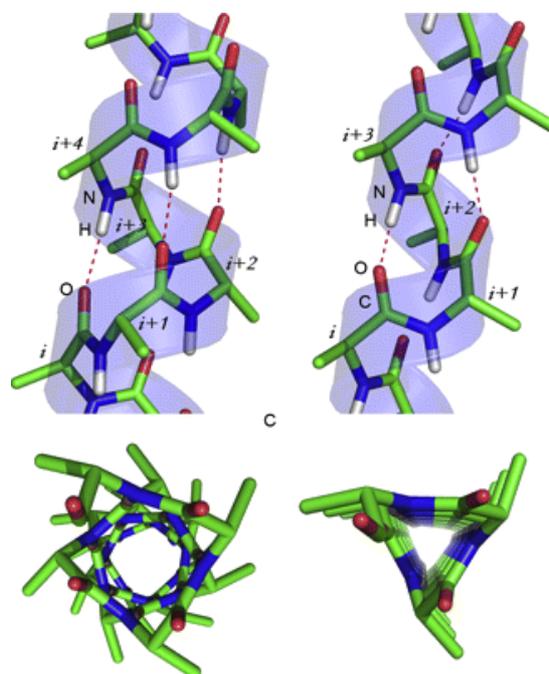


Figure B.2 Helical conformations. Comparison of the α -helix (left) and the 3_{10} -helix (right). The 3_{10} -helix is tighter wound with residue i interacting with residue $i+3$, versus the $i+4$ interaction in the α -helix. Figure adapted from [13].

The 3_{10} -helix is narrower and longer than the α -helix, including 3 residues and 10 backbone atoms per turn, hence the origin to its name. The 3_{10} -helix is a rarely observed secondary structure element in particular because of the $i \rightarrow i + 3$ hydrogen bond pattern which renders this helix energetically unfavorable due to the poor backbone dipoles arrangement.

The π -helix on the other hand is shorter and wider than the α -helix, with 4.4 residues per turn. The π -helix is also an energetically unfavorable structural element because of its $i \rightarrow i + 5$ hydrogen bond pattern. Both the 3_{10} and π -helix mainly appear at the edges of α -helices, consisting of no more than a few residues.

B.2.2 The β sheet and β turn

Contrary to the compact structure of the α -helix, the β -strand adopts a less compact structure with the amino acid residues preferring conformations with $\phi = -120^\circ$ and $\psi = 120^\circ$. β -strands typically have between 5 and 10 residues, with the distance between the C_α atoms of the adjacent residues over twice as long as in the α -helix, thereby making the β -strand a more extended conformation. The extended conformation of the β -strand does not permit intermolecular hydrogen bonds within the same strand, or stabilizing van der Waals contacts. Consequently, it is rare that β -strands appear in isolation and instead are found to form flat structures arranged alongside one another, known as the β -sheet.

There are two types of β -sheets, the anti-parallel and the parallel. In the anti-parallel β -sheet the strands are separated from one another by a few residues, that form a loop known as the β -turn, allowing the β -strands to line side by side in opposite directions. Conversely, in the parallel β -sheet, the residues connecting each β -strand are long enough to allow for the arrangement of the strands in the same direction in a parallel fashion.

The hydrogen bond network in β -sheets is between strands where most of the backbone groups are involved in hydrogen bonds. The preferred amino acids in β -sheets are Asp, Asn, Ser, and Pro which appear in their C-termini [14], whereas the amino acid preference in the N-termini are Lys and Arg. In general, β -sheets are more accommodating to residues with bulkier side chains, owing to their less compact structure.

B.2.3 Loops

The segments of proteins that lack a regular secondary structure are referred to as loops. This absence of secondary structure makes loops more flexible, which also means that the loop backbones are not involved in a hydrogen bond network like helical and β conformations [15]. Loops consist mainly of polar residues, and are more prone to appear on the surface of the protein as opposed to the hydrophobic core.

B.3 Tertiary and quaternary structures

Tertiary structure refers to the three-dimensional folding of proteins, which normally determines the function of a protein. Most of the interactions between the amino acids in a protein are non-covalent, i.e. van der Waals, electrostatic or nonpolar [16]. This is particularly important since the large number of interactions contributes towards stabilizing the tertiary structure of proteins, but since these interactions are relevantly weaker than covalent bonds, protein structures remain dynamic rendering them indispensable to their functions. In the core of the protein that is mainly built of nonpolar residues, it is primarily nonpolar interactions that are the strongest. However, on the surface of the protein, it is mostly electrostatic interactions that are prevalent, where polar residues interact with each other or the surrounding water molecules and ions.

The structural arrangement of proteins with more than one independent subunit is described as quaternary structure. When the subunits are identical the protein is a homomer, and when they are different it is called a heteromer. Quaternary structures are stabilized mainly by non-covalent interactions, and the formation of these structures is driven by the hydrophobic effect [17].

Chapter C

Repeat Proteins

Repeat proteins are non-globular folds that are built from repeated motifs of between 20 and 40 amino acids stacked together to form extended super-helical structures [18,19]. These proteins differ from globular proteins by their repetitive and elongated structures that are governed mainly by local and short-range interactions. Examples of a few repeat protein families with different motifs are: β strand – link – β strand motif such as hexapeptide (HPR) and WD₄₀ repeats, α -helix – link – α -helix motif such as TPR and HEAT repeats, α -helix – link – β -strand motif such as leucine-rich (LRR) repeats, α -helix – β -hairpin or loop – α -helix motif such as ankyrin (ANK) repeats [20].

No single motif is yet known to form a stable folded unit, as stacks of these structural units are needed to form stable folded structures. What unifies different repeat proteins is their ability to mediate protein-protein interactions. By forming scaffolds, these proteins are able to bind varying binding partners, thereby making these proteins open to varied cellular functions such as protein transport, protein complex assembly, and protein regulation. Next, I will turn my focus to ankyrin repeats and discuss in further detail the nature of one of the most frequently observed motifs.

C.1 Ankyrin repeats

The ankyrin repeat (AR) is present in about 6 % of eukaryotic protein sequences, making it one of the most common sequence motifs. The ankyrin repeat is a 33 residue sequence motif that was first discovered in the yeast cell cycle regulator Swi6/Cdc10 and the *Drosophila* signaling protein Notch [21] and takes its name from the human erythrocyte protein ankyrin [22]. While ankyrin repeats are found in all three kingdoms including

bacteria, archaea, and eukarya in addition to a few viral genomes, the majority of these repeats are present in eukaryotes. These modular protein domains, which act as scaffolds for molecular interactions, are deemed crucial for the development of many signaling pathways [23].

The PFAM and SMART databases reveal proteins containing up to 33 ankyrin repeats with the majority having six or fewer repeats. Identification of terminal repeats has proven to be a challenge because of the divergence of many well-conserved hydrophobic residues. Terminal repeats often deviate from the customary consensus sequence because these positions have through evolution been replaced with polar residues to enhance favorable interactions with the solvent. Also, as a result of truncation, terminal repeats are often not accounted for in protein sequences. This opens up the possibility of detected ankyrin repeat proteins having one or two more additional repeats in their sequences in reality. Ankyrin repeats can either exist as single protein domains or in coexistence with other domains in the same protein [24].

Individual ankyrin repeats fold into two antiparallel α -helices followed by a β -hairpin or a long loop. The repeats are assembled in an array and form an L-shaped domain, resembling a cupped hand with the fingers represented by the hairpins and the palm by the helices [20,25]. Structures of proteins with many ankyrin repeats reveal an overall curved shape. Ankyrin repeats are quite well-conserved and rarely deviate from the canonical motif, with the exception of insertions mainly in the loop areas. The helices in the repeats are primarily stabilized by hydrophobic interactions, whereas the β -hairpins and the loops are connected by a hydrogen bonding network. The inner helix is seven residues long extending from positions 5-12 in the canonical sequence, while the outer helix being nine residues long spans the positions 15-24. In positions 4-7 is the prevailing TPLH motif, with the proline causing a tight turn leading to the L-shaped structure of the ankyrin repeat [24,25].

The ankyrin repeat is present in many biologically important proteins that are implicated in protein functions such as cell-cell signaling, cytoskeleton integrity, transcription and cell-cycle regulation, inflammatory response, and development. A few families containing ankyrin repeats include the INK4 tumor suppressor family, Notch the signaling protein involved in development, the TRP cation channel family, and NF- κ B, the transcription factor regulating inflammatory response inhibited by the ankyrin repeat protein I κ B α . Before we shift our attention to the ankyrin repeat protein I κ B α , we will first take a look at the NF- κ B activation pathway and understand its significance.

C.2 NF- κ Bs and their ankyrin repeat inhibitors

C.2.1 NF- κ B proteins

Initially NF- κ B was identified as a transcription factor binding to the enhancer element of the immunoglobulin (Ig) κ light-chain gene in B cells [26,27]. However, it later became apparent that it is a major regulator of innate and adaptive immunity and inflammatory responses present in every cell type, retained in the cytoplasm in an inactive mode bound to the I κ B inhibitors [28,29]. When in an active DNA-binding mode, NF- κ B is a heterodimer

composed of a combination of members from the NF- κ B/Rel family. The known members of this family are: NF- κ B1 (p50 and its precursor p105), NF- κ B2 (p52 and its precursor p100), c-Rel, RelA (p65), and RelB [30]. A well-conserved 300 amino acid Rel homology region (RHR), composed of two immunoglobulin domains, is common between these proteins. The RHR region is involved in DNA binding, dimerization, and interaction with the I κ B proteins, and contains a nuclear localization sequence (NLS) [31]. Knockout studies in mice have identified the p65/RelA protein to be crucial for survival. The p65/p50 heterodimer is the most occurring NF- κ B member throughout different cells and is the protein complex in question when referring to NF- κ B [30,32].

C.2.2 The I κ B family

The members of the I κ B family include I κ B α , I κ B β , I κ B γ , I κ B ϵ , Bcl-3, p105, p100, and the *Drosophila* protein Cactus [30,33]. The I κ B members are ankyrin repeat proteins each containing six or seven ankyrin repeats which enable binding to the RHR and masking of the NLS in NF- κ B. The most vital NF- κ B regulators are I κ B α , I κ B β , and I κ B ϵ all of which have a regulatory N-terminal region that is required for stimulus induced degradation, an important step in NF- κ B activation [34].

One of the first and best-characterized I κ B family members is I κ B α [35]. Upon synthesis, I κ B α enters the nucleus and binds to NF- κ B enabling its dissociation from DNA, leading it back to the cytoplasm through its nuclear export sequence (NES) [36]. The three classical I κ Bs experience signal induced proteasomal degradation with varying kinetics [37]. As opposed to much slower kinetics in I κ B β and I κ B ϵ degradation, I κ B α displays higher kinetics in response to inflammatory stimuli such as TNF- α and lipopolysaccharide (LPS) and via a negative feedback loop, synthesized I κ B α enters the nucleus to bind to deacetylated p65/p50 dimers and carry them back to the cytoplasm [36,38].

C.3 The canonical NF- κ B activation pathway

The most well known NF- κ B activation pathway is the so called 'canonical' pathway, which through a variety of stimuli activates NF- κ B and in turn induces the rapid degradation of I κ B α and causing the nuclear accumulation of NF- κ B. One of the primary and critical events in this activation pathway is the phosphorylation of I κ B α at serines 32 and 36 in the N-terminal regulatory domain by the I κ B kinase IKK [39,40].

IKK consists of two catalytic kinases, the helix – loop - helix (HLH) serine-threonine kinases, IKK α and IKK β [40,41], and a noncatalytic signal recognition scaffolding protein known as IKK γ [42], the IKK associated protein IKAP [43], and FIP the 14.7 interacting protein with the role of a signaling adapter and scaffolding protein. IKK β activation is identified as the main regulatory step in the canonical pathway since IKK β is known to have a 30-fold higher activity toward I κ B α [41,44] and its genetic deficiency blocks I κ B α proteolysis by various inducers [45]. IKK β is triggered to interact with its substrate through its activation by phosphorylation of two activation loop serines [46].

Phosphorylation of I κ B α leads to the substrate being recognized by the F-box/WD40 E3RS^{I κ B}/ β -TrCP that initiates the polyubiquitination of I κ B α at lysines 21 and 22

by the Skp1-Cullin-F-box(SCF)-type E₃ ubiquitin ligase which results in the rapid degradation of I κ B α through the 26S proteasome [35,47-49]. Upon degradation of its inhibitor, the free NF- κ B NLS is exposed and targeted for binding to karyopherins and its translocation to the nucleus [50].

C.4 Medical relevance of malfunctioning in the NF- κ B/I κ B α complex

Misregulation of the NF- κ B activation pathway has implications in many chronic diseases such neurodegenerative diseases, auto-immune diseases and cancer [51-53], making it a suitable therapeutic target. Any of the events of the signaling pathway can be a target for therapeutic intervention whether it is the phosphorylation, ubiquitination, or the degradation of the I κ B proteins.

C.5 Phosphorylation meets the degron recognition motif

In resting cells, the half-life of a stable I κ B α is 138 min, but upon phosphorylation of residues 32 and 36, the half-life of I κ B α in these stimulated cells is reduced to 1.5 min [54]. Phosphorylation of I κ B α induces recognition by the SCF-type E₃ ligase. The E₃ recognizes a variety of different substrates allowing for specificity but only phosphorylated substrates through its F-box component, the variable receptor subunit [55]. The SCF complex thereafter forms a bridge between the substrate I κ B α and the E₂ ubiquitin-conjugating enzyme, which enables the ubiquitination of I κ B α at lysine sites 21 and 22 [56].

The F-box domain of E₃, β -TrCP, was identified by mass spectroscopy and interacts specifically with a phosphorylation-based motif through its WD₄₀ domain [57]. This phosphorylation-based motif, DpSG ϕ XpS, is also known as the degron motif (ϕ represents a hydrophobic residue and X stands for any residue).

The two best characterized substrates of β -TrCP are I κ B α and β -catenin both sharing the degron motif. It was Pavletich and colleagues that shed light on the structural basis for recognition of a phosphorylated I κ B α by the E₃ ligase by solving the crystal structure of β -TrCP in complex with the phosphopeptide β -catenin [56]. The interacting face of β -TrCP is the β -propeller shaped by the seven WD repeat domain of β -TrCP where a central groove going through the β -propeller bonds with the degron-containing segment of the substrate. The three residues accommodated between the two phosphoserines insert the farthest into the groove making intermolecular contacts in a rather buried environment. The two phosphorylated serines on the other hand bind at the rim of the groove through hydrogen bonds and electrostatic interactions.

C.6 $\text{I}\kappa\text{B}\alpha$ structural map

$\text{I}\kappa\text{B}\alpha$ contains a signal receiving domain (SRD), six to seven ankyrin repeat units [58] and a largely unstructured PEST (enriched in amino acids proline (P), glutamate (E), serine (S) and threonine (T)) domain at the C-terminus. The C-terminal PEST domain is also the site of post-translational modifications due to the casein kinase II (CK2) phosphorylation at positions 283, 288, 291, 293 and 299 [59,60].

The two crystal structures of $\text{I}\kappa\text{B}\alpha$, determined by Huxford *et al.* [61] at 2.3 Å and by Jacobs and Harrison [62] at 2.7 Å resolution, illustrate how the six repeating ankyrin domain assumes the shape of an arched cylinder assembled on top of the interface of the NF- κB heterodimer. Every repeat unit in $\text{I}\kappa\text{B}\alpha$ is composed of two α -helices connected to each following repeat with a loop of varying size and a β -hairpin turn containing short β -strands. However, repeats 1, 3, and 4 deviate from the canonical 33 amino acid repeat unit. These repeats are longer than the repeat units in the ankyrin consensus sequence, with the insertions contained in the loop sections, as these regions are those with the lowest sequence similarity among all ankyrin repeat proteins. Lack of homology is also observed in the sixth and last repeat unit, where the dissimilarity falls in after the second helix clearing the last 11 residues of any secondary structural elements [61].

C.7 Free $\text{I}\kappa\text{B}\alpha$

Free $\text{I}\kappa\text{B}\alpha$ (67-317) was characterized by circular dichroism (CD) spectroscopy, 8-anilino-1-naphthalenesulphonic acid (ANS) binding, differential scanning calorimetry (DSC), and amide hydrogen/deuterium exchange experiments [63]. The CD spectrum of free $\text{I}\kappa\text{B}\alpha$ is nearly identical to the CD spectrum of the $\text{I}\kappa\text{B}\alpha$ /NF- κB complex but it shows significant ANS binding and rapid amide exchange over much of the protein. These findings suggest that the secondary structure of $\text{I}\kappa\text{B}\alpha$ is formed but the tertiary structure may not be compact. The β -hairpins of AR2 and AR3 were remarkably resistant to exchange, whereas AR5 and AR6 exchanged completely within the first minute in free $\text{I}\kappa\text{B}\alpha$. When bound to NF- κB , the β -hairpins of AR5 and AR6 showed dramatically less exchange in the bound state [64].

C.8 The SRD in $\text{I}\kappa\text{B}\alpha$

The SRD of $\text{I}\kappa\text{B}\alpha$ is the central signal receiving and transmitting domain when activating NF- κB . It contains sites for post-translational modifications (phosphorylation by kinases IKK α and IKK β [61] at Ser32 and Ser36; and Lys21 and Lys22 as the sites for subsequent ubiquitination by SCF(β -TrCP), respectively [62,65]). The SRD was always assumed to be unstructured or highly disordered based on the failed attempts to crystallize full-length $\text{I}\kappa\text{B}\alpha$ in complex with NF- κB . The instability of free $\text{I}\kappa\text{B}\alpha$ in solution and the absence of significant SRD contributions to the interaction energy of the protein-protein complex of $\text{I}\kappa\text{B}\alpha$ /NF- κB lead to the hypothesis of the SRD not being critical for this complex formation.

Detailed knowledge of the NF- κ B/I κ B α interaction comes from protein crystallography and high resolution NMR experiments [66]. However, these results do not include any structural information about the SRD (residues 1-72) of I κ B α . Previous investigations by molecular dynamics (MD) simulations of NF- κ B/I κ B α focused on the amide proton/deuterium exchange kinetics of four central ankyrin repeat units of co-crystallized I κ B α by accelerated molecular dynamics (aMD) simulations [65], a truncated free-I κ B α [67] and the structure of a free, doubly phosphorylated 24 amino acid peptide of the SRD [68].

In studies I and II of this thesis, I have explored the nature of secondary structure elements in the SRD of I κ B α in addition to studying the structural effects that arise as a response to phosphorylation on this ankyrin repeat protein.

Chapter D

Voltage-gated Ion Channels

D.1 The biological membrane

The cell membrane is the permeable structure that envelops a cell and acts as a protection barrier between the inside, the cytosol, and outside, the extracellular environment, deciding what enters and leaves the cell. The membrane consists of phospholipids, proteins, and carbohydrates. The phospholipids form the surfaces of the membrane bilayer with their polar head groups and their hydrophobic acyl chains tuck in between and shape the lipid bilayer. The proteins crowding the surface of membranes are peripheral proteins, whereas the proteins embedded into the membrane bilayer are known as integral membrane proteins, which act as channels and transporters for ions and hydrophilic substances. Membrane bilayers are highly dynamic and fluid structures, which enable interaction among proteins and between proteins and lipids. The permeability property of the membrane enables it to maintain charge and concentration gradients crucial to the function of the cell such as ATP synthesis, flow of solutes across the membrane, and in nerve and muscle cells produce and transmit electrical signals.

Because of difficulties with purification and crystallization of membrane proteins, the atomic structure of less than 10 % of known membrane proteins has been determined so far. The transmembrane (TM) peptide chains of integral membrane proteins are largely composed of nonpolar residues as these are the segments that are in contact with the hydrophobic environment of the lipid bilayer, interacting with the hydrophobic acyl groups of the membrane phospholipids and often take the secondary structure shape of α -helices. A common membrane protein in eukaryotic cells are ion channels; channels that are characterized by ion specificity and gating and respond to ligands or voltage across the

membrane bilayer. The characteristics of ion channels have been better understood in part because of the high-resolution structures of voltage-gated potassium channels [69].

D.2 The membrane potential

A neuron is an electrically excitable cell that processes and transmits information by electro-chemical signaling and is the principal element of the nervous system. Signals coming in from other neurons are normally received through the dendrites, and outgoing signals flow out through the axon of a neuron. The rapid communication between neurons is achieved through electrical signals. Communication then takes place at synapses through transmission of chemicals known as neurotransmitters.

It was in 1952 when the first quantitative description of the electrical events, forming the basis of action potential generation, was proposed by Alan Hodgkin and Andrew Huxley - using data from the squid giant axon, they proposed a model that accurately predicted the shape of the action potential [70]. When at rest, neurons generate a constant voltage across their membranes called the resting membrane potential that is typically between -40 to -90 mV. In response to stimuli, the resting membrane potential is changed and electrical signals are produced by the neurons. The electrical signals that are propagated through their axons are called action potentials [71]. The resting membrane potential is maintained via an unequal distribution of ions across the membrane that is created by the Na^+/K^+ - ATPase.

These ATP-dependent pumps exchange internal Na^+ for external K^+ , consequently leading to a concentration of K^+ ions inside the neuron and Na^+ ions outside. A high concentration of Na^+ on the outside of the neuron creates a concentration gradient, causing Na^+ to diffuse into the neuron creating a net positive electrical charge. The influx of Na^+ into the neuron makes the inside more positive - this is known as depolarization and leads to a reduction and consequently reversal of the transmembrane potential. As the electrical gradient across the membrane is reduced the influx of Na^+ ions to the neuron is diminished. When there is no flow of Na^+ ions inside the neuron, an equilibrated state of membrane potential, known as equilibrium potential, is reached. The equilibrium potential is deduced from the Nernst equation and for Na^+ and K^+ lies around +55 mV and -103 mV, respectively [72].

The action potential can be sectioned into three different stages. At the first stage, the depolarization stage, the membrane potential rapidly goes from -60 to +40 mV. In the second stage, the repolarization stage, the membrane potential returns to a resting potential. In the third stage, known as the after-hyperpolarization, the membrane potential slowly recovers from the resting potential [72]. The generation of electrical signals in neurons is conveyed through the concentration gradient maintained by ions. The membrane proteins that create and maintain ion gradients are known as active transporters, and the membrane proteins that selectively permeate ions across the membrane are called ion channels. The ion channels that are able to sense the electrical voltage across the membrane are called voltage-gated ion channels, which open and close in response to the membrane potential [71].

D.3 Voltage-gated potassium channels

Voltage-gated ion channels are the third largest group of signaling molecules encoded by the human genome, after protein kinases and G protein-coupled receptors [73]. K^+ channels constitute about half of this superfamily which has 78 members. The K^+ channel family is divided into four structural types depending on their mode of activation and the number of transmembrane segments: the inwardly rectifying 2-transmembrane K^+ channels (K_{ir}), 2-pore-4-transmembrane K^+ channels (K_{2P}), Ca^{2+} -activated 6-transmembrane/7-transmembrane K^+ channels (K_{Ca}), and the largest group in this family the voltage-gated 6-transmembrane K^+ channels (K_V) [74]. The K_V channel group is encoded by 40 genes in the human genome and is divided into 12 subfamilies, known as K_{V1} to K_{V12} [75]. K_V channels are normally expressed together with either or both voltage-gated Na^+ and Ca^{2+} (Ca_V) channels in excitable cells such as neurons or cardiac myocytes, and are responsible for repolarization after action potential firing, and both voltage-gated Na^+ and Ca^{2+} channels have been found to be homologous to the K_V channels [74].

It was in 1987 that the first structure of a voltage-gated K^+ channel, the *Drosophila Shaker* channel, was cloned and expressed [76]. Since then, X-ray crystallographic structures have revealed the atomic structures of several K^+ channels, shedding light into the basic mechanism of K^+ channel function. The *Streptomyces lividans* K^+ channel ($KcsA$) solved by MacKinnon has contributed greatly into mapping the structural details of a K^+ channel [77].

The K^+ channel is a tetrameric channel with four identical subunits shaping a central water-filled ion-conducting pore. Each subunit has six transmembrane α -helices, S_1 - S_6 , with both N- and C- termini on the intracellular side of the membrane. The selectivity filter, which is the formed by the narrowest part of the pore, is shaped by a loop connecting S_5 and S_6 . The pore domain (PD) is formed by S_5 , a pore helix (PH), and S_6 and helices S_1 - S_4 comprise the voltage sensor (VSD). All K^+ channels have the signature sequence, TMxTVGYG, between the two most carboxy-terminal transmembrane helices, the pore loop [78]. The selectivity filter has a diameter of 3 Å and is highly hydrophilic, made up of twelve backbone carbonyl groups, three from each of the subunits. K_V channels have another conserved domain, the tetramerization domain (T1) at the N-terminus [79,80], which determines the specificity of subunit assembly [81] in addition to acting as an anchoring point of potential β -subunits [79].

D.4 The voltage sensor and voltage sensing mechanism

Next to ion permeation as a main function, speed and specificity are two other important factors when K_V channels move ions through their pores. Ion permeation is regulated by the opening and closing of the pore, which lead to conformational changes known as gating. This gating is coupled to a sensing mechanism detecting changes in the membrane voltage. As depolarization prompts the opening of the channel and hyperpolarization induces its closing, the channel requires a voltage-sensor to detect the potential across the membrane.

The architecture of an activated (open or slow inactivated) chimaeric K_V channel called the paddle-chimera, in which the voltage sensor paddle has been transferred from

K_v2.1 to K_v1.2, and crystallized in complex with lipids at 2.4 Å resolution, has revealed the pore and voltage sensors embedded in lipids mimicking a membrane bilayer environment [82]. The VSD is composed of helices S₁-S₄ that are arranged in an anti-parallel fashion. The VSD is coupled to the S₆ gate through the S₄-S₅ linker helix and it is this coupling that enables channel opening and closing. The S₄ helix contains the gating charges, five positive residues mainly arginines thought to be involved in voltage gating of which the first four are suggested to be the most important – this marks S₄ as the voltage sensor [83,84]. The positively charged residues in S₄ interact with negatively counter charges in S₁-S₃ within the membrane, while the charged residues exposed to the membrane interact with the lipid head groups [82,85,86]. This leads to a neutral interior of the channel. A highly conserved phenylalanine in S₂ forms a barrier for the voltage sensor charges [87].

As no crystal structure on the closed state of voltage-gated K⁺ channels has yet been solved, the closing of the channel is explained by a few potential models, which describe the voltage sensor movement. At negative membrane voltages, S₄ is trapped in a down position closer to the intracellular side of the channel. At positive membrane potentials, S₄ is driven in an outward direction through the membrane towards the extracellular side of the channel. This concerted motion opens the inner gate of the pore allowing ions to pass through the channel.

Activation can be divided into two main parts: the independent outward movement of the four S₄ helices in a channel, followed by a concerted opening step when all four S₄ helices move together [88-92]. However, exactly how S₄ moves can be explained by three main theories, as reviewed by Börjesson and Elinder [93]. The first suggested model is the helical-screw or sliding helix, which proposes that the positive charges in S₄ make contact with negative counter charges in other parts of the channel as S₄ moves 4.5 Å and rotates 60° along its length axis [94,95]. The second model, the transporter-like, helical-twist, or rocking-banana model, suggest a large rotation but without little translational movement required to transfer charges from the intracellular to the extracellular side [96]. The third model is the paddle model, which assumes that the voltage sensor paddle, i.e. S₄ and S_{3b}, are in close contact and never cease this contact during gating [97].

D.5 Modulation of voltage sensor in the Shaker

Ion channel modulation can be achieved by a direct effect on the ion-conduction pore, such as local anesthetics and the neurotoxins tetrodotoxin from the pufferfish. Another way of channel modulation is by affecting the channel's voltage-dependence. By moving the voltage dependence of a depolarized activated channel in a positive direction along the voltage axis, the probability of an open channel is decreased, however if you move it in a negative direction the probability of an open channel is increased [93]. There are a number of mechanisms that can induce this shift.

Modulation can occur from the intramembrane side and directly affect the voltage sensor and the gate. The membrane phospholipids for example interact with the channel by being in direct contact with S₄ [98], or the phospholipid phosphatidylinositol-4,5-bisphosphate (PIP₂) that is located in the cytoplasmic leaflet of the membrane and acts electrostatically on the cytoplasmic parts of the channel mediating its effect [99].

The voltage dependence of ion channels can also be modified from the extracellular side through a direct effect on the voltage sensor. For example there are toxins that bind to the extracellular VSD and trap the voltage sensor either in a resting or an activated state [100,101]. The tarantula toxin, hanatoxin, has been found to bind to the S₃-S₄ linker of K_V channels and by preventing the movement of S₄ shift the voltage dependence towards more positive voltages [101,102]. Another type of channel modifier are free fatty acids, charged lipophilic molecules that interact directly or electrostatically with the channel by inserting into the membrane near the channel or into hydrophobic cavities on the channel itself, such as the polyunsaturated fatty acids (PUFAs) [103-108].

D.6 Disease and dysfunctional K_V channels

K_V channels are important therapeutic drug targets because of their implications in diverse diseases ranging from cancer to autoimmune diseases to metabolic, neurological and cardiovascular disorders [74]. In specific, the dysfunction of K_V channels has been linked to epilepsy, a brain disorder that is characterized by recurrent seizures caused by a synchronous firing of neurons [109]. When antiepileptic drugs do not respond in about 25-30 % of children with epilepsy [110], a ketogenic diet has proven to be an effective treatment [110-112]. The ketogenic diet is high in fat, ample protein and low amounts of carbohydrates.

The mechanisms by which this diet has an effect are not clear, but a proposed mechanism is the direct inhibitory effect of PUFAs [113]. PUFAs are vital for normal brain development and function [114] and have been found in higher concentration in the blood serum [115] and brain [116] following a ketogenic diet treatment. While PUFAs are important regulators of neuronal excitability by modulating sodium and calcium currents [117,118], less is known about their effects on K_V channels. It has been reported that epilepsy is caused by mutations in the K_{V1}-type and KQT-type K_V channels leading to an increase in K⁺ current [119-124]. The molecular mechanism of how PUFAs exert their effects on K_V channels has been the subject of debate.

D.7 PUFA mediated modulation

PUFAs have a lipophilic acyl tail with two or more double bonds and a negatively charged carboxyl head group. Different methods of PUFA modulation have been proposed. PUFAs binding directly to the channel or PUFAs affecting the channel by inducing changes to the membrane properties such as fluidity have been suggested [104,105,107,125-127].

PUFAs have been suggested to electrostatically affect K_V channels by binding to a hydrophobic environment. The charge of the PUFA head group has proved necessary in the shift of the voltage dependence of the channel. Also acyl chain properties such as the requirement of at least two double bonds in the *cis* geometry are needed for the PUFA effect on K_V channels [128]. The site of PUFA modulation in K_V channels has been proposed to be close to the lipid facing side of S₃ and S₄ helices on the extracellular side of the channel. Also, PUFAs are thought to act on the final voltage sensor transition and with this effect relying on certain amino acid residues on the extracellular side of S₄ [129].

Studies III and IV of this thesis have been dedicated to study the interactions between PUFAs and the *Shaker* K_V channel. I have focused on finding a PUFA binding site, characterize the nature of these interactions, and the effects that PUFAs infer on the energetics of the channel in going from an open to an intermediate closed state.

Chapter E

Methodology

E-1 Molecular modeling

Experimentally obtained structures of biological macromolecules is large and ever increasing. The three-dimensional biological structures of all proteins and other large molecules determined from all experimental techniques are stored online in the Research Collaboratory for Structural Bioinformatics' repository the Protein Data Bank (PDB).

A predominant and well-established experimental technique to decipher protein structure is X-ray crystallography. In this method, the molecule is bombarded with X-rays, which diffract into many specific directions. From the angles and intensities of the diffracted beams a three-dimensional description of the density of the electrons are produced. Through this density, the positions of atoms and the chemical bonds that connect them are determined. What are needed are well-ordered crystals and the right crystallization conditions, which can be difficult for membrane proteins especially. A low resolution crystallographic protein structure only indicates the overall shape of a protein, while a higher resolution ($> 2 \text{ \AA}$) structure reveals the atomic positions of the protein. A crystallographic structure only represents an average over all the molecules in the crystal so there is always uncertainty of the quality of the structure due to the dynamic nature of proteins. Another method that is widely used for structure determination is nuclear magnetic resonance (NMR). In NMR, the magnetic spin properties of atomic nuclei are used to set up a list of distance constraints between atoms from which the three-dimensional structure of the protein can be resolved. This method is restricted to the determination of the structure of smaller proteins. And as opposed to X-ray crystallography that requires crystals, NMR structures are obtained from concentrated protein solutions [130].

Still little or no structural information is available for many proteins. Consequently, modeling which is the process of generating a so called 'model' or an idealized description of a system has provided valuable when structure determination has been beyond the reach of experiments. Therefore methods for predicting tertiary and secondary structure from amino acid sequence have been called for. The *ab initio* method attempts to solve structure prediction from first principles, however, one needs to explore the entire conformational space of the molecule in order to find the best possible structure. Finding a structure with the lowest energy among an infinite number of conformations has been next to hopeless in the present time. In search for more attainable prediction methods, a rather common and reliable technique for modeling is comparative modeling that relies on detecting evolutionary similarity between the sequences of an unknown and known protein structures. Besides the comparative modeling approach, in particular with cases where low sequence similarity hinders detection of related homologues, protein structures can be predicted using threading [131].

E-1-1 Comparative modeling and threading

The modeling procedure requires a series of consecutive steps usually repeated iteratively until a reasonable model is obtained: finding and selecting one or more suitable template proteins related to the target; alignment of the target to the template sequences; building an initial model of the target based on the three-dimensional structure of the templates; *ab initio* modeling of side chains and loops in the model that differ from the template; refining and evaluating the final model.

Comparative modeling is based on the general understanding that evolutionary related sequences have similar three-dimensional structures. To identify a statistically significant relationship between the target and templates a sequence alignment method is applied. The search for related sequences is done normally by PSI-BLAST [132]. It creates a multiple sequence alignment by an iterative process, where each time a new sequence is identified it is included in the next query to the sequence database. It is this process which enables the detection of distantly related sequences. A further improvement in the search of related sequences is to prepare sequence profiles for all known structures and do a pairwise comparison; another way to find conserved motifs among sequences has been the implementation of profile-based Hidden Markov Models (HMM) improved with the incorporation of predicted secondary structure elements [133,134].

When the sequence similarity between proteins is more than 70 % it is easier to determine an alignment that leads to a reliable model. However, as this sequence similarity decreases it becomes increasingly harder to fully trust the model and as the sequence similarity falls below 20-30 % one enters the so called *twilight zone*. At this stage, threading can be a suitable option. Threading is based on the pairwise comparison of a protein sequence and a protein structure, where the target sequence is matched to a library of three-dimensional profiles or threaded through a library of three-dimensional folds [135].

After finding related sequences, selecting the optimal templates is the next important step. As a matter of fact, in selecting templates, the use of several templates can improve the quality of the model. By combining multiple templates, different domains of the target sequence can be aligned to the different templates with little or no overlap in between allowing the generation of a model for the whole target sequence. Also, by

aligning the entire target sequence to all templates, a model can be improved by local selection of the best template.

In essence, when constructing of a three-dimensional model first the amino acid backbone for the structurally conserved regions is generated to which loops are added followed by the placement of side chains. This model then undergoes refinement by being subjected to energy minimization. The procedure that Modeller [136] adopts in constructing the model is deriving a large number of restrains. These constraints are normally distances between alpha-carbon atoms, residue solvent accessibilities or side chain torsion angles.

E.2 Molecular dynamics simulations

Proteins are inherently flexible molecules and can experience complex internal motions. These motions include larger global deformations of the entire protein induced by contact with its surrounding environment whether it is the solvent or the lipid bilayer or the folding and unfolding of a protein or the transitioning from one state to another. But these motions can also be local caused by internal movement triggered by constant atomic motions or side chain fluctuations. The function of proteins can be coupled to the dynamics of these systems, in which structural fluctuations are closely connected to the activity of the molecule.

Simulations have played a crucial role in studying the dynamics of these macromolecules. The properties of a system can be explained by the detailed description of particle motions as a function of time that is obtained by molecular dynamics simulations. Quantum mechanics methods give the most accurate results when studying the dynamics of a system; in quantum mechanics electrons are treated explicitly and the energy of the system is calculated by solving the Schrödinger equation, however, these methods are limited to small scale systems usually involving a few atoms only and to much shorter time scales, normally within picoseconds. Methods based on molecular mechanics (MM) on the other hand permit the study of much larger systems involving millions of atoms up to microsecond time scales normally. Molecular mechanics covers a range of simulation techniques, but MD is one of the most common techniques applied when studying biological macromolecules.

E.2.1 Theory

The first molecular dynamics simulation of a biologically relevant macromolecule was carried out by McCammon *et al.* [137] in 1977. Performed in vacuum, the bovine pancreatic trypsin inhibitor (BPTI) was simulated for 9.2 ps and revealed the flexible nature of proteins as opposed to the relatively rigid structures they had been thought of. In molecular dynamics, atomic motion is simulated by solving Newton's equation of motion simultaneously for all atoms in the system. In classical molecular dynamics, electrons are not treated explicitly, but rather electronic interactions are estimated and taken into account implicitly. In molecular dynamics, forces between atoms are calculated using an empirically derived force field, which describe the interaction between different types of atoms using a set of mathematical functions and parameters.

In molecular dynamics, the calculations of a system consisting of N particles are split into a series of very small time steps, typically between 1 femtosecond and 10 femtoseconds. At every step and for each atom i , forces on the atom with mass m is calculated and combined with the positions and velocities to determine new positions and velocities by directly solving Newton's second law of motion:

$$\vec{F}_i = m_i \cdot \vec{a}_i = m_i \cdot \frac{d^2 \vec{r}_i}{dt^2} \quad (\text{E.1})$$

The force on atom i is calculated by differentiating the potential energy function V with respect to each atom's position:

$$\vec{F}_i = -\nabla_i V(\vec{r}_1, \dots, \vec{r}_N) \quad (\text{E.2})$$

Finite difference methods are used to integrate the equations of motion. The basic idea is that integration is divided into many small steps each separated by a fixed time. Integration of Equation E.2 given the position of atom i at time $\vec{r}_i(t)$, gives the positions and velocities at a later time step, $\vec{r}_i(t + \Delta t)$, approximated by a standard Taylor series expansion:

$$\vec{r}_i(t + \Delta t) = \vec{r}_i(t) + \frac{d\vec{r}_i(t)}{dt} \Delta t + \frac{d^2 \vec{r}_i(t)}{dt^2} \frac{\Delta t^2}{2} + \dots \quad (\text{E.3})$$

There are many algorithms that exist for integrating the equations of motion using the Taylor series expansions. A commonly used integration algorithm in molecular dynamics simulations is the Verlet algorithm [138]. The Verlet algorithm uses the positions $\vec{r}_i(t)$ and accelerations $\vec{a}_i(t)$ at time i and the positions from the previous step $\vec{r}_i(t - \Delta t)$ to calculate new positions:

$$\vec{r}_i(t + \Delta t) = \vec{r}_i(t) + \vec{v}_i(t) \Delta t + \frac{1}{2} \vec{a}_i(t) \Delta t^2 + \dots \quad (\text{E.4})$$

$$\vec{r}_i(t - \Delta t) = \vec{r}_i(t) - \vec{v}_i(t) \Delta t + \frac{1}{2} \vec{a}_i(t) \Delta t^2 - \dots \quad (\text{E.5})$$

The addition of the above two equations gives:

$$\vec{r}_i(t + \Delta t) = 2\vec{r}_i(t) - \vec{r}_i(t - \Delta t) + \frac{\vec{F}_i(t)}{m_i} \Delta t^2 \quad (\text{E.6})$$

The acceleration expressions as part of Equations E.4 and E.5 have been substituted with the force and the mass of the particle in the additive Equation E.6 as stated by Equation E.1. The Verlet algorithm is simple to implement, accurate and has been proven to be stable. However, one disadvantage with the Verlet algorithm is that velocities are not directly generated. Therefore, many variations on the Verlet algorithm have been

developed. One such alternative is the leap-frog algorithm [139]. The leap-frog algorithm calculates velocities at half time steps:

$$\vec{v}_i\left(t + \frac{1}{2}\Delta t\right) = \vec{v}_i\left(t - \frac{1}{2}\Delta t\right) + \vec{a}_i(t)\Delta t \quad (\text{E.7})$$

By evaluating this equation, velocities at time i can be calculated from:

$$\vec{v}_i(t) = \frac{1}{2}\left[\vec{v}_i\left(t + \frac{1}{2}\Delta t\right) + \vec{v}_i\left(t - \frac{1}{2}\Delta t\right)\right] \quad (\text{E.8})$$

The leap-frog algorithm results in greater numerical accuracy as it does not require the calculation of subtraction of large numbers. The most challenging task in molecular dynamics simulation is calculating the force on each particle in the system. This therefore calls for an integration algorithm that conserves energy and momentum, is time-reversible, and allows for a longer time step, Δt , to be used. With longer time steps, fewer iterations are needed to sample a configurational space and obtain a trajectory of the motions of every atom in the system. The software package used throughout this work for performing molecular dynamics simulations in an efficient manner is GROMACS [140]. Other commonly used packaged are NAMD [141] and CHARMM [142].

E.2.2 Force fields

At the heart of a classical molecular dynamics simulation lies the force field, which is a description of the potential energy of a system relating the interactions between the particles in that system. The main components of a force field consist of an energy function together with the parameters as part of the function. Many force fields embody distinct components in the system; the bonded interactions mediated by bonds, angles, and torsions and non-bonded interactions related by van der Waals and electrostatics:

$$V_{TOTAL} = \sum_{BONDS} k_b (b - b_0)^2 + \sum_{ANGLES} k_\theta (\theta - \theta_0)^2 + \sum_{TORSIONS} k_\omega (1 + \cos(n\omega + \delta)) + \sum_{vdW} \sum_{COULOMB} \left(\epsilon_{ij} \left[\left(\frac{R_{min,ij}}{r_{ij}} \right)^{12} - 2 \left(\left(\frac{R_{min,ij}}{r_{ij}} \right)^6 \right) \right] + \frac{q_i q_j}{4\pi\epsilon_0\epsilon_1 r_{ij}} \right) \quad (\text{E.9})$$

The first term in the Equation E.9 models the covalent bond interaction modeled by a harmonic potential that gives the increase in energy as the bond length b diverges from the reference bond length, b_0 . The second term, also modeled by a harmonic potential, sums the angles where θ is the angle formed between atoms $A - B - C$. The third term is the energy associated with torsional rotation around the middle bond of atoms $A - B - C - D$. The rotational angle is determined by ω and k_ω sets the energy maximum. The fourth term accounts for the non-bonded contributions in the system. Calculated for all pairs of atoms i and j , the non-bonded interaction is calculated for atoms that are separated by at least three bonds in the same molecule or are in different molecules. In its simple form, this term

is modeled by a Lennard-Jones potential for the van der Waals interactions and a Coulomb potential for the electrostatic interactions. The Lennard-Jones is characterized by an attractive and a repulsive part. The attractive contribution is due to the dispersive force. Also known as the London force, instantaneous dipoles form due to fluctuations in the electron clouds and they in turn trigger the dipole formation in neighboring atoms causing the attraction. The repulsive contribution is due to short-range repulsive forces also known as the overlap forces that restrict electrons from occupying the same inter-nuclear region. The energy minimum is specified by ϵ and is located between atoms at a distance R_{min} . The Coulombs potential models the electrostatic potential between particles, where q_i and q_j are the charges, ϵ_0 is the vacuum permittivity and ϵ_1 the relative permittivity of the environment.

The summation of all these terms gives the total potential energy V . Normally high amounts of energy are required for bonds stretching and angles bending to deviate from their reference values. Indeed, it is energies concerned with torsional and non-bonded interactions that mainly contribute to the divergence in structure and relative energies. The performance of a force field relies upon the quality of its parameters. Parameterization is a complicated process where values for the parameters are obtained by experimental data and quantum mechanics calculations. The optimization of the parameters relating to the torsional and non-bonded interactions is in particular important as these parameters mainly affect the force field.

E.3 Free energy calculations

The expression *free energy calculations* refers to a set of simulations that calculate the free energy difference between different states of biological systems. Calculating this free energy difference has many applications, for example by calculating the free energy of receptor-ligand binding, one can obtain the dissociation constant of the receptor.

Free energy by definition is a thermodynamic quantity that is equivalent to the amount of energy accessible in a system to do work. It describes the stability of a system since the free energy of a system is minimized only when the system is at equilibrium. The free energy is either expressed as the Helmholtz function, A , or the Gibbs functions, G . Systems that have a constant number of particles, temperature and volume are suitable for the Helmholtz free energy. Most molecular dynamics experiments are however performed under conditions of constant number of particles, temperature and pressure, making the Gibbs free energy more suitable.

Gibbs free energy, also known as ΔG , allows the prediction of the spontaneity of a chemical reaction. If $\Delta G < 0$, then the reaction is spontaneous, whereas if $\Delta G > 0$, the reaction is nonspontaneous. Gibbs free energy combines enthalpy (H) and entropy (S) and is defined as:

$$\Delta G = H - TS \quad (\text{E.10})$$

where T is the absolute temperature. The Helmholtz function, which gives the free energy of a system at constant number of particles, volume and temperature, is given by:

$$A = -k_B T \ln Q \quad (\text{E.11})$$

where Q is the partition function of the system and k_B is Boltzmann's constant. Free energies cannot be accurately calculated from a molecular dynamics simulation since these simulations do not sample the regions from phase space that play an important role in the free energy of the system. Molecular dynamics simulations search for the lower energy phase space and since they do not sample the high energy regions of the phase space, calculating energies with the traditional simulation method end up in poor results of free energies. Conventional free energy calculation methods use molecular simulations to generate independent samples from the equilibrium distribution of the system. These samples are consequently analyzed with statistical mechanics methods that give estimates of free energy differences.

E.3.1 Sampling along a reaction coordinate

When studying the behavior of a system in specific states or along particular paths, one is often interested in calculating the free energy along a reaction coordinate. The reaction coordinate could be the distance between two molecules, a torsion angle along the protein backbone, or the relative orientation of a helix with respect to the membrane. The free energy or potential of mean force (PMF) [143] along a reaction coordinate is given by:

$$A(\xi) = -k_B T \ln P(\xi) \quad (\text{E.12})$$

where $P(\xi)$ is the probability distribution of the reaction coordinate ξ . To sample important values of the reaction coordinate, methods such as umbrella sampling [144] can be used to access regions of low probability in the conformation space. In umbrella sampling, the interval along the reaction coordinate is split into subintervals, called windows, in which a biasing potential can be used to improve sampling.

E.3.2 Accelerated weight histogram

A drawback with umbrella sampling for example is that this method only samples one pathway at a time except if the sampling is carried out multiple times. In a transition, however, there could be many different pathways involved and as such, efficient sampling is vital in mapping out an accurate free energy landscape. Here is where a recently developed approach, the accelerated weight histogram (AWH) method [145] is advantageous. The AWH method is an extended ensemble method that uses an adaptively biasing approach. A simulation in extended ensemble performs a random walk in the energy space and overcomes the difficulty of getting stuck in local energy minima. By using a Gibbs sampler and a probability weight histogram the AWH method enables larger transitions in parameter space. The use of the weight histogram allows for efficient utilization of the transition history in order parameter space in the free energy updates by

adapting the bias as well in keeping a track of the simulation. Also, by using a weight histogram the need of discretization of the parameter space is removed. AWH estimates the free energy by:

$$g(\lambda) = f(\lambda) + \ln \rho(\lambda) \quad (\text{E.13})$$

where $g(\lambda)$ is the biasing function, $f(\lambda)$ the estimate of the true free energy and $\rho(\lambda)$ the target distribution. During the simulation, the biasing function is adjusted to fit a specified target distribution. To begin with, $f(\lambda)$ is initialized but throughout the sampling it is iteratively refined. The main benefit of the AWH method lies with the iteration of the process of sampling along the reaction coordinate, permitting the sampling of different pathways.

Chapter F

Study I

F.1 Structural modeling of the N-terminal signal-receiving domain of I κ B α

A detailed understanding of molecular mechanism of NF- κ B activation, regulation and the protein-protein communication with partners may assist to design and develop novel chronic inflammation modulators as well as anti-cancer drugs. The insight gained from structural biology of NF- κ B and I κ B α proteins and its implications for the signaling process control have been reviewed extensively by *i.e.* [146-148].

From the structure of I κ B α in complex with NF- κ B, a valuable level of insight was rendered into the regulation of NF- κ B by I κ B α and the nature of their association [61,149]. Each unit of the complex was partially truncated leading to a missing I κ B α N-terminal segment comprising ~70 residues. This N-terminal SRD receives the phosphorylation and ubiquitination signals and targets the protein to the proteasome for degradation [150], but has no measureable effect on binding of I κ B α to NF- κ B [151]. While SRD plays a crucial part in activation of NF- κ B, it has not been found to be engaged in enabling the complex formation of I κ B α /NF- κ B [152-154]. Protein crystallization and structure determination were unsuccessful for free I κ B α due to its short lifetime and degradation within minutes. This led to the suggestion of conformational disorder in the free protein [65]. For I κ B α in complex with NF- κ B, however, there are two protein crystal structures available (PDB IDs 1IKN and 1NFI). The truncated I κ B α sequences in 1IKN (residues 73-292) [61] and 1NFI (residues 71-280) [149], however, did not reveal information about possible secondary structure elements (SSEs) in the SRD.

The concept of conservation of secondary structure elements [155] in families can be used to identify proteins only distantly related in sequence, which may, however, still share a higher degree of conservation of SSEs. Recent approaches have demonstrated that the use of multiple tools of secondary structure prediction and the use of a 'consensus' of methods yields more reliable results than single algorithms [59,60].

In this study, I present for the first time, the structural elements of the full length SRD of I κ B α in complex with NF- κ B and in free I κ B α . I clearly show that the SRD displays well-defined secondary structure elements and cannot be considered as 'unstructured'. In contrast, it contains 3 α -helical regions which are stable during molecular dynamics simulations. Also, in free I κ B α the SRD is structured albeit displaying a larger degree of flexibility and larger fluctuations. This represents the first step in an approach to model the signal transduction cascade of the NF- κ B/I κ B α complex from IKK phosphorylation to degradation.

F.2 Methods

F.2.1 Structural modeling

The secondary structure prediction of the full I κ B α sequence was performed with SYMPRED [156] which builds upon results from PROF [157], SPRO [158], YASPIN [159] and PSIPRED [160]. In addition, JPRED3 [161], JUFO [161], NetSurfP [162], PORTER [163], PredictProtein [164] and ScratchProteinPredictor [165] were also used. All secondary structure prediction algorithms correctly identified and positioned the six ankyrin repeat units in the crystal structure (PDB ID 1IKN) in addition to four additional α -helical regions in the N-terminal SRD, which is not resolved in the protein crystal structure. The consensus of predicted secondary structure elements was used for structural modeling of the SRD.

To identify a suitable structural template for modeling the SRD of I κ B α , pDomThreader [166] was used; a profile based recognition fold method incorporating domain superfamily discrimination, which distinguished 46 probable structural templates. The fourth ranked template, 1N11, was chosen as a suitable template basing the decision on a top alignment score, the degree of coverage and the structural alignment of 286 out of 317 residues in I κ B α . pDomTHREADER [166] makes use of the CATH database of annotation of protein structural superfamilies from the PDB [167]. It is an implementation of GenTHREADER, a method which predicts protein fold from sequence by integrating profile-profile alignments, secondary structure gap penalties and both classic pair and solvation potentials employing an optimized regression SVM model. pDomTHREADER is thus able to discriminate between different structural superfamilies from the protein sequences and to detect distant homology to proteins of known structure.

The structural model was generated with Prime [168,169]. A manually constructed sequence alignment of our templates 1IKN and 1N11 was used. A two-template composite model was thereby constructed; residues 73-292 were based on the crystallized I κ B α protein 1IKN while the secondary structure elements of ankyrin protein 1N11 served as basis for residues 1-98. The Build process involves coordination of the copying of the backbone atoms for aligned regions and side chains of conserved residues, building

insertions and deletions in the alignment, optimization of side chains not found in template and energy minimization of those residues not derived from the templates. The Prime Build process applies the OPLS_2005 all-atom force field for energy scoring and the Surface Generalized Born (SGB) continuum solvation model for treating solvation energies and effects. Additionally it utilizes the residue-specific side-chain rotamer and backbone dihedral libraries, derived from the non-redundant data sets extracted from the PDB.

F.2.2 System assembly and protocol for MD simulations

The MD simulations were carried out using Gromacs 4.5 [170,171] employing the GROMOS96 43a1 force field [172]. The all-atom structural model of I κ B α bound to the X-ray crystallographic structure of NF- κ B included 6945 atoms in total. The protein complex was immersed in a rectangular box of dimensions 78 x 89 x 145 Å³ solvated with 29686 SPC water molecules together with 117 Na⁺ and 91 Cl⁻ ions in order to neutralize the net system charge. The structural model of the free I κ B α was immersed in a slightly smaller rectangular box of dimensions 66 x 68 x 114 Å³ solvated with 15757 SPC water molecules together with 70 Na⁺ and 47 Cl⁻ ions, in total containing 50270 atoms. The LINCS algorithm [173] was applied for constraining bond lengths. Electrostatic interactions were calculated every step with the Particle-Mesh Ewald algorithm [174]. Neighbor lists were saved and reused for 5 steps. The simulations were performed at constant pressure of 1.0 bar with Parrinello-Rahman pressure coupling and the isotropic pressure scaling, time constant of 1.0 ps, and a system compressibility of 4.5e-5 bar⁻¹. The temperature of the system was coupled to 300 K using the velocity-rescaling algorithm with a time constant of 0.1 ps. Newton's equations of motion were integrated using the leap-frog algorithm with a 2 fs time step.

The solvated system was first minimized with the steepest descent algorithm until a maximum force of < 100.0 kJ/mol was reached. Equilibration of the system was initiated by 10000 steps of position-restrained MD by relaxing the solvent and keeping the non-hydrogen atoms of the system fixed. With the system relatively free of strain an NVT equilibration phase followed by an NPT phase of 10000 steps each was then carried out. Coordinates were saved every 2 ps for analysis and the production phase of the simulation ran for a total of 200 ns. Fan and Mark have shown that molecular dynamics simulation in explicit water are able to refine homology-based protein structures within a short period of simulation [175]. For small to medium-sized proteins (50-100 amino acids), the first 1-5 ns were able to remove initial distortions and only in few cases simulations of > 100 ns were necessary to obtain a significant reduction of RMSD. This was taken as a lower threshold and added a factor of two considering the complexity of the system. Three independent replicates of the system were simulated for 600 ns in total, each starting with different initial velocities. Simulating independent replicates is a rather cost-effective way to sample conformational space [176].

F.3 Results and Discussion

F.3.1 Structural elements in the signal-receiving domain (SRD)

In order to better understand the effects of phosphorylation and the mechanisms, which govern recognition of phosphorylated I κ B α , and consequentially initiate ubiquitination, one requires a structural model of the complete N-terminal protein SRD. A BLASTp search of the first 72 residues of I κ B α did not yield any significant sequence similarity with other known proteins.

A consensus-based secondary structure annotation of the full length I κ B α sequence with SYMPRED [156] was performed which builds upon results from PROF [157], SSPRO [158], YASPIN [159] and PSIPRED [160] (Figure I.1).

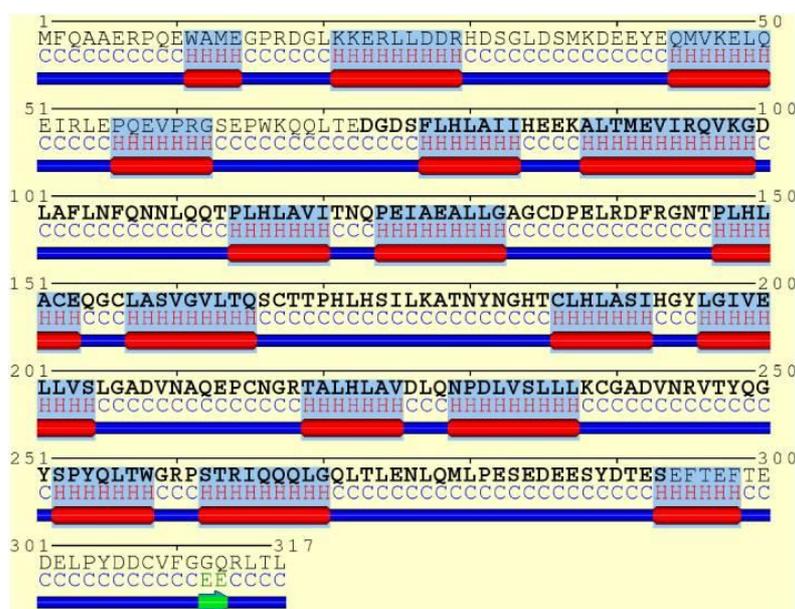


Figure I.1 Consensus secondary structure annotation of full length I κ B α (residues 1–317). Truncated, as crystallized I κ B α from 1KN (residues 73–293), is shown in bold letters. The six ankyrin repeat units of the ARD are recovered, correctly annotated and positioned. Two additional α -helix-loop- α -helix regions were detected in the N-terminal SRD. The C-terminal PEST domain displays less structural features.

In addition, JPRED₃ [161], JUFO [161], NetSurfP [162], PORTER [163], PredictProtein [164] and ScratchProteinPredictor [165] were also employed and give close to identical results.

All six ankyrin repeat units in the crystal structure (PDB ID: 1KN [61]) are recovered, correctly annotated and positioned. In addition four α -helical regions were detected in the N-terminal SRD, which is not resolved in the protein crystal structure (Figure I.1). This indicates that the SRD region may contain secondary structured subregions with a high α -helical content (residues 11-14, 21-29, 44-50, 56-62) not covered in any of the available I κ B α crystal structures and not investigated in any of the NMR studies of free or complexed I κ B α . The position of these α -helices is not fixed with respect to each other and may obstruct protein crystallization of full length I κ B α .

The initial finding prompted the generation of a full-length I κ B α model including the SRD and the investigation of its spatial and temporal integrity and stability. Due to the absence of sequence similarity of the SRD region to any structurally resolved protein in the PDB (~12 %) sequence-based comparative modeling is not a feasible approach here. As an alternative, the choice of template was based on identification of a remote structurally

related protein template with a similar secondary structure fold. The conservation of SSEs in protein superfamilies can guide the design of a structural model. Even when the structure of only a single member of a superfamily is known the conservation of SSEs can be used to predict the structure of other superfamily members [177,178]. Such information is useful when modeling the structure of other members of a superfamily or identifying structurally and functionally important positions in the fold. An efficient template detection allows the structural modeling to be extended even in the twilight zone of 10–30 % sequence identity [178].

pDomTHREADER [166] identifies 46 possible structural templates with reliable secondary structural similarity. Based on a top alignment score, the degree of coverage and the structural alignment of 286 out of 317 residues in I κ B α , one of the top ranked structures (1N11) was chosen as a template for modeling I κ B α .

As an alternative approach, a combination between comparative modeling and *de novo* protein structure prediction was performed using Robetta. For proteins with detected PDB homologs, comparative models are built based on templates that are found and aligned with incorporated versions of HHSEARCH/HHpred, RaptorX, and Sparks-X. Protein domains with no close PDB homologs are generated with the Rosetta *de novo* protocol [179,180]. A structure prediction carried out by Robetta [181] for the full I κ B α sequence also yielded 1N11 as the top-ranked template of choice for the generation of its structural models. Figure I.2 shows the alignment of secondary structure elements of I κ B α and 1N11 in the SRD region.

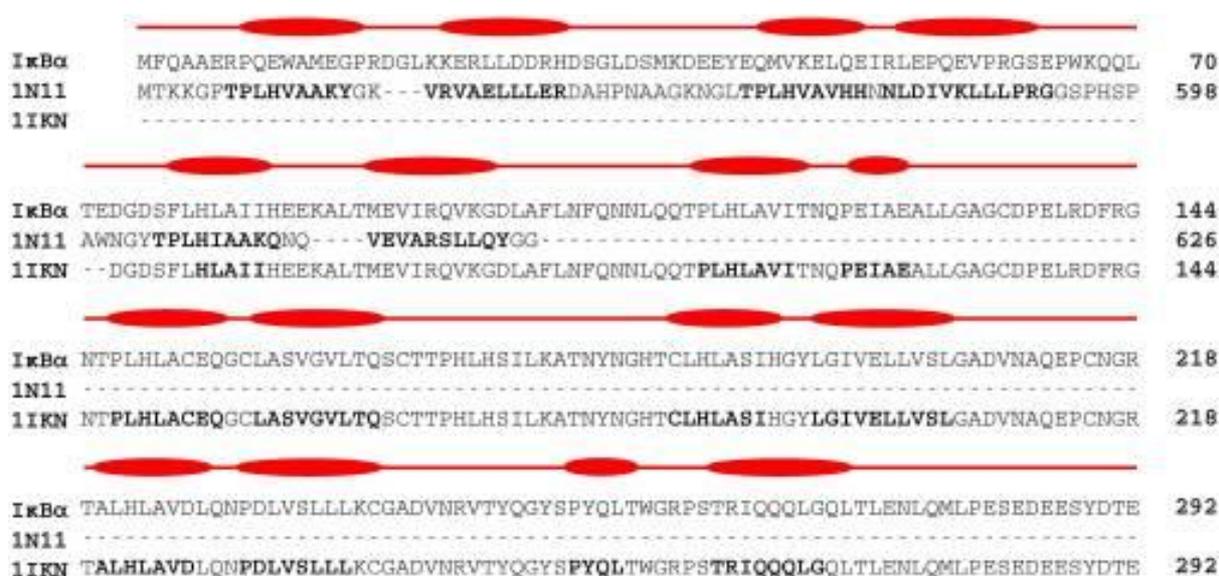


Figure I.2 Two-template sequence alignment used for the generation of a composite structural model of the full-sequence I κ B α . The bold segments in each template correspond to the α -helical regions forming the ankyrin repeat units present in the crystal structures of 1IKN and 1N11. The curved boxes in red display the helical segments in our generated structural model of I κ B α .

Despite an overall low primary sequence identity of only 23 %, the alignment of secondary structural elements is striking. 1N11 is the crystal structure of a 12 ankyrin repeat units stack from the human ankyrinR. AnkyrinR belongs to a family of adaptor proteins that

mediate anchoring between integral membrane proteins and the spectrin-actin cytoskeleton. The membrane-binding domain of ankyrins contains 24 ankyrin repeats of which the crystal structure of the human ankyrinR maps the D₃₄ region. This region, which consists of repeats 13-24, is stacked contiguously in the shape of a left-handed superhelix [182].

A composite model from crystallized I κ B α (67-317) 1IKN and ankyrinR 1N11 PDB structures was generated. Residues at positions 73-292 were taken from the crystallized I κ B α protein (PDB ID: 1IKN) and for residues 1-98 SSEs of the SRD were taken from the X-ray structure 1N11. For an overlapping stretch of residues 73-98, two α -helices forming one ankyrin repeat in the 1N11 template was taken to remove any possible artifacts from truncated sequence crystallization.

F.3.2 Structural refinement by molecular dynamics simulations

The protein-protein complex model was used as a starting configuration for subsequent MD refinement. The stability of the suggested secondary structural elements in the SRD and the dynamics of possible rearrangements were investigated.

In order to achieve a reliable full-sequence structural model, three independent MD simulations of I κ B α in complex with NF- κ B were performed for 200 ns each in a neutralized solvent box of about 30000 explicit water molecules. Thus, a total production simulation time of 600 ns was achieved. After energy minimization, a stepwise relaxation of the simulation setup and careful equilibration first in an NVT and then in an NPT ensemble, the general behavior of all simulation runs reveals well-behaved and stable systems. This is reflected in the conservation of total energy and temperature of the entire system, which is kept at a constant room temperature of 300 K throughout the whole 200 ns simulation runs.

The structural stability of the I κ B α /NF- κ B complex is also monitored by calculating the root mean square displacement (RMSD) from the starting protein-protein complex structure. The RMSD increases sharply to 3.5-4.5 Å for the three replicate runs during the first 100 ns of the simulations, and settles at roughly 4.5-5.5 Å for the last 100 ns, indicating a well-structured stable complex. The results discussed herewith are the average findings of the three replicate runs unless otherwise stated.

In order to investigate the secondary structural profile of the initial I κ B α model and possible structural re-arrangements, the 200 ns simulation was sectioned into two equal parts. This provides a comparison of results at the beginning and end of the production run periods.

To better understand the inherent flexibility of our protein, the root mean square fluctuations (RMSF) of the backbone C α atoms of I κ B α around the average structure were calculated (Figure I.3 A). The SRD N-terminal segment comprising ~75 residues clearly stands out as the most flexible region, particularly in the initial 100 ns of the simulations. Although not as expressed as in the initial 100 ns of simulations, the flexibility in the subsequent 100 ns run region is still comparatively high. One encounters the two most flexible helical regions of the whole protein, namely helices 1 and 2, also in this region. This result indeed explains the difficulty to crystallize the SRD region. Instead an N-terminal truncation of I κ B α was necessary to obtain protein single crystals [65]. One sees, in general, the retention of all crystallographically resolved six ankyrin repeat units in I κ B α during the entire simulation runs (Figure I.3 A). While the peaks mark the hairpin loop segments

connecting the α -helices in each ankyrin repeat unit, the troughs of the RMSF plot correspond to helical regions. This result shows that while the helical regions are stable and not so flexible, greater flexibility is observed in the β -loop segments. This is in agreement with the amide $^1\text{H}/^2\text{H}$ exchange experiments followed by MALDI-TOF mass spectrometry (MS) in bound and free $\text{I}\kappa\text{B}\alpha$ [63]. The β -hairpins of some ankyrin repeats readily exchange amide protons for deuterons (1^{st} , 5^{th} and 6^{th} ankyrin units) whereas other units (2 – 4) are less solvent accessible. In particular, ankyrin repeat unit 1 remains highly solvent accessible even in the complex. The solvent accessibility of the β -hairpin in ankyrin repeat unit 1 (AR 1) decreases slightly upon NF- κB binding [64].

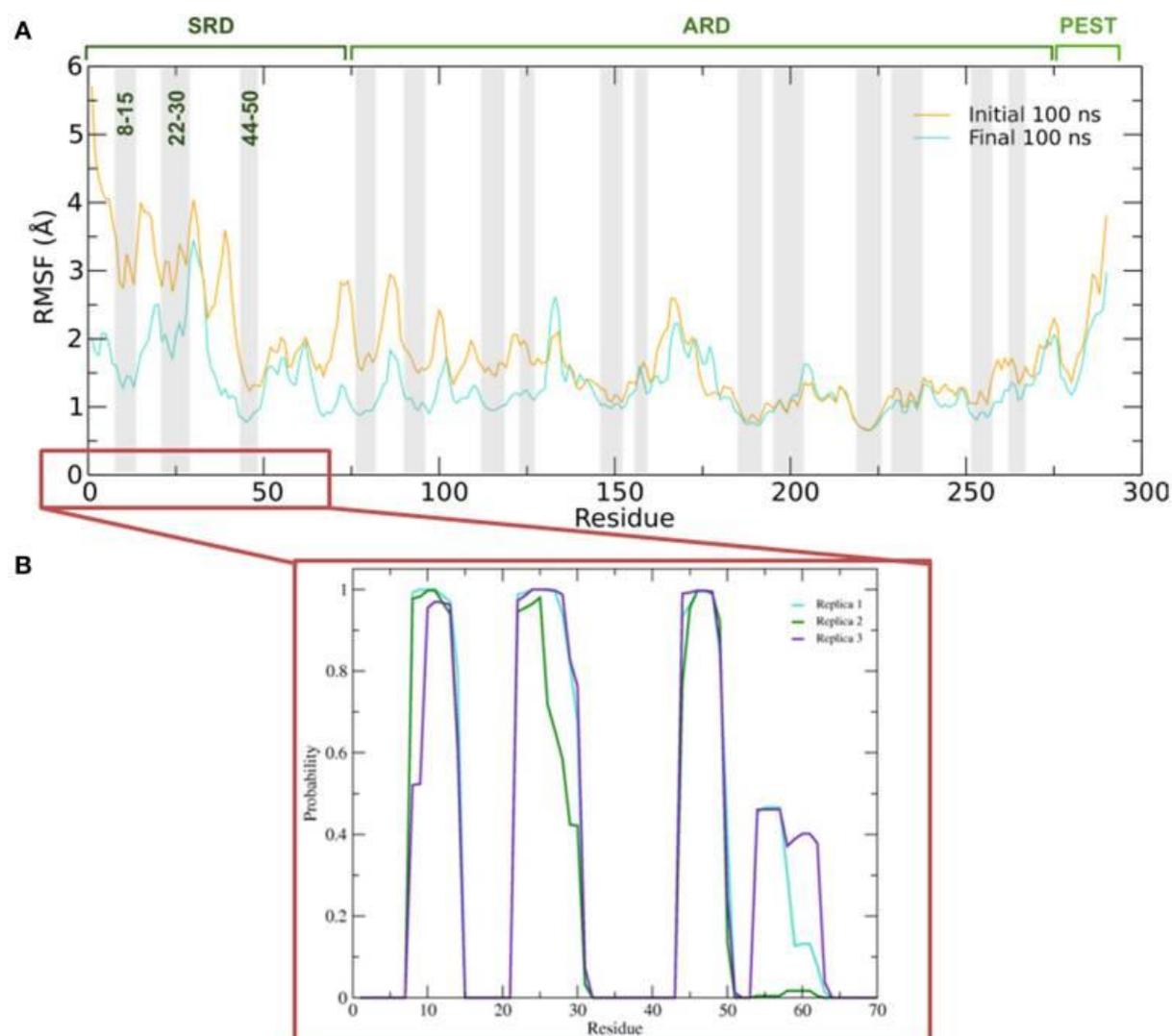


Figure I.3 (A) Average root mean square fluctuations (RMSF) of the backbone of $\text{I}\kappa\text{B}\alpha$ for the initial and final 100 ns of the simulation. Shaded areas depict α -helical regions at the end of the three independent 200 ns simulation periods. (B) Probability distribution of α -helix formation of the first 70 residues of the SRD of $\text{I}\kappa\text{B}\alpha$ in complex with NF- κB .

Figure I.3 B gives the probability distributions of helical formations in the SRD of $\text{I}\kappa\text{B}\alpha$. Together with the RMSF of Figure I.3 A, a consistent picture of stable vs. flexible subregions in the signal receiving domain is obtained. Residues 31-37 in the SRD immediately adjacent to the second α -helix in the N-terminal region represent the most

flexible part of $\text{I}\kappa\text{B}\alpha$, in the case of disregarding residues beyond 275 (Figure I.3 A). It is natural to discard residues beyond 275 from the comparison as they form a long loop and constitute a rather disordered region void of any tertiary structure. We do see the conservation of three α -helices, residues between 8-15, 22-30 and 44-50 within the SRD region. These values are in agreement with the predicted secondary structure models, which identified the three α -helices to lie between residues 10-13, 22-29 and 44-50. The last two α -helices align perfectly, while SYMPRED predicts a somewhat shorter α -helix compared to that observed in the refined structure. Furthermore, the fourth α -helical element, which was positioned from residues 54-63 from the 1N11 template, no longer adopts an α -helical shape but acquires instead a less ordered loop conformation (Figure I.3 B and Figure I.4, below). Here, obviously the refinement by MD simulations is sufficient to remove the ambiguous assignment of secondary structure elements and provide a more stable conformation of this stretch of 10 amino acid residues in length. All other secondary structure elements are retained during the MD simulations. This gives the confidence in the reliability of the protein-protein complex model and the existence of well-defined secondary structural elements in the SRD of $\text{I}\kappa\text{B}\alpha$ when it is in complex with NF- κB .

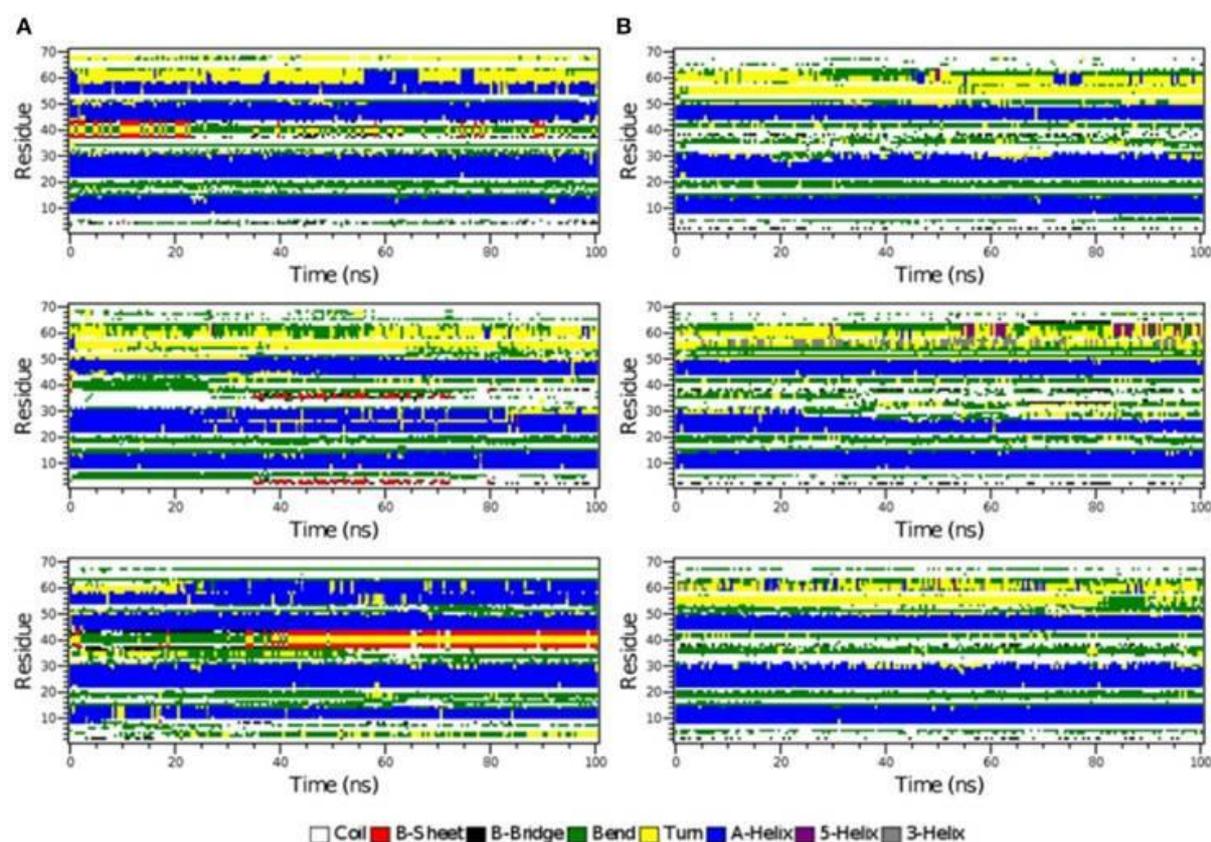


Figure I.4 The secondary structure elements of the first 70 N-terminal residues of $\text{I}\kappa\text{B}\alpha$ in complex with NF- κB as calculated by DSSP for the three system replicas during the initial 100ns (A) and final 100 ns (B) of the simulation.

The time-evolution of secondary structure elements in the N-terminal SRD during the MD refinement is then analyzed in detail. The DSSP-annotated SSEs of the first 70 amino acid residues in $\text{I}\kappa\text{B}\alpha$ for each of the replica systems is plotted in order to analyze the SSEs of the first 70 amino acid residues during the MD trajectory frames (Figure I.4). The

first three α -helices, residues between 8-15, 22-30, and 44-50 retain their α -helical structure (blue regions) during the initial 100 ns MD simulations in all three systems replicas. (Figure I.4 A). They are followed by a recurring β -sheet turn β -sheet formation (green-yellow-green). This region is followed by an unstable α -helix that is formed between residues 52-62. This short helix is observed only in two of the replicas (top and bottom). This segment mainly adopts the turn/bend secondary structure in the third replica. The structural stability is observed for the first three α -helices throughout the entire simulations during the final 100 ns of the simulation runs. (Figure I.4 B) The temporarily formed fourth α -helix, however, observed in the first 100 ns, is no longer formed and the sequence instead remains variable in its secondary structure. During most of the production runs, it takes a turn-like secondary structure (yellow) or bend (green) with short interludes of stretches of 3_{10} -helices (grey) and π -helices (purple).

Figure I.5 summarizes the results from secondary structure prediction, initial model generation and secondary structure elements of the full-length I κ B α obtained after MD refinement. Four helical stretches were detected from consensus SSE prediction and thus also represented the starting SRD model (top line, up to residue 70). After MD refinement, three helical stretches are structurally retained and the fourth one was not stable and adopts a disordered conformation, after the MD simulations. The ankyrin repeats of the ankyrin repeat domain (ARD) are structurally stable during the MD simulations of the protein-protein complex and well positioned with respect to the crystal structure.

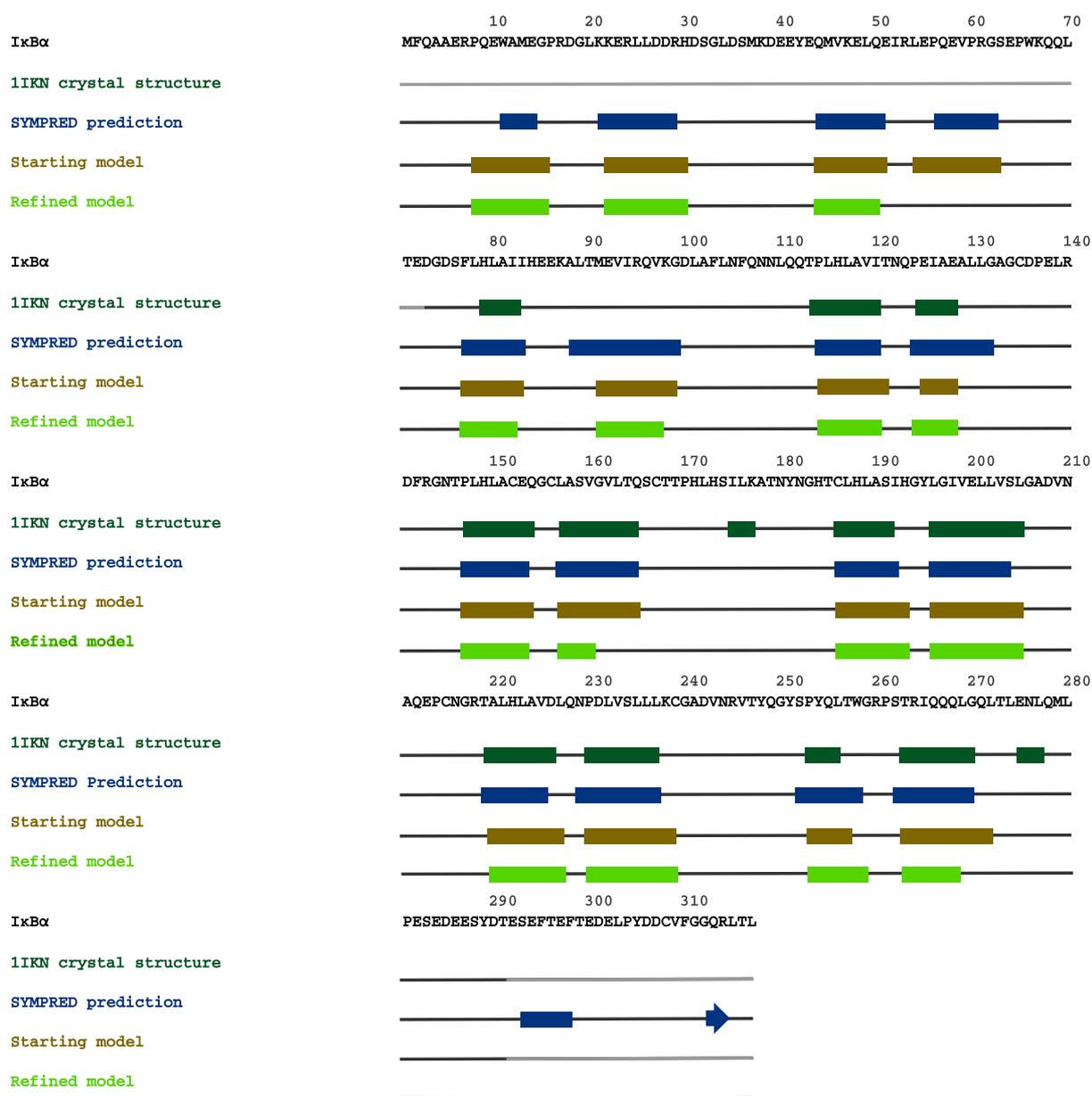


Figure I.5 A graphical map of the secondary structure elements of $\text{I}\kappa\text{B}\alpha$, displayed on its complete sequence. The boxes highlight the α -helical regions, and the arrows indicate β -strands. Dark green designates secondary structures determined in the crystal structure 1IKN, blue denotes secondary structures predicted by SYMPRED, and brown and fluorescent green indicate secondary structures suggested by our initial and refined structural models.

In Figure I.6, I present the refined structural model of the $\text{I}\kappa\text{B}\alpha$ -NF- κB complex (blue) portrayed together with the initial structural model (purple). The refined representative structure is depicting the last frame of a system replicate that has the lowest RMSD with respect to the average structure. This model reveals three helical structures in the previously not resolved SRD unit in addition to the six ankyrin repeats in the ANK protein domain. While the inner helix is nine residues long and extends from positions 22-30, the initial helix in the first pair of helices is eight residues long, spanning from positions 8-15 in the $\text{I}\kappa\text{B}\alpha$. The α -helix pair is followed by a 13-residue long loop, joining this element with the

consecutive α -helix of seven residues long covering positions 44-50. The lengthy loop linking the third helix to the subsequent ankyrin repeat domain comprises 26 residues, and connects the unresolved N-terminal segment of I κ B α to the crystallized ankyrin repeat domain of this protein. The structural superpositioning of the initial and refined models of I κ B α bound to its partner, NF- κ B, reveals an ANK domain that is partly rigid and well-structured. Ankyrin repeats 4-6 remained intact and display greater stability when bound to NF- κ B, while ankyrin repeats 1-3 show increased flexibility. This is in agreement with the analysis of residual dipolar coupling (RDC) of free and bound I κ B α which showed that the helix 2 from ankyrin repeat 3 differed most in the free and bound forms [65].

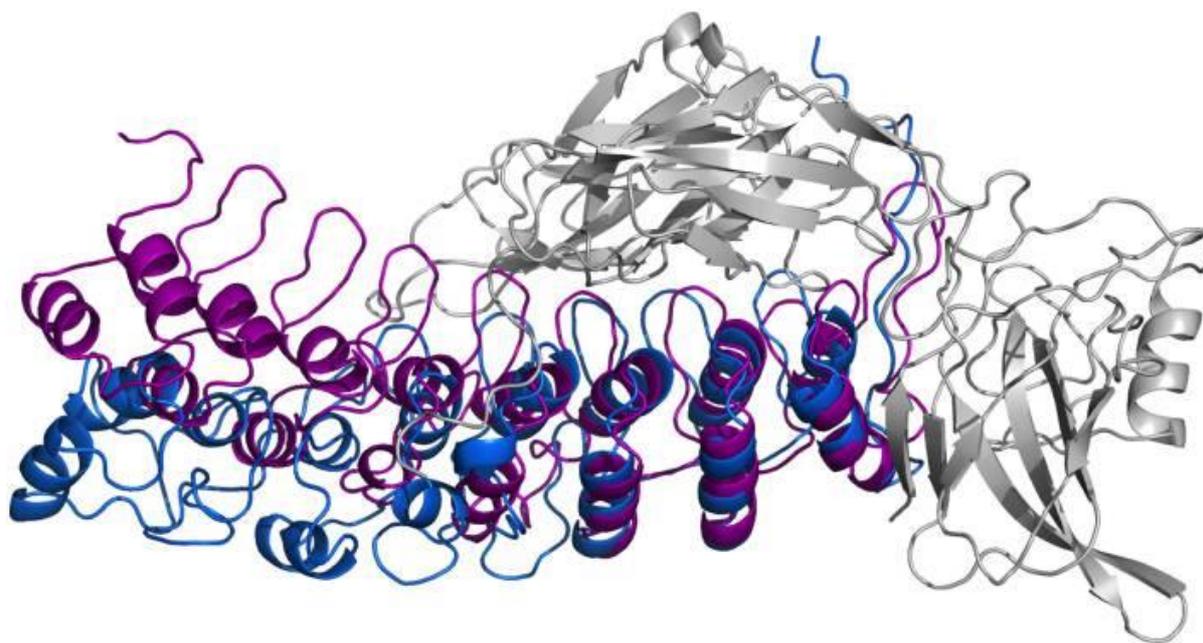


Figure I.6 Ribbon diagrams of the three-dimensional initial structure (purple) and the refined structure after a 200 ns MD simulation (blue) of I κ B α . The structures are shown in comparison by superpositioning I κ B α 's binding partner NF- κ B (gray).

In particular, ankyrin repeat 1 shows the greatest displacement, which together with the SRD segment move away from NF- κ B and deviate the most from the initial structure. This is in agreement with experimental studies which could show that the SRD does not contribute to the overall NF- κ B binding affinity to I κ B α [183]. Also, NMR studies of I κ B α in complex with its binding partner, NF- κ B, show a more flexible ankyrin 1-4 domain in comparison to rather rigid ankyrin repeats 5-6 [184]. An earlier amide H/D exchange study [64] indicated that when in complex with NF- κ B, ankyrin repeats 5 and 6 fold into compact domains upon binding to NF- κ B. Along with ankyrin repeats 5 and 6, ankyrin repeat 1 is another region seen to display greater conformational flexibility as observed here in the refined structure of I κ B α .

F.3.3 Conformational change induced in I κ B α in its bound form to NF- κ B

Thus, so far I have looked at the structural elements in I κ B α only. In the crystal structures and in the simulations however, I κ B α is in complex with NF- κ B (the RelA/p50 heterodimer) and for this reason it is imperative to look at the conformation of I κ B α in relation to its binding partner, NF- κ B, and see how the nature of this association was affected. The protein-protein surface area of interaction is larger than 4000 Å² and all six ankyrin repeat units are involved in forming a noncontiguous contact surface. Here, I discuss in particular electrostatic and hydrogen bonding interactions between I κ B α and RelA/p50. The hydrogen bonds that are discussed here remain intact for longer than 10 % of the simulation time and occur in at least two of the replicate simulations.

The I κ B α /RelA interface

I κ B α binds to RelA by forming a number of hydrogen bonds between different regions of each protein (Table I.1). Several residues situated on ankyrin repeats 5 and 6 form hydrogen bonds with residues located on both the RelA dimerization subunit and the RelA amino-terminal. The I κ B α carboxy-terminal residues are in close contact with regions on the amino-terminal and dimerization subunit of RelA and form several hydrogen bonds.

Table I.1 Hydrogen bond contacts between I κ B α and the p50/RelA subunits of NF- κ B.

Hydrogen bonds					
I κ B α	RelA	I κ B α	p50	RelA	p50
GLY155	ARG297	ASP73	LYS352	ARG198	HIS304
LEU157	ARG297	GLN107	LYS352	ASN200	ASP254
ASN216	ASP243	ASN109	ASP353	ASP243	ARG252
ASP226	SER238	ASN109	LYS352	HIS245	CYS270
THR247	ASP243	TYR181	THR256		
GLN249	ASP243	ASN182	THR256		
TRP258	GLN26	GLY183	THR256		
GLY259	GLN241	CYS215	MET253		
GLN266	ILE24	TYR248	LYS343		
GLN267	GLU22	TYR248	GLU341		
GLN271	LEU179	GLN249	ARG252		
GLN271	VAL21				
LEU280	GLN29				
SER283	GLU222				
ASP285	GLN247				
GLU286	GLN247				
GLU286	THR191				
SER288	GLN247				

Residues found in interactions in the crystal structure 1IKN are shown in bold. All Bonds are present for more than 10 % of the total simulation time in at least 2 of the replicate simulations.

The other major source of stabilization is via electrostatic interactions from the salt bridge interactions between the carboxy-terminal of I κ B α and different regions of RelA

(Table I.2). The ARD region of I κ B α contributes to the I κ B α /RelA stabilization by forming salt bridges between ASP226 and ARG218 on ANK5 and between ARG253 and ASP243 as well as GLU211 on the dimerization component. In addition, ARG264 on ANK6 interacts with GLU22 on the amino-terminal of RelA. Specifically the interaction between ARG218 and ASP243 is also observed to form in the crystal structure of I κ B α [61].

Table I.2 Salt bridge formations between I κ B α and the p50/RelA subunits of NF- κ B.

Salt bridges					
I κ B α	RelA	I κ B α	p50	RelA	p50
GLU85	ARG302	GLU41	LYS354	ARG198	ASP302
GLU85	LYS301	GLU72	LYS352	ARG198	GLU265
GLU86	ARG302	GLU72	LYS354	ARG201	ASP254
GLU125	ARG302	ASP73	LYS354	ARG201	GLU265
ARG143	ASP294	GLU138	LYS323	GLU211	ARG252
GLU153	ARG295	ARG143	GLU350	ASP217	ARG305
GLU153	ARG297	GLU286	ARG305	ASP243	ARG252
ARG218	ASP243	GLU286	LYS272	ARG246	ASP271
ARG218	GLU211	GLU287	LYS272		
ASP226	ARG253	ASP290	LYS272		
ARG264	GLU22	GLU292	LYS249		
GLU282	ARG30	GLU292	LYS272		
GLU282	ARG158				
GLU282	LYS79				
GLU284	ARG246				
GLU284	LYS79				
ASP285	LYS218				
GLU286	LYS218				
GLU287	ARG246				
GLU287	LYS218				
ASP290	LYS221				
GLU292	ARG246				

Residues found in interactions in the crystal structure 1IKN are shown in bold. All salt bridges occur in at least 2 of the replicate simulations throughout the whole simulation.

The elongated and relatively flexible 13 residue carboxy-terminal of RelA, known as the NLS polypeptide, extends across ankyrin repeats 1-3 and makes several contacts with residues present on the loops and helical regions of these ankyrin repeats, forming both hydrogen bonds and salt bridges.

The I κ B α /p50 interface

A number of residues on ankyrin repeats 4-6 interact with the dimerization domain on p50 by forming hydrogen bonds. Among these interactions, TYR181 has previously been shown to be a key player in the interaction between NF- κ B and I κ B α [61]. Eminently, residues

CYS₂₁₅, TYR₂₄₈, and ARG₂₅₂ on the p50 subunit are among those reported to form interactions in the crystal structure of I κ B α . The amino acid residues LYS₃₅₂-ASP₃₅₃ located on the carboxy-terminal of p50 engage in additional hydrogen bond interaction with the residues ASP₇₃, GLN₁₀₇, and ASN₁₀₉ situated on ankyrin repeats 1 and 2. The interaction between I κ B α and p50 is further stabilized by electrostatic interactions. The carboxy-terminal PEST sequence residues GLU₂₈₆-GLU₂₈₇, ASP₂₉₀, GLU₂₉₂ in I κ B α take part in forming salt bridges with the residues LYS₂₄₉, LYS₂₇₂, ARG₃₀₅ on the amino-terminal and the interconnecting loops on the 'top' of the p50 subunit. Ankyrin repeats 1-3 and the SRD in I κ B α and the carboxy-terminal and an interconnecting loop at the 'bottom' of p50 participate in another set of salt bridge network involving residues GLU₄₁, GLU₇₂-ASP₇₃, GLU₁₃₈ and LYS₃₂₃, LYS₃₅₂, LYS₃₅₄, on respective chain. Notably, with one single exception, the acidic residues are contributed by I κ B α , whereas the basic residues are to be found on the p50 subunit.

The RelA/p50 interface

The dimerization interface takes part in several hydrogen bonds formed by 8 residues including an ASP₂₅₄(p50)/ASN₂₀₀(RelA) hydrogen bond. This hydrogen bond can also be found in the crystal structure and is considered one of the most critical interactions in discriminating subunit dimerization specificity among NF- κ B dimers [61,185,186]. The other hydrogen bonds include HIS₃₀₄(p50)/ARG₁₉₈(RelA), ARG₂₅₂(p50)/ASP₂₄₃(RelA), CYS₂₇₀(p50)/HIS₂₄₅ (RelA). The RelA/p50 dimer interface is additionally stabilized by electrostatic interactions. Several residues form salt bridges between the two subunits. Two of these include salt bridges that are also reported for the crystal structure namely residues ASP₂₁₇ and ASP₂₇₁ on the p50 subunit and ARG₃₀₅ and ARG₂₄₆ on the RelA component, respectively [61].

F.3.4 Free I κ B α vs bound I κ B α

All the simulations discussed above were describing the stable, long-living complex of I κ B α with its binding partner NF- κ B, as revealed in their crystal structures. All efforts to crystallize I κ B α in its unbound state have been unsuccessful [63]. For this reason, additional simulations of full-length free I κ B α in solution were performed and compared with the more stable NF- κ B-bound state.

The simulation setups followed the same procedure as for the bound I κ B α , resulting in three replicated systems of 200 ns each. Conservation of total energy and temperature of the three simulations points to systems that have reached a stable state. In contrast to the bound I κ B α , the RMSF of the free state of I κ B α remains on average ~ 1 Å higher compared to its complexed state. This points to a higher degree of flexibility of free I κ B α compared to its complexed state. The RMSF of the backbone of the protein around the average structure in the modeled SRD remains the most flexible domain throughout the protein in addition to the unstructured C-terminal region (Figure I.7 A). The probability distributions of the helical propensity in the SRD of the bound I κ B α reveal (Figure I.7 B) the first three helical segments to be stable throughout the whole simulation. In free I κ B α , although the first three helical segments are present in all three replicate simulations, we

observe different probabilities across the different replicate simulations. Bound I κ B α displays a narrower distribution of probabilities of helical regions and this indicates to a stabilization of the SRD upon complexation with NF- κ B. The fourth initially assigned helix in the SRD varies in both length and probability in both the bound and free forms of I κ B α , indicating that this fourth helix is not well defined and not stable during MD refinement.

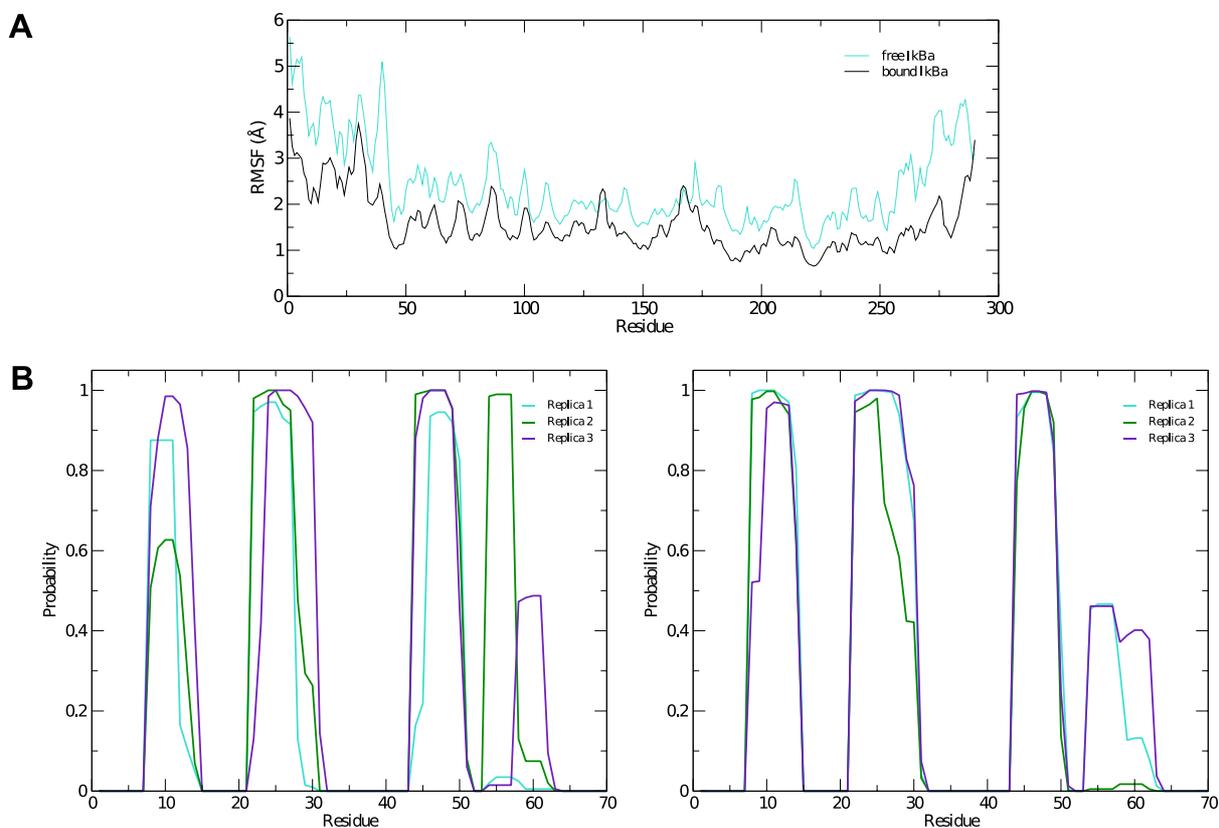


Figure I.7 (A) Average root mean square fluctuations (RMSF) of the backbone of the free I κ B α (cyan) in comparison to the one in complex with NF- κ B (black). (B) Probability distributions of α -helix formation of the first 70 residues of the SRD. Left: free I κ B α . Right: I κ B α in complex with NF- κ B.

The secondary structure evolution of the first 70 amino acid residues in the SRD of the free I κ B α (Figure I.8 A) reveals greater differences in the SRD in terms of secondary structure element evolution in comparison to the bound I κ B α . The first helix in the free I κ B α is considerably shorter than its counterpart in the bound I κ B α . During the first 100 ns of the simulations, this helix can be clearly distinguished whereas it is only present in 2 of the replicate runs in the final simulation period. The second and third helices remain intact throughout the entire 200 ns simulations in all three replicate runs, which is very similar to the pattern seen in the bound I κ B α simulations. In contrast to the bound I κ B α , here one observes the formation of a 4 residue long fourth helix in 2 of the replicate runs; in one of the simulations this helix is present during the entire simulation, whereas in the other run it appears in the last 100 ns of the simulations with irregular intervals. In a previous study [68], the conformations of a short 24 amino acid peptide (residues 21-44) of the doubly phosphorylated free I κ B α were characterized by NMR spectroscopy and MD simulations and compared to its β -TrCP bound state using saturation transfer difference NMR. The

conformational observation agreed on the presence of a bend between residues 30 and 36 in both states of the phosphorylated peptide, a trend which I also observe throughout the simulations of the free and NF- κ B bound states of I κ B α . While the N-terminal of amino acids 30 to 36 is preceded by a short α -helix and the C-terminal succeeded by a region of β -sheet–turn– β -sheet flanked by bends in the free and bound states of I κ B α in this study, Pons et al. observed disordered N- and C-terminal segments in the free I κ B α versus the adoption of turns in the bound state I κ B α . This difference in results can be rationalized from the truncation of the peptide which could have influenced the conformational integrity of the N- and C-terminals, an effect which would not be detectable in our structural models of the full-length I κ B α .

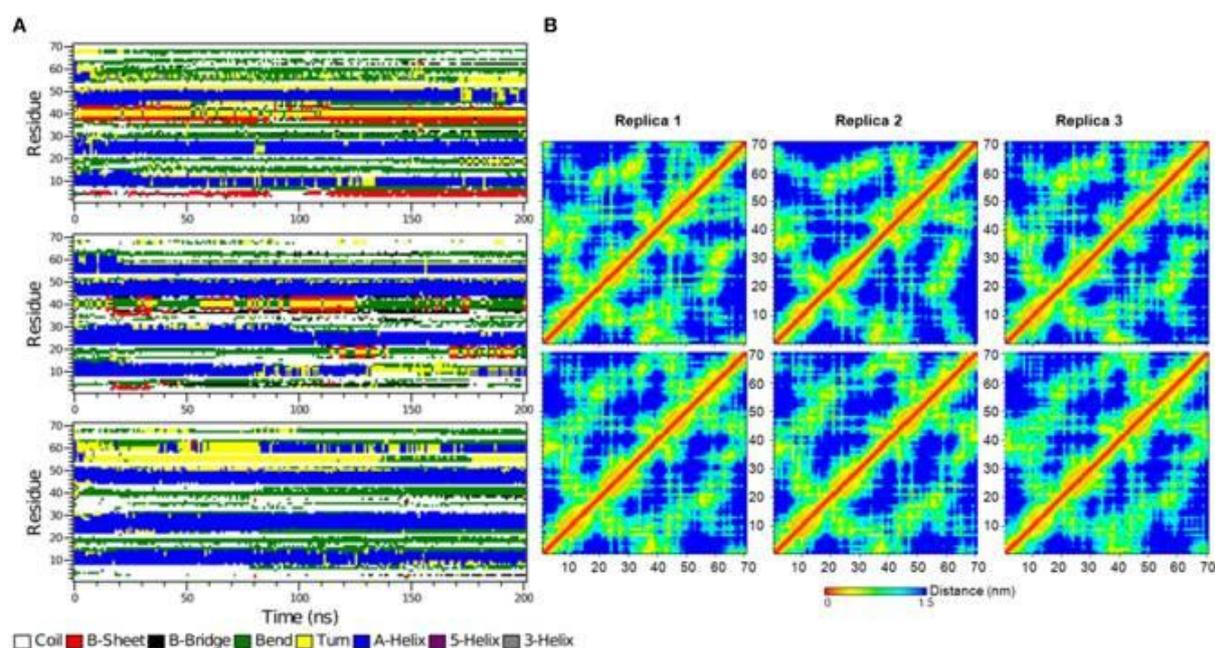


Figure I.8 (A) The secondary structure elements of the first 70 N-terminal residues of free I κ B α calculated by DSSP for the three system replicas for the entire simulation. (B) Interatomic distance matrices for the first 70 N-terminal residues of free I κ B α (top) and in complex with NF- κ B (bottom).

Figure I.8 B shows the interatomic distance matrices depicting the smallest distance between residue pairs in the SRD of I κ B α for both free (top) and complexed I κ B α (bottom). The distance matrices of all three replicates are very similar and there are no large differences in interatomic distances upon NF- κ B binding. The red and yellow colors indicate shorter distances between the residues and are more detectable for regions where helical segments are present in the SRD. In both the unbound and free forms of I κ B α , the fourth segment is less apparent across the replicates.

There are, however, also apparent stretches of amino acids which display a higher degree of flexibility upon NF- κ B complexation (Figure I.7 A). The residues around positions 133 and 167 become more flexible upon protein-protein complex formation. These positions correspond to loop regions following the outer helices in AR2 and AR3. This was also found by analyzing residual dipolar coupling (RDC) of ARs 1-4 [65].

Another interesting comparison between the free and bound I κ B α structures is the solvent accessible surface area (SASA) or the relative solvent accessible area (RSA) of the

phosphorylation and ubiquitination sites located on the SRD (Table I.3). These sites (SER32 and SER36 for phosphorylation and LYS21 or LYS22 for ubiquitination) ought to become accessible by the kinase IKK and the E3 ligase, respectively, in the complexed form of I κ B α . The RSA is computed by the SASA of the residue normalized by the accessible surface area of that residue in its extended tri-peptide (Gly-X-Gly) conformation. By setting a threshold of < 20 % for buried residues, SER32 and SER36 are both surface-exposed in the bound I κ B α , while in its free state only SER32 lies above the threshold. SER36 in the free state has an RSA of 9.3 %, which is considerably lower than the threshold and can be considered to be a buried residue. As regards to the ubiquitination sites, LYS21 stays well buried in both the free and bound states of I κ B α . However, LYS22 with an RSA of well over 60 % in both states of I κ B α remains surface exposed. Thus, in the bound-form of I κ B α the phosphorylation SER32 and SER36 sites are accessible by the IKK and we suggest LYS22 to be the putative site of ubiquitination.

Table I.3 Solvent accessible surface area (SASA) and relative surface area (RSA) of the free and bound I κ B α .

	SASA (Å ²)		RSA (%)	
	Bound I κ B α	Free I κ B α	Bound I κ B α	Free I κ B α
SER32	61.2±18.5	43.4±15.7	50.2±15.2	35.6±12.8
SER36	50.6±14.0	11.3±9.1	41.5±11.5	9.3±7.4
LYS21	18.5±4.8	39.2±16.4	8.8±2.3	18.6±7.8
LYS22	135.6±6.0	127.5±17.9	64.3±2.8	60.4±8.5

The accessible surface areas of serine and lysine are 122 Å² and 211 Å² respectively as calculated by Miller et al. [187]. The SASA values are the averages of the three replicate simulations over the 200 ns of total simulation time.

Chapter G

Study II

G.1 Double phosphorylation-induced structural changes in the signal-receiving domain of I κ B α in complex with NF- κ B

The findings in study I provide the starting platform for investigating and understanding the mechanism of phosphorylation in I κ B α by IKK and the effect of single and double phosphorylation by IKK. In an approach to unravel the molecular basis of NF- κ B signaling by multiscale molecular simulations, the structural model of full length I κ B α obtained in Study I was further refined for this study (Figure II.1 A).

Mutation of either serine residue at sites Ser32 and Ser36 (Figure II.1 C) have been shown to disrupt polyubiquitination of I κ B α and abolish the stimulus-induced degradation of I κ B α , thereby requiring the strict phosphorylation of both serine residues as a recognition and initial step in NF- κ B activation [47,188-192]. Although phosphorylation of I κ B α precedes its dissociation from NF- κ B, it is not sufficient and requires additional ubiquitination to enable its dissociation and subsequent degradation [189].

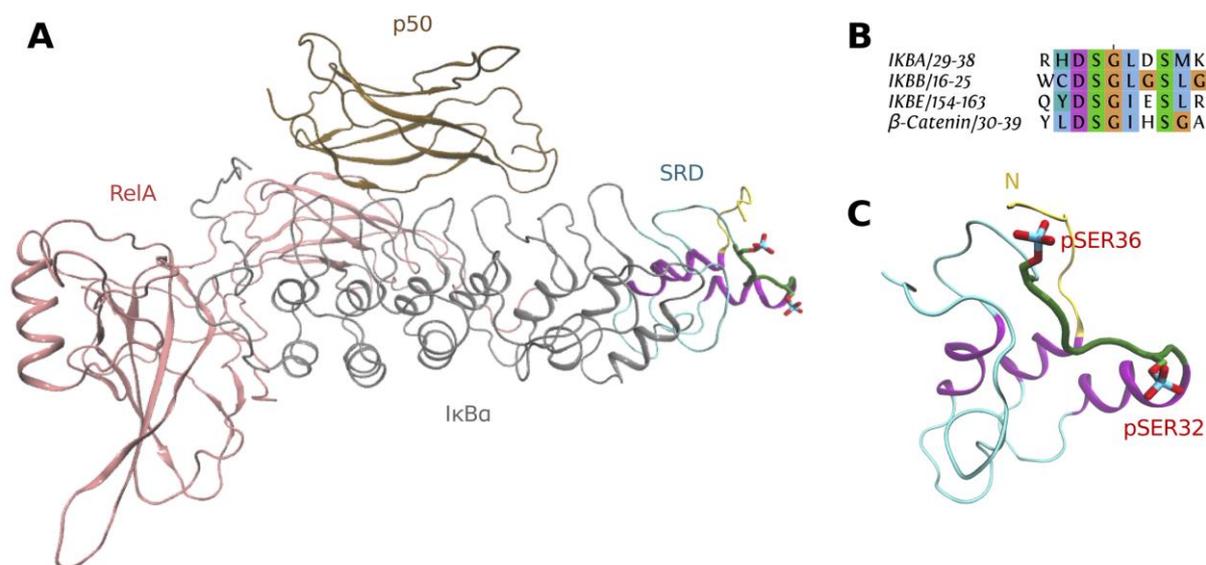


Figure II.1 (A) A ribbon representation of the structure of a complete $I\kappa B\alpha$ and its phosphorylation sites bound to the partnering NF- κB , composed of the RelA and p50 subunits. RelA is colored in pink and p50 colored in brown. The ankyrin repeat domain and the C-terminal of $I\kappa B\alpha$ are displayed in grey with the modeled signal receiving domain (SRD) as part of the N-terminal region colored according to its secondary structure elements with the phosphorylation sites pSer32 and pSer36 displayed in atomic detail. (B) A multiple sequence alignment of the N-terminal regulatory region, the degron motif constituted of residues DSG ϕ XS, with ϕ corresponding to a hydrophobic residue (here isoleucine) and X to any amino acid. (C) A close-up of the $I\kappa B\alpha$ SRD displaying the modeled helical structural elements displayed in purple. The phosphorylation sites, Ser32 and Ser36, flank the degron motif highlighted in green.

To paint a complete picture of how phosphorylation guides and regulates protein function, the rub lies in understanding the molecular basis of this phenomenon. With phosphorylation being the most well-studied posttranslational modification, computational simulation methods on the atomistic scale have proven valuable in offering explanations of the structural basis of protein regulation by phosphorylation in addition to complementing experimental studies [193-195].

The crystal structure of the β -TrCP/ β -catenin complex has revealed the basis of substrate recognition by the WD₄₀ domain in β -TrCP and shed light on the intimate interaction between degron motif containing substrates (Figure II.1 B) and the SCF ubiquitin ligase [56]. Of similar interest would be the interactions between β -TrCP and the phosphorylated $I\kappa B\alpha$; however prior to this, the subject of interest is the structural rearrangement triggered by phosphorylation of the SRD in the ankyrin repeat protein which drives the recognition mechanism by SCF. To this end, I have in this study examined the role of mono- and double-phosphorylation by probing into the structural effects that it induces on $I\kappa B\alpha$ bound to its partner, NF- κB . By applying long scale MD simulations of 500 ns, phosphorylation of Ser32 and Ser36 of $I\kappa B\alpha$ in complex with NF- κB , both in their double- and mono-states of phosphorylation were studied. This study identified a partially more stable region locally by the site of phosphorylation in the double-phosphorylated $I\kappa B\alpha$. Moreover, double-phosphorylation induced a rather extended conformation in

vicinity of the phosphorylation site as compared to a more bent shape observed in the unphosphorylated state. The mono-phosphorylated states each fell into an intermediary state between the unphosphorylated and the double-phosphorylated states, only half initiating this structural transition. In addition, distinct variations between pSer32 and pSer36 were observed. pSer32 emerged as the more solvent exposed residue and subsequently burying Ser36 slightly causing the formation of a new hydrogen bond pattern stabilizing the N-terminal tail and the region closest to pSer36. The findings also revealed a more pronounced electrostatic effect upon double-phosphorylation, which induce structural rearrangements that change the surface charge potential and create a greater acidic environment around the phosphorylation site. Overall, these findings explain the prerequisite for double-phosphorylation on a detailed molecular level and offer an insight into the structural rearrangements that take place, hence laying the ground work into future studies of I κ B α / β -TrCP recognition and binding mode.

G.2 Methods

G.2.1 Structural model of I κ B α /NF- κ B

The model of the full-length I κ B α /NF- κ B complex was obtained as described in the previous study (Study I), containing the complete signal receiving domain (SRD; residues 1-67) of I κ B α and the ankyrin repeat domain (ARD; 68-280) thereby providing a suitable working model to investigate the phosphorylation in the SRD region of I κ B α . A model of a double-phosphorylated simulation system is presented in Figure II.1 A.

G.2.2 Molecular simulation setup

The well-equilibrated structural model of the unphosphorylated I κ B α /NF- κ B complex from Study I of this thesis, was used as an initial structure for the simulations. To investigate the role of phosphorylation, four distinct system setups were constructed: an unphosphorylated (nl κ B α) system, two mono-phosphorylated complexes where either of serines 32 (p32I κ B α) or 36 (p36I κ B α) were post-translationally modified into phosphoserines, and a double-phosphorylated system (ppI κ B α) in which serine residues 32 and 36 were both transformed into phosphoserines simultaneously. The phosphoserine mutations were introduced with PyMol [196]. The simulations were carried with Gromacs 4.5 [140] with the modified GROMOS96 43a1p force field [172] containing the parameters for phosphoserine. A time step of 2 fs was used. The LINCS algorithm [173] was applied for constraining bond lengths. Electrostatic interactions were calculated with the Particle-Mesh Ewald algorithm at every step [174]. A 1.0 nm cutoff was used both for electrostatics and van der Waals interactions, with neighborlists updated every 10 steps. The simulations were performed at constant pressure of 1.0 bar with Parrinello-Rahman pressure coupling [197] and the isotropic pressure scaling, time constant of 1.0 ps, and a system compressibility of 4.5e-5 bar⁻¹. The temperature of the system was coupled to 300 K using the velocity-rescaling algorithm [198]. The nl κ B α system containing 6945 protein atoms

was solvated with 31,361 SPC water molecules, 218 out of which were replaced with 122 sodium and 96 chloride ions to neutralize the total net system charge and obtain a physiological salt concentration of 0.15 M. The total number of atoms in the wild-type system reached 100,226 all of which were placed in a rectangular cell of an approximate system size of $9.4 \times 8.0 \times 13.7 \text{ nm}^3$. For the PTM-modified systems, sodium ions were replaced with corresponding number of water molecules to maintain a neutralized system charge depending on the phosphorylation state of the systems. The systems were energy minimized with steepest descent until the maximum force reached $< 100 \text{ kJ/mol/nm}$ and were subsequently equilibrated for 100 ps while keeping the protein position restrained ($F_c = 1000 \text{ kJ/mol/nm}^2$). Eventually, all position restraints were removed and the production runs were performed for 500 ns each. Three independent replicates of each system were simulated, each starting with different initial velocities, amounting to a total simulation time of 6 μs , thus allowing a thorough exploration of the extensive conformational space.

6.3 Results

In order to study the effects of phosphorylation of Ser32 and Ser36 located on the SRD of $\text{I}\kappa\text{B}\alpha$, I carried out MD simulations of the unphosphorylated and the double-phosphorylated $\text{I}\kappa\text{B}\alpha$ in complex with the transcription factor NF- κB . To further characterize the solitary roles of mono-phosphorylation of Ser32 and Ser36, two additional mono-phosphorylated systems were studied and compared with the unphosphorylated and the double-phosphorylated $\text{I}\kappa\text{B}\alpha/\text{NF-}\kappa\text{B}$ complex. The reported results are based on an average finding of the three independent replicate runs of each system, which were simulated for 500 ns each.

6.3.1 Local divergence in structural stability promoted by double-phosphorylation

The degree of flexibility of the protein can be evaluated by measuring the backbone atom root mean square fluctuation (RMSF) around the initial structure. The stretch of residues, $\text{I}\kappa\text{B}\alpha^{28-40}$, encompassing the sites of phosphorylation Ser32 and Ser36 as shown in Figure II.2 A, reveal a sharp contrast between the unphosphorylated and the double-phosphorylated system. The RMSF of this segment varies between 2-4 Å in the unphosphorylated system, while remaining rather stable in the double-phosphorylated system maintaining an RMSF between 2 and 3 Å (Figure II.2 B).

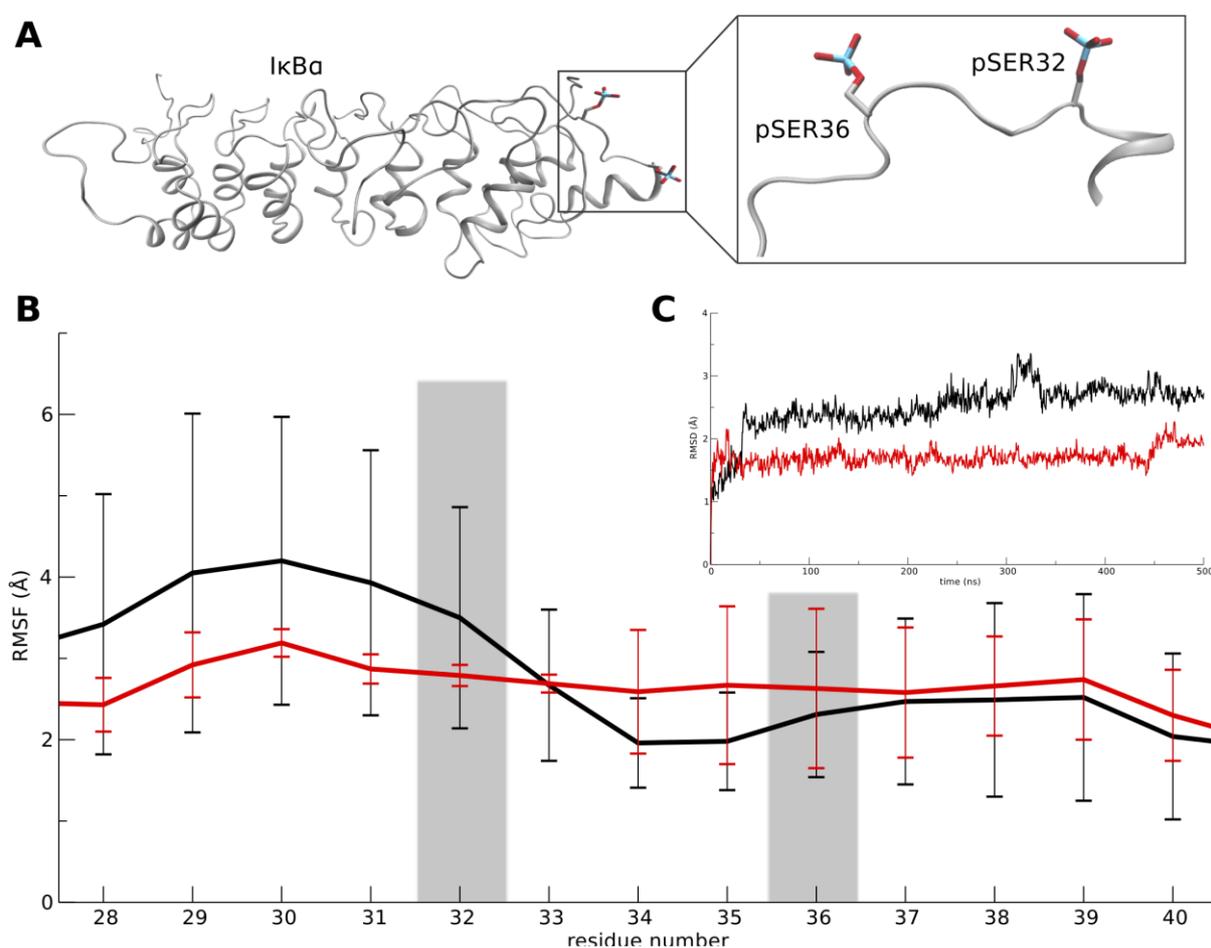


Figure II.2 (A) A ribbon representation of $I\kappa B\alpha$, magnifying the stretch of residues $I\kappa B\alpha^{28-40}$ situated on the SRD encompassing the two phosphorylation sites, SER32 and SER36. (B) Root mean square fluctuation (RMSF) of the backbone atoms of segment $I\kappa B\alpha^{28-40}$ in $nI\kappa B\alpha$ (black) and the $pI\kappa B\alpha$ (red) systems. The highlighted regions display the sites of phosphorylation, residues 32 and 36. (C) Root mean square deviation (RMSD) of backbone atoms for segment $I\kappa B\alpha^{28-40}$ during 500 ns for the $nI\kappa B\alpha$ and $pI\kappa B\alpha$ systems, calculated relative to the initial conformation after a least-squares fitting of $I\kappa B\alpha^{28-40}$. A more stable and lower RMSD is observed in the double-phosphorylated system.

In particular around residues preceding the initial site of phosphorylation at pSer32, residues 28 to 32, the RMSF reveals a more stable segment in the double-phosphorylated system. The RMSF of these residues lies around 2.5-3 Å in the double-phosphorylated system with very small fluctuations across the replicated runs as displayed by the error bars, whereas in the unphosphorylated system a higher RMSF of 3.2-4 Å is observed with much greater fluctuation across the different replicates. However, interestingly, the residues preceding the second phosphorylation site, pSer36, residues 33 to 35 including pSer36, display a somewhat lower RMSF in the double-phosphorylated system as compared to the unphosphorylated system. The general observations in the same stretch of residues are supported by the root mean square deviation (RMSD) of the backbone atoms of this region (Figure II.2 C). The RMSD of this segment keeps steady at roughly 2 Å in the double-phosphorylated complex, in comparison to the unphosphorylated system that has an RMSD of > 2.5 Å, a difference of > 0.5 Å observed between the two systems.

In the mono-phosphorylated systems, the p36I κ B α system is structurally more stable compared to the p32I κ B α system, with a sustained difference of 1 Å in the RMSF of the residues surrounding the phosphorylation site, a trend also reflected in the RMSD in particular in the final 300 ns of the simulations (Figure II.3). An opposing effect of the two sites of phosphorylation is revealed in the mono-phosphorylated states when compared to the unphosphorylated state: pSER32 modification leads to an increase in RMSF in immediate proximity of the phosphorylation site, whereas pSER36 causes a drop in RMSF in surrounding residues.

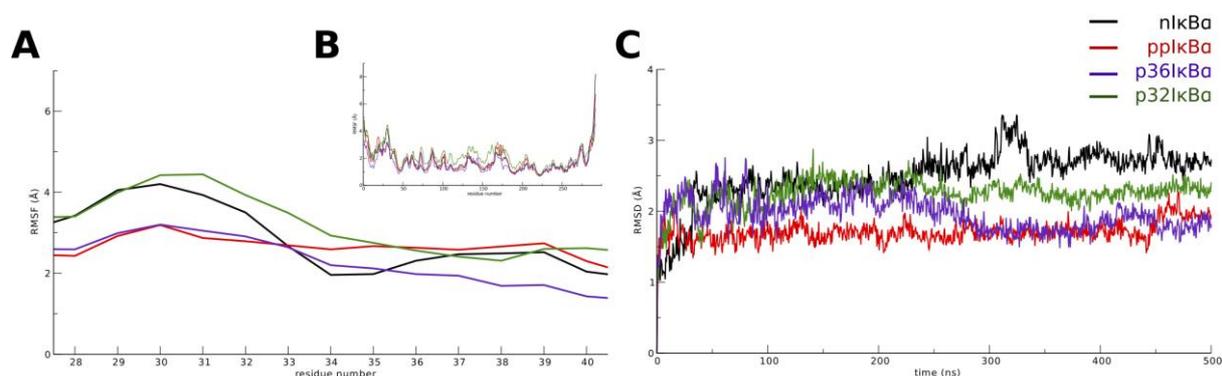


Figure II.3 Root mean square fluctuations (RMSF) of the backbone atoms of segment I κ B α ²⁸⁻⁴⁰ (A) and the entire protein (B) in all states. (C) Root mean square deviation (RMSD) of backbone atoms for segment I κ B α ²⁸⁻⁴⁰ during 500 ns for all systems, calculated relative to the initial conformation after a least-squares fitting of I κ B α ²⁸⁻⁴⁰.

6.3.2 Double-phosphorylation induces an extended N-terminal conformation of SRD

To further characterize induced structural variation due to two-fold phosphorylation, the C α -C α distance between Ser32 and Ser36 was measured (Figure II.4 A). By mapping the C α distances between these residues, one is able to investigate in detail the effect of mono-phosphorylated pSer32/nSer36, nSer32/pSer36 and double-phosphorylated pSer32/pSer36 and compare with the not-PTM nSer32/nSer36 'unphosphorylated' system.

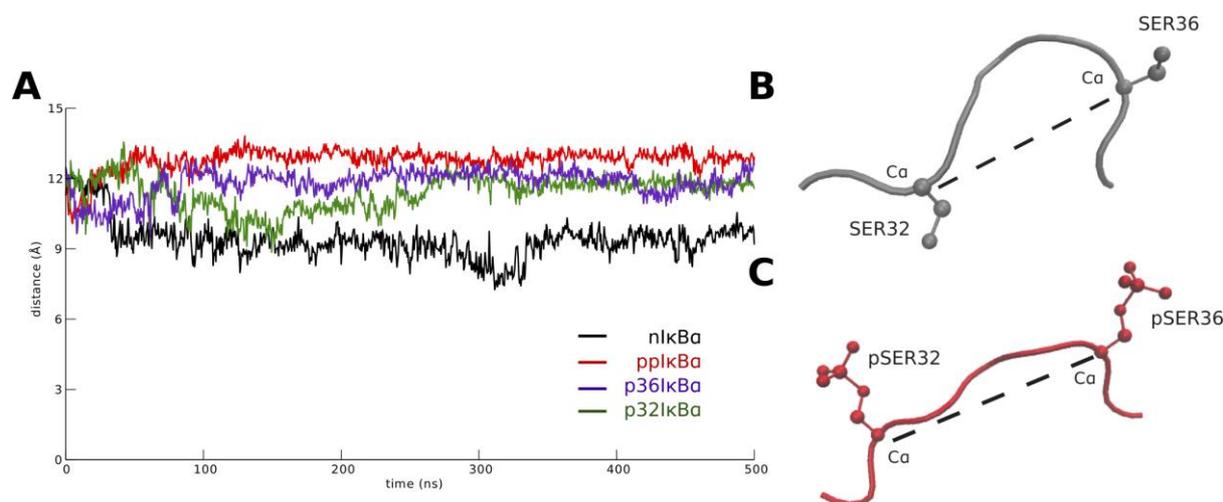


Figure II.4 (A) The Ser32^{Cα}-Ser36^{Cα} distance compared between the different simulation systems. (B) Representative structure of the degron containing segment IkBa³¹⁻³⁷ in the unphosphorylated and (C) the double-phosphorylated system.

A remarkable deviation in the Ser32^{Cα}-Ser36^{Cα} distance could be observed across the unphosphorylated and the double-phosphorylated complexes. In the doubly phosphorylated state, the Cα-Cα distance increases gradually in the first 30 ns of the simulation, after which it reaches a distance of 13 Å and maintains it throughout the rest of the simulation. In contrast, the Cα-Cα distance in the not phosphorylated state drops sharply down to 9 Å in about 30 ns and sustains this distance at about 9-10 Å for the remaining of the 500 ns simulation duration averaged over three independent simulation runs. In both of the mono-phosphorylated systems, the Cα-Cα distance fluctuates between 9-12 Å in the initial 300 ns of the simulation, however, into the final 200 ns both systems keep a rather constant distance at approximately 12 Å. This shows that two-fold phosphorylation is more effective than mono-phosphorylation of either Ser32 or Ser36 residues but that the structural effect is not additive.

This variation in the Ser32^{Cα}-Ser36^{Cα} distance between the unphosphorylated and the double-phosphorylated states is exhibited in the structural conformation of the phosphorylation region. The representative conformations of a stretch of residues 31 to 37 in the unphosphorylated (Figure II.4 B) and the double-phosphorylated (Figure II.4 C) complexes, clearly portrait a region with a defined bend in between the sites of phosphorylation in the unphosphorylated system, whereas a more extended structure is revealed upon double-phosphorylation.

G.3.3 Variation in solvent exposure in Ser32 and Ser36

The impact of phosphorylation on sites Ser32 and Ser36 can be assessed by calculating the relative solvent accessible surface area (SASA) of the individual serine residues.

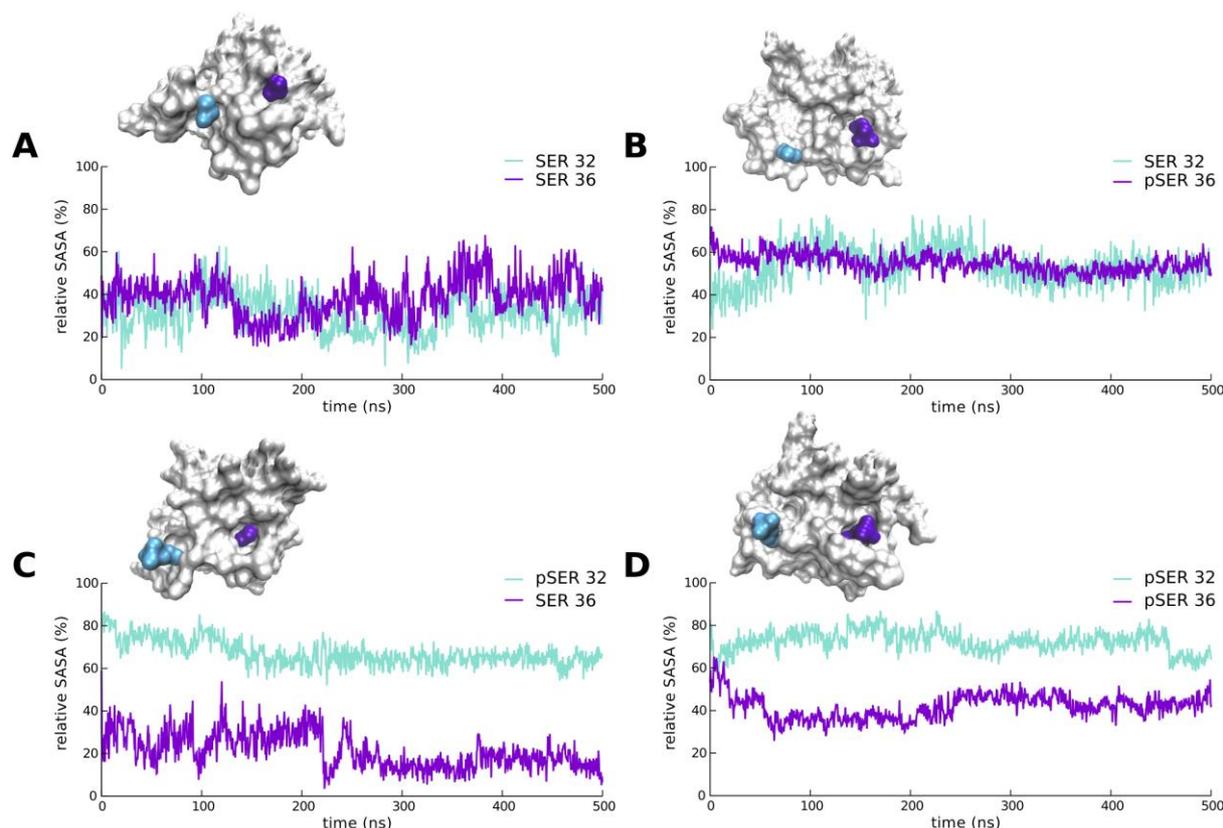


Figure II.5 Relative solvent accessible surface areas (SASA) of Ser32 and Ser36 in the unphosphorylated (A), the pSer36 mono-phosphorylated systems (B). pSer32 mono-phosphorylated (C), and double-phosphorylated (D); Ser32 is depicted in cyan, and Ser36 is colored in purple.

As evident from Figure II.5 A, in the unphosphorylated complex, both Ser32 and Ser36 remain surface exposed (staying above a threshold of 20 %), with Ser36 holding the upper hand with a slight margin fluctuating around a relative SASA of 40 %. Intriguing are also the effects of mono-phosphorylation on each respective unphosphorylated partner. Mono-phosphorylation of pSer32 sets in motion structural rearrangements, which in turn induce a partial burial of Ser36 as shown in Figure II.5 B. However, the mono-phosphorylation of pSer36 triggers firstly, the increase of the relative SASA of pSer32 as compared to the unphosphorylated structure, and secondly, it allows a higher surface exposure of pSer36, even higher than the pSer36 relative SASA in the double-phosphorylated state (Figure II.5 C). In clear contrast to the unphosphorylated state, double-phosphorylation leads to a striking difference in the relative SASA between pSer32 and pSer36 (Figure II.5 D). Although phosphorylated, pSer36 maintains a similar but less fluctuating relative SASA of 40 %, suggesting partial burial of this residue. However, pSer32 maintains a highly exposed configuration with a relative SASA of 80 % during a larger part of the simulation, allowing the potential for a wide range of interactions. These observations are suggestive of the distinctive roles that pSer32 and pSer36 could play upon phosphorylation: a rather exposed pSer32 which extends beyond the surface of the protein to act as an anchoring point and engage in inter-protein interactions, and a comparatively less exposed pSer36 that could be contribute to key-specific intra-protein interactions.

G.3.4 Double-phosphorylation stabilizes region by novel hydrogen bond interactions

The conformational differences caused by phosphorylation may be accounted for by a new hydrogen bond formation pattern in the double-phosphorylated complex. Based on our simulations, out of the two phosphorylation sites, it is in particular pSer36 that contributes to establishing newly formed interactions.

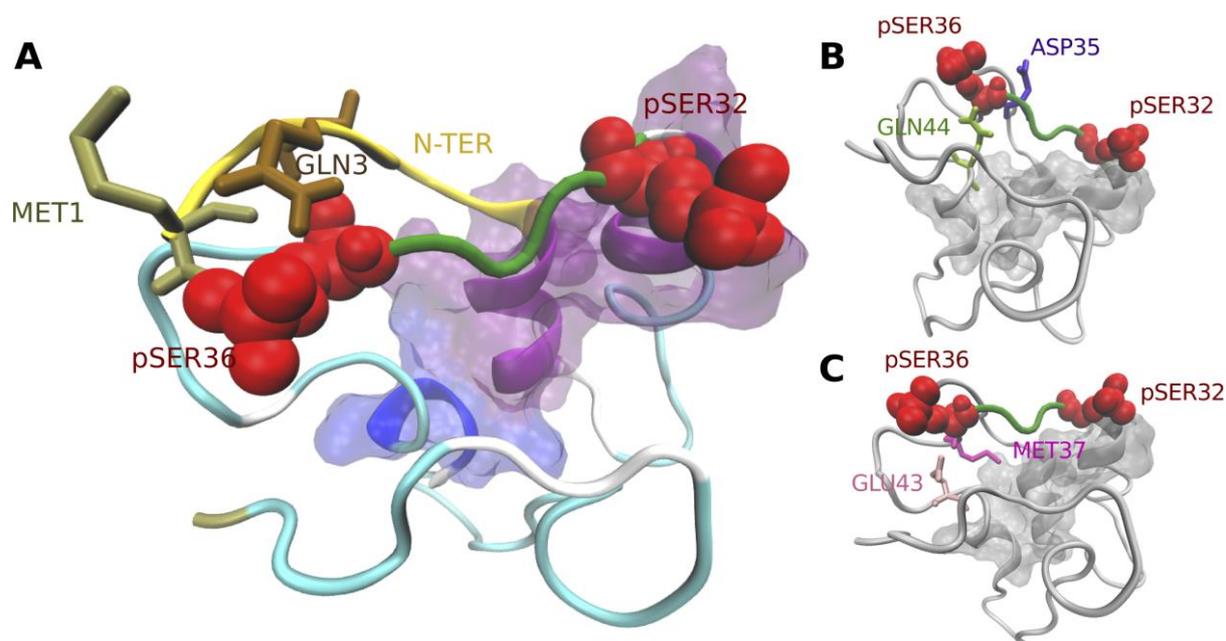


Figure II.6 (A) Cartoon representation of double-phosphorylated $I\kappa B\alpha^{1-70}$ color coordinated according to secondary structure elements. The green shaded segment flanked with the red colored spherical phosphoserines denotes the degron in $I\kappa B\alpha$. The N-terminal, depicted in yellow, is shown to interact closely with pSER36 that forms hydrogen bonds with residues Met1 and Gln3 situated right by the N-terminal tail. Additional unique hydrogen bonds formed upon phosphorylation in the vicinity of the phosphorylation site involve residues Asp35 and Gln44 (B) and Met37 and Glu43 (C). The hydrogen bond interactions deemed significant are present in at least 30 % of the simulation period in at least 2 out of 3 of the replicate runs.

Residues Met1 and Gln3 located at the tip of the N-terminal shift away from pSer32 and closer to pSer36 and form unique hydrogen bonds: The backbone amide of Met1 interacts with the phosphate group of pSer36, whereas both the backbone and side chain amides of Gln3 form hydrogen bonds with the backbone carbonyl group of pSer36; these shifts together with other local conformational changes lead to the stabilization of the N-terminal tail and concomitant other regional changes in the structure of $I\kappa B\alpha$ (Figure II.6 A). Additional residues that come to form hydrogen bonds in vicinity of the phosphorylation site are Asp35/Gln44 (Figure II.6 B) and Met37/Glu43 (Figure II.6 C). The backbone carbonyl of Asp35 engages in hydrogen bond interaction with the side chain amide of Gln44, and the

backbone amide of Met37 forms a similar interaction with the side chain carboxyl group of Glu43. Interestingly, Asp35 and Met37 both enclose the pSer36 phosphorylation site, and the adjoining residues Gln44 and Glu43 are localized on the third α -helix in the SRD of $\text{I}\kappa\text{B}\alpha$. This newly formed interaction pattern present uniquely in the double-phosphorylated while absent in the unphosphorylated states, highlights the role that phosphorylation possesses in establishing stability and order in the region, as reflected previously in the lower RMSF exhibited by the double-phosphorylated state. Specifically, it appears that phosphorylation of Ser36 is the principal contributor to this effect.

6.3.5 Electrostatic effects

Part of the effects of phosphorylation may be explained by differences in electrostatic potential surrounding the site of phosphorylation. The degron motif in $\text{I}\kappa\text{B}\alpha^{31-36}$ holds the sequence of residues DSGLDS, which encompasses the integral site of phosphorylation Ser32 and Ser36 each of which are preceded by an acidic residue, aspartic acid. The electrostatic potential of this region is illustrated in Figure II.7, with red colored surface correlated with potentials of -10 kT and blue colored surface correlating with potentials of +10 kT.

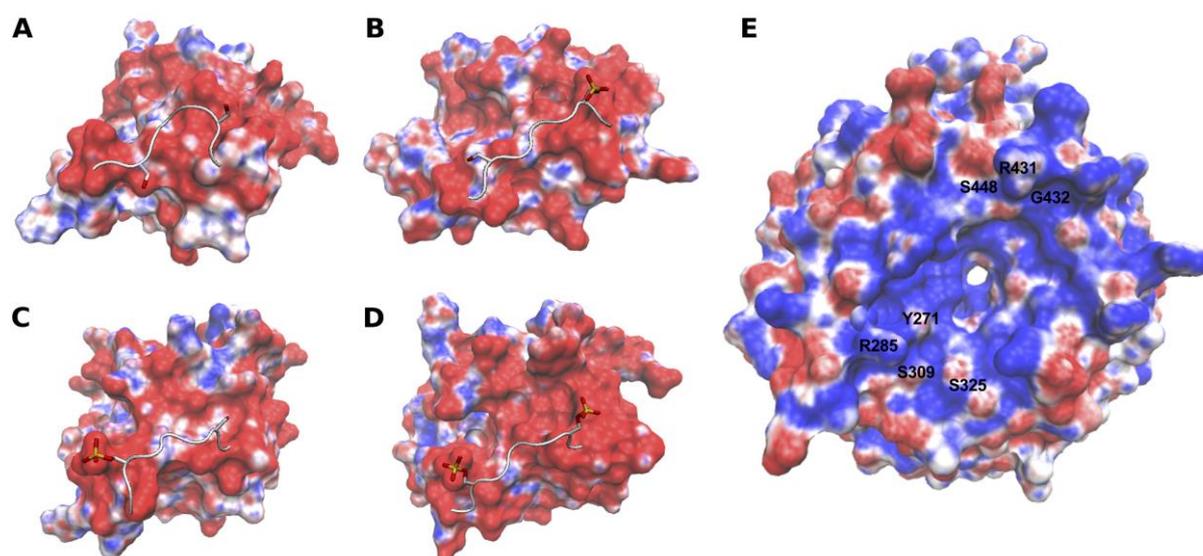


Figure II.7 Electrostatic potentials of the SRD mapped onto the van-der-Waals protein surface of $\text{I}\kappa\text{B}\alpha$ calculated by APBS [199] in the $\text{nI}\kappa\text{B}\alpha$ (A), $\text{pSer36I}\kappa\text{B}\alpha$ (B), $\text{pSer32I}\kappa\text{B}\alpha$ (C) and the $\text{ppI}\kappa\text{B}\alpha$ states (D). The position of the site of phosphorylation is indicated by the cartoon figure of segment $\text{I}\kappa\text{B}\alpha^{31-37}$. (E) The electrostatic surface potential on the WD40 domain of β -TrCP, showing a top view of the binding interface to a double-phosphorylated $\text{I}\kappa\text{B}\alpha$. Negative potentials of -10 kT are depicted in red, and positive potentials of +10 kT are depicted in blue. The residues making intermolecular contacts with the phosphoserines in the degron motif are indicated in their respective positions.

In the unphosphorylated state (Figure II.7 A), the negative potential patches at or near the protein surface surrounding the sites of phosphorylation are fostered by the presence of strong acidic residues, namely Asp27, Asp28, Asp31 in the α -helical region

preceding Ser32 and residues Asp39, Glu40, Glu41, Glu43 that are localized on the loop segment following Ser36. In the double-phosphorylated state, the previously negatively charged patch grows even stronger covering a more extended part of the protein surface (Figure II.7 D). The introduced negative charges upon phosphorylation, jointly with induced structural rearrangements alter the distribution of surface charges potential in a visible manner creating an even more pronounced negatively charged protein surface. The mono-phosphorylated states, as displayed by Figures II.7 B and C, do not exhibit the same electrostatic effects as their double-phosphorylated counterpart. Since, double-phosphorylation is a prerequisite for β -TrCP binding, the electrostatic surface potential of the binding area of β -TrCP is shown (Figure II.7 E). The surface potential of the top narrow part of the channel being the binding surface reveals an extensive positive blue colored patch indicative of a basic environment. The electrostatic complementarity of double-phosphorylated I κ B α and β -TrCP protein surfaces may be the critical recognition mechanism to initiate the formation of the tertiary protein complex.

6.4 Discussion

Phosphorylation is the most prevalent post-translational mechanism regulating protein function throughout the cell. Protein kinases carry out phosphorylation by modifying existing serine, threonine and tyrosine side chains with the addition of a phosphate group. At physiologically relevant pH, a phosphate group usually carries a -2 charge. Introducing such a negative charge often leads to electrostatic perturbations directly affecting protein energy landscapes, which exercise control over protein-protein interactions and conformational dynamics [200]. As with many other cellular functions, it is challenging to derive an atomic-level understanding of how phosphorylation alters protein structure and function. Computational methods, such as all atomistic MD simulations, are a major contributor in filling this knowledge-gap and throwing light on the structural changes upon protein phosphorylation (as reviewed in [200]).

Many experimental studies on the importance of I κ B α phosphorylation have been done in the past [47,188-192]. Yet little is known about the molecular details, which render this post-translational modification indispensable for recognition by the SCF complex. With MD simulations, I was able to characterize local structural aspects induced by phosphorylation of Ser32 and Ser36. Although no striking differences were observed in secondary structure elements upon phosphorylation, aspects of local stability, in the progression of disorder to order surrounding the sites of phosphorylation were evidenced. Double-phosphorylation leading to the uniform ordering of the segment I κ B α ²⁸⁻⁴⁰ was in part brought forth by the stabilization of residues preceding the initial phosphorylation site, namely residues 28-32. Moreover, increased N-terminal stabilization induced by forming hydrogen bond interactions with the phosphorylation site was another direct effect of double-phosphorylation. In contrast, the unphosphorylated I κ B α experienced a more variable structural stability across this region. Post-translational phosphorylation has often previously been associated with sparking both ordering and disordering in proteins. For the degron of I κ B α in complex with NF-kappaB, we observe an opposite effect which was also observed, for example for smooth muscle myosin [201]. Such a residue-to-residue decrease in phosphorylated serine amino acid residues was also observed in the α -amino-3-hydroxy-

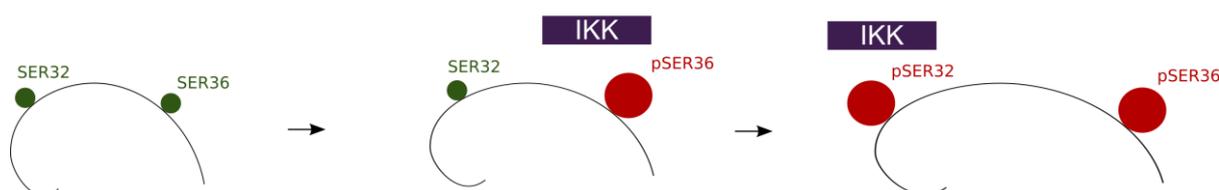
5-methyl-4-isoxazolepropionic acid (AMPA) receptor and interpreted as a coil-to-helix compressed secondary structure change [202].

A Monte Carlo/Stochastic Dynamics simulations study revealed how electrostatic interactions navigated the stabilization of a helical conformation at the N-terminus upon phosphorylation in model peptides consisting of a serine in the N-terminus followed by nine alanines [203]. Via docking methods, phosphorylation of the phenylalanine hydroxylase (hPAH) demonstrated in another study an increase in stability of the N-terminal tail through local conformational changes as a result of electrostatic interactions [204]. In an investigation employing the Car-Parrinello *ab initio* MD method, phosphorylation of the CREB-CBP complex involved in activation of the DNA transcription machinery results in stability of the complex by a new hydrogen bond interaction [205]. In another study employing molecular mechanics-based methods on kinases and prokaryotic response regulators, local conformational changes specific to areas in close proximity to the phosphorylated amino acid were observed [206]. Structural transitions between order and disorder induced by phosphorylation have also been mapped by previous experimental work. An order to disorder transition of a helix containing a phosphorylated serine was documented in the Oncoprotein 18/stathmin (Op18) phosphoprotein [207], in contrast to a disorder to order adaptation of the protein kinase B/Akt which, after phosphorylation, experiences disruption of the α C-helix causing global restructuring of the protein [208].

The unphosphorylated state simulations in this study indicated a bent shape between the two serine residues. However, an impact of double-phosphorylation was the structural reconfiguration of the segment connecting the phosphoserines resulting in an extended conformation and verified by a larger Ser32^{C α} -Ser36^{C α} distance. Findings in a study by Pons *et al.* [209] proposed a random coil conformation withholding a weaker bend of a free phosphorylated 24 amino acid segment, I κ B α ²¹⁻⁴⁴ characterized by NMR spectroscopy and MD simulation, in contrast to the conformation of the bound segment to β -TrCP obtained by transfer nuclear Overhauser effect spectroscopy (tr-NOESY) presenting an apparent bend between the two serines. The occurrence of a well-defined bend connecting pSer32 and pSer36 in the bound structure could be initiated by structural rearrangement upon binding to the ubiquitin ligase.

A distinct pattern of conduct of the phosphorylated serines, revealed by the simulations, is indicative of the individual roles played by each serine residue. I propose that phosphorylation is commenced at site Ser36, which in turn increases the solvent surface exposure of Ser32 increasing its accessibility to the kinase IKK, although partially burying pSer36 (Scheme II.1).

Hypothetical model of I κ B α sequential phosphorylation



Scheme II.1 A hypothetical model of the sequential phosphorylation of I κ B α by the protein kinase IKK. Based on these simulations, Ser36 is phosphorylated first by the IKK, which in

leads to an increased solvent accessible surface area of the second phosphorylation site, Ser32. With an enhanced exposure to the solvent, Ser32 is in turn phosphorylated by the kinase, with the effect of an extended structural loop conformation of the degron segment.

This relative partial burial of pSer36 leads to a new hydrogen bond network stabilizing the N-terminal segment of I κ B α and the region proximal to the pSer36 phosphorylation site. With pSer36 contributing to local structural stability of the protein, pSer32 with a relatively large SASA is able to interact freely with β -TrCP. This view is supported by the resolved crystal structure of the bound β -TrCP to β -catenin showing how pSer33 (homologous to pSer32 in I κ B α) is the residue that makes the largest number of contacts with β -TrCP, that is residues Tyr271, Ser309, Ser325 and Arg285, while pSer37 (homologous to pSer36 in I κ B α) forms comparatively fewer interactions with residues Ser448, Gly432 and Arg431 (Figure 1.7 E) [56]. These residues are located on the rim at opposite sides of the channel. Moreover, molecular docking data of a phosphorylated 11 amino acid I κ B α peptide bound to β -TrCP demonstrated that pSer32 establishes the same interactions as β -catenin [210].

Long-range electrostatic effects are a major determinant of protein-protein recognition and association. The electrostatic complementarity of protein surfaces was identified to be a major regulator of protein-protein complex formation, see for example [211-213]. Post-translational introduction of phosphate groups carrying a double negative charge by IKK enhances the negative electrostatic potential in the double-phosphorylated state of I κ B α , an effect remarkably different from the unphosphorylated and the mono-phosphorylated states. The elicited negatively charged protein environment adjacent to the phosphorylation sites is also of critical importance in the recognition by the β -TrCP SCF complex.

Chapter H

Study III

H.1 The Molecular Basis of Polyunsaturated Fatty Acid Interactions with the Shaker Voltage-Gated Potassium Channel

Polyunsaturated fatty acids (PUFAs) being essential parts of cell membrane phospholipids of heart cells and neurons [214] have been found to shift the voltage dependence of the *Shaker* voltage-gated K⁺ (K_v) channel to enable channel activation via a proposed electrostatic mechanism [108,128]. While the fatty acid modulatory effects have been extensively studied, in particular on voltage-gated ion channels [215], the molecular mechanism by which PUFAs modulate channel functioning remains an open question.

PUFA molecules are embedded within the lipid bilayer, and the modulation of K_v channel activity can in principle occur via indirect or direct effects, or a combination of the two. For example, addition of PUFAs to cardiac myocytes blocked sodium currents in Na⁺ channels by increasing the membrane fluidity, which is suggestive of an indirect effect [104]. On the other hand, direct PUFA interactions have been observed for several voltage-gated ion channels [103,105,107,125-127] and are often inferred by exclusion of indirect lipid bilayer effects. In this way, direct PUFA-mediated inhibition have been observed in K_v1.5 channels [103], Na_v channels [125,127], and Ca_v channels [126] and both activation and inactivation on the K_v1.5 and K_v2.1 channels [105]. In addition, mutational analyses pinpointed direct effects of PUFAs to inhibit Na_v channels targeting the S6 helix in the pore [107] and to activate K_v channels via the extracellular part of the VSD [129].

Because K_v channels are integrated components of the nervous system, their malfunction is often connected to disease. As a result, refractory epilepsy [110-112], which is triggered by K_v1 -type or K_v7 -type channel malfunction [119-124], can effectively be treated by ketogenic diets containing a high PUFA content [108,110]. Even though ketogenic diets have been prescribed to patients since the 1920s [216], the underlying mechanisms by which it operates and prevents the epileptic seizures remain elusive. However, given the wide scope of PUFA effects on a range of voltage-gated ion channels, the mechanism of ketogenic diets likely include PUFA-channel interactions.

The PUFA head group (Figure III.1 A) was observed to activate *Shaker* K_v channels, which is referred to as the lipoelectric mechanism [128,217] and can be abolished or reversed by neutralization of the negative PUFA head group charge or introducing a positive charge [128]. Similarly, the acyl chain properties have also proven important. While the acyl tail length does not seem to be critical, the number and geometry of the double bonds have significant effect on the modulatory properties of the PUFA [128], e.g. a minimum of two double bonds, particularly in the *cis* arrangement, substantially increase channel currents. PUFAs have been reported to modulate the gating mechanism of the *Shaker* K_v channel by partitioning into the lipid bilayer and interacting with the extracellular halves of the VSD helices S_3 and S_4 [129]. Specifically, a series of cysteine point mutations on the S_3 , S_4 , S_5 , and S_6 helices revealed that residues on helices S_3 - S_4 altered the sensitivity to the docosahexaenoic acid (DHA) PUFA. In addition, introducing positively charged cysteine-specific MTSEA⁺ probes identified four high-impact residues on the lipid-facing side of the VSD cytoplasmic region; I325 and T329 (helix S_3) and A359 and I360 (helix S_4) (Figure III.1 B). Indeed, channel VSDs make significant contacts with the surrounding lipid bilayer [97,98,218] and MD simulations show specific interactions between the gating-charge R1 and R2 arginines on helix S_4 on $K_v1.2$ channels to salt-bridge to lipid headgroups [85]. In addition, S_4 gating charges have been observed to interact with lipid head groups in open, resting and intermediate states of the $K_v1.2$ and the paddle- $K_v1.2$ chimera channels [219-221]. Finally, the S_3 b- S_4 paddle in VSDs of K_v channels has been identified as the key interaction point between the lipids and the channel [222]. Hence, a picture emerges where charged PUFAs partitioned into the lipid bilayer exert modulatory effects by contributing specific interactions with the VSD domains of ion channels.

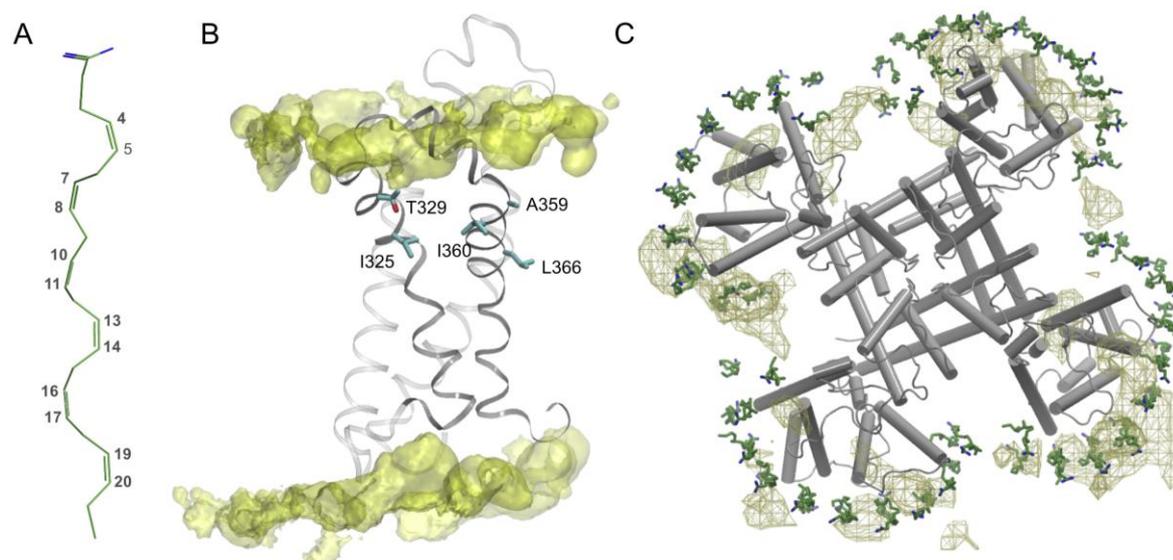


Figure III.1 Molecular PUFA-VSD interactions in the *Shaker* channel. (A) Close-up of a PUFA (DHA) in its initial conformation. The numbers marked in grey depict the carbons forming the *cis* double-bonds. (B) Side-view of one VSD equilibrated in a POPC lipid bilayer represented by a yellow iso-density surface corresponding to the positions of lipid nitrogens in the simulation at 5 % occupancy. The residues shown in experimental studies to be close to the interaction site of DHA, namely residues I325, T329 located on S₃, and A359, and I360 located on S₄, are colored in cyan [129]. (C) Top-view of the *Shaker* tetramer with PUFAs in their starting positions. The PUFA carboxyl head group and carbon tail are colored in blue and green, respectively. The simulated dynamics of the PUFAs surrounding the *Shaker* tetramer is represented by a brown mesh iso-density surface at 27 % occupancy. The cut-off was chosen to visualize the differences between the PD and VSD interactions with the PUFAs.

This study characterizes PUFA-K_v channel interactions in the open and closed states of the channel using atomistic MD simulations. The study explores the interaction between the *Shaker* K_v channel in an open state and a PUFA-enriched lipid bilayer and specifically characterized PUFA enrichment regions on the VSD. The open state *Shaker* K_v channel was modeled based on the high-resolution experimental structure of the chimera channel K_v2.1/K_v1.2 [82]. A potential PUFA-K_v channel site of interaction was found to be located on the lipid-facing side of a pocket connecting the extracellular halves of S₃ and S₄ helices, which is supported from experiments. In general, the lipophilic PUFA tail covered a wide range of non-specific hydrophobic interactions along helices S₃ and S₄, while the carboxylic head group formed fewer and more specific electrostatic interactions with the top regions of the S₃-S₄ helices and the S₃-S₄ linker. In addition, by performing simulations of saturated fatty acids (SFA)-K_v channel systems, the prerequisite of a polyunsaturated carbon tail is explained by simulations as suppression of flexibility in a saturated carbon tail. The closed state simulations revealed an interaction pattern in which both the PUFAs and SFAs formed fewer interactions compared to the corresponding open state simulations. Together, the results explain the selective stabilization of the open state of a K_v channel, identify a putative PUFA interaction site at atomic detail and thereby provide novel K_v

channel interaction points that can be tested experimentally and aid in the design of pharmaceutical compounds for the treatment of epilepsy.

H.2 Methods

H.2.1 Shaker channel system setup in open state

The K_v *Shaker* channel has been extensively studied with respect to PUFA modulation (gating function) [108,128,129]. Because of a 66 % sequence identity in transmembrane (TM) helices S₀-S₆ between the *Shaker* K⁺ channel and the high-resolution X-ray structure of the $K_v2.1$ paddle- $K_v1.2$ chimera channel (PDB ID 2R9R) [82], the structural model of the *Shaker* in an open state was based on this structure. A *Shaker* channel with partially truncated S₃-S₄ linker (residues 337-350) has been shown to have similar biophysical properties to the wild-type channel [223,224]. Therefore, I opted for removing this extended loop, which is also missing in the chimera crystal structure. The three-dimensional structure of the *Shaker* channel was constructed with Modeller 9.12 [136]. Initially, five models were built and the optimal model was identified using a combination of the following model quality scores: The Discrete Optimized Protein Energy (DOPE) score, a statistical potential used to assess homology models [225], and the GA₃₄₁ score, a method used for model assessment based on the percentage sequence identity between the template and the model [226].

The ion channel was immersed in a pure POPC bilayer [227] consisting of 256 lipids per leaflet using the *g_membed* tool in the GROMACS tool repository [228]. After inserting the protein into the bilayer 438 POPC lipids remained. The system was energy minimized with steepest descent until the maximum force reached < 1000 kJ/mol/nm and was subsequently equilibrated for 10 ns while keeping the protein and the waters position restrained ($F_c = 1000$ kJ/mol/nm²). With the position restraints removed, the system was relaxed in a 50 ns simulation. The Amber ff99SB-ILDN force field [229] was used for the construction of the membrane-protein system, in combination with the Berger force field [227] for the lipids and the TIP3P water model [230] and the CHARMM36 force field [231] was used for the 1 μ s production run.

H.2.2. PUFA and SFA system setup in open state simulations

Force field parameters for the PUFA (docosahexaenoic acid), SFAs (docosanoic acid) and the MTSEA⁺ compounds were obtained from the multipurpose atom-typer for CHARMM (MATCH) server [232], which utilizes libraries of topology and parameter files in existing force fields for extrapolation to the new molecules consistent with the parameterization strategy within a given force field. The default Charmm General Forcefield (top_all36_cgenff) was used for deriving partial charges and parameterization. The calculated order parameters for DHA (Figure III.2 A) are in agreement with those obtained previously both by computational and experimental efforts [233]. In addition, the order parameters for the SFA agree with the increased level of disorder towards the end of the

chain reported for the saturated palmitic acid [233]. Together, these comparisons validate the parameterization procedure.

The simulations were performed using the all-atom CHARMM36 force field [231] and a development version of Gromacs [170,171]. The LINCS algorithm [173] was applied for constraining bond lengths. Electrostatic interactions were calculated with the Particle-Mesh Ewald algorithm at every step [174]. P-LINCS [234], a non-iterative parallel constraints algorithm allowing replacement of hydrogens with virtual interaction sites, enabled 5 fs time steps [140]. A 1.2 nm cutoff was used both for electrostatics and van der Waals interactions, with neighborlists updated every 10 steps. The simulations were performed at constant pressure of 1.0 bar with Parrinello-Rahman pressure coupling [197] and the semiisotropic pressure scaling, time constant of 2.0 ps, and a system compressibility of $4.5 \times 10^{-5} \text{ bar}^{-1}$. The temperature of the system was maintained at 300 K using the velocity-rescaling algorithm [198].

PUFAs were placed around each VSD on each lipid leaflet (Figure III.1 C) and clashing POPC lipids were removed resulting in 80 PUFAs and 275 POPC lipids surrounding the tetrameric *Shaker* channel. The system was subsequently solvated with 33,389 TIP3P waters and 192 sodium and 104 chloride ions replaced water molecules to neutralize the net system charge (and obtain a salt concentration of 0.1 M). The total number of atoms in the system reached 163,139. Initially, the system was equilibrated for 30 ns while keeping the protein, waters, and PUFAs frozen. Position restraints were then applied to the protein, PUFAs, and z-coordinates of the waters followed by 100 ns equilibration. Hereafter, all position restraints were removed except for the PUFAs and the system was allowed to equilibrate for an additional 10 ns. This series of equilibration steps allowed POPC lipids to pack around the PUFAs while preventing water molecules from penetrating into the membrane bilayer. The SFA-channel system was created by replacing the PUFAs in this equilibrated system with SFAs followed by a 200 ps equilibration step.

Two initial PUFA distribution patterns were tested. In the first, 32 PUFAs were evenly distributed across the lipid bilayer (Figure III.3 A) and sampling was performed in a 5 μs production run. In the second setup, center of mass (COM) pulling was applied using an umbrella potential between each VSD and the surrounding PUFAs in order to allow for close (but non-specific) protein-PUFA packing. Here, the PUFAs were pulled in the x and y dimensions with a harmonic force constant of $1000 \text{ kJ mol}^{-1} \text{ nm}^{-2}$ at a rate of 0.08-0.01 nm/ps depending on the initial distances between the PUFAs and the VSD followed by a 1 μs production run. The COM pulling procedure was repeated to create the SFA system.

H.2.3 Modeling of Shaker channel in closed state

A model of the *Shaker* channel in a closed state was built based on a previous Rosetta model of the channel in the C3 state [235] with a partially truncated S3-S4 linker using Modeller 9.12 [136]. This closed-state model replaced the open state channel in the previously set up PUFA and SFA systems. The new system configurations were relaxed for 10 ns followed by COM pulling to position PUFA and SFA molecules in initial positions close to the channel and subsequent 1 μs production simulation runs.

To assess the structural integrity of the structures, the root mean square deviation (RMSD) of the protein backbone was calculated against the average structure for all four systems (PUFA and SFA channel systems in open and closed states) and compared to a

channel-only system. Overall, the channels maintain stable structures, settling around an RMSD of 1.5 – 2.0 Å throughout the simulation time.

H.2.4 In silico mutagenesis of the VSD

Residues I325, T329, A359, I360, and L366 were mutated to cysteines and modified to include the reagent MTSEA⁺ to characterize PUFA interaction sites. The mutations were carried out using PyMOL (The PyMOL Molecular Graphics System, Version 1.3 Schrödinger, LLC). Each mutated system was minimized with a steepest descent algorithm until a maximum force of < 1000.0 kJ/mol was reached followed by a short 200 ps equilibration. Production runs were 1 μs long for each of the MTSEA⁺ mutated systems, except for the L366 system which was simulated for 500 ns. In addition, the MTSEA⁺ modified systems were mutated back to cysteines after 500 ns of simulations and continued for 500 ns.

H.2.5 Small-scale PUFA and SFA systems

Two membrane patches were built using the membrane generator MemGen [236] with 24 lipids per leaflet. In each of the bilayer patches, DHA or DA, were inserted in each lipid leaflet and were set up as the described PUFA system. The systems each contained ~12,000 atoms with roughly 48 POPC lipids, 1,660 water molecules and 20 sodium and chloride counter ions. Each system was subjected to steepest descent energy minimization until the maximum force reached a value < 1000 kJ/mol/nm followed by 1.5 ns of equilibration. Production runs were 1 μs for each system configuration.

H.3 Results

In a set of independent simulations, I explored dynamics of polyunsaturated (docosahexaenoic acid, DHA) (Figure III.1 A) and saturated (docosanoic acid, DA) fatty acids in neat lipid bilayers as well as bilayers containing the *Shaker* K_v channel in both its open and closed states.

H.3.1 Saturation levels in the carbon tail affect the structural dynamics

To determine how saturation levels in the fatty acid tail influenced the general structural dynamics, single PUFA (DHA) and SFA (DA) molecules in two separate 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) membrane patches were simulated. As expected, the observed order parameters for both the PUFA and SFA molecules decreased gradually from the carboxyl end located in the bilayer interface to the end of the acyl chain tail in the bilayer center (Figure III.2 A). While the order parameters were quite similar near the carboxyl head and at the methyl end of both fatty acids, the PUFA order parameters were generally lower and displayed a different overall shape reflecting the positions of the

cis double-bonds. Thus, the PUFA molecule exhibited greater conformational mobility compared to SFA in the POPC membrane patch.

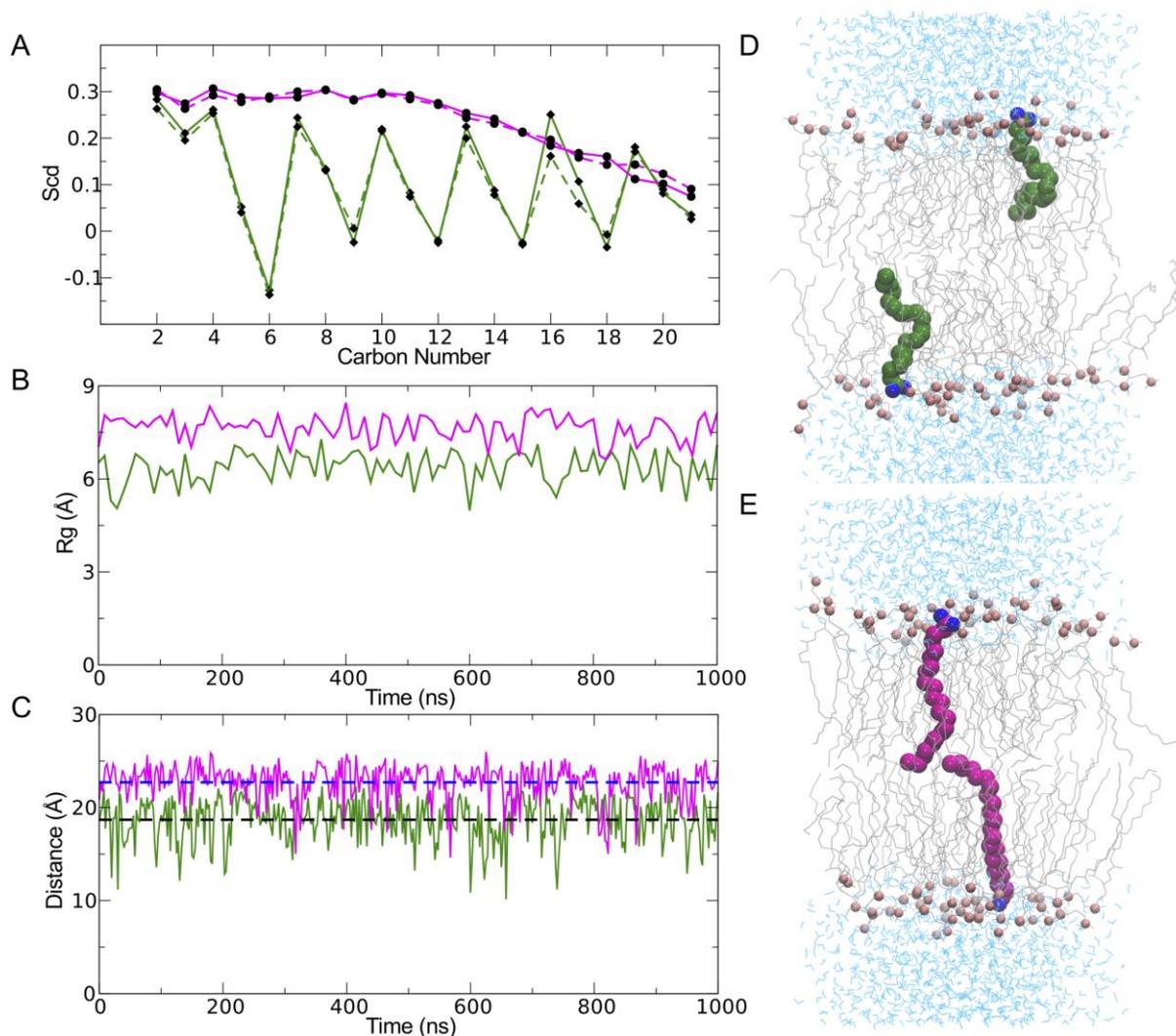


Figure III.2 Structural flexibility in lipid-partitioned PUFA and SFA. (A) Deuterium order parameters of PUFA (green) and SFA (purple) carbon chains. (B) The radius of gyration of PUFAs (green) and SFAs (purple). (C) Distances between the head group oxygen and the last carbon in the PUFA (green) and SFA (purple) chains. The dotted lines denote the overall average distance for the PUFAs (black) and the SFAs (blue), respectively. Representative structures from 1 μ s simulations of PUFAs (D) and SFAs (E) are shown embedded in POPC bilayers.

Furthermore, I characterized differences in the overall shape and packing properties by measuring the radius of gyration of the carbon tail and head-to-tail length. The polyunsaturated chains displayed a significantly lower radius of gyration as compared to the fully saturated acyl chains (Figure III.2 B). Similarly, the distance between the carboxyl head group oxygen and the final methyl carbon in the tail ($O-C_{22}$ distance) was 5 Å shorter in PUFAs compared to the SFAs (Figure III.2 C). In addition, the fluctuations in the $O-C_{22}$ distance were more pronounced in the polyunsaturated carbon chains with extreme values

between 10 Å and 20 Å. Hence, while the PUFA molecule tends to twirl and curl up (Figure III.2 D), the average SFA conformation is more extended (Figure III.2 E).

H.3.2 PUFA interactions with the K_v Shaker channel in the open state

To characterize PUFA-protein interactions in the membrane, I initially distributed 32 PUFAs evenly across a POPC bilayer containing the *Shaker* tetramer, 16 on each bilayer leaflet (Figure III.3 A) and simulated for 5 μ s (Figure III.3 B). Amino acid residues that resided within 3.5 Å of the PUFAs for more than 300 ns were mapped out in order to visualize the region of interaction between the channel and PUFAs on the outer leaflet of the membrane. The observed interactions differed significantly between the PUFA head and tail groups with the PUFA tails exhibiting approximately twice the number of contacts compared to the head groups. The contacting residues were distributed on the S₁-S₂ and S₃-S₄ segments of the VSD. Residues close to the PUFA head groups were either charged or polar and located on the outer regions of the VSD whereas residues contacting tail groups were predominantly hydrophobic and covered the inner parts of the S₂ and S₃ helices (Figure III.3 C). While this approach provided us with an initial idea of the PUFA interaction patterns, translational diffusion might prevent identification of realistic PUFA-channel interactions.

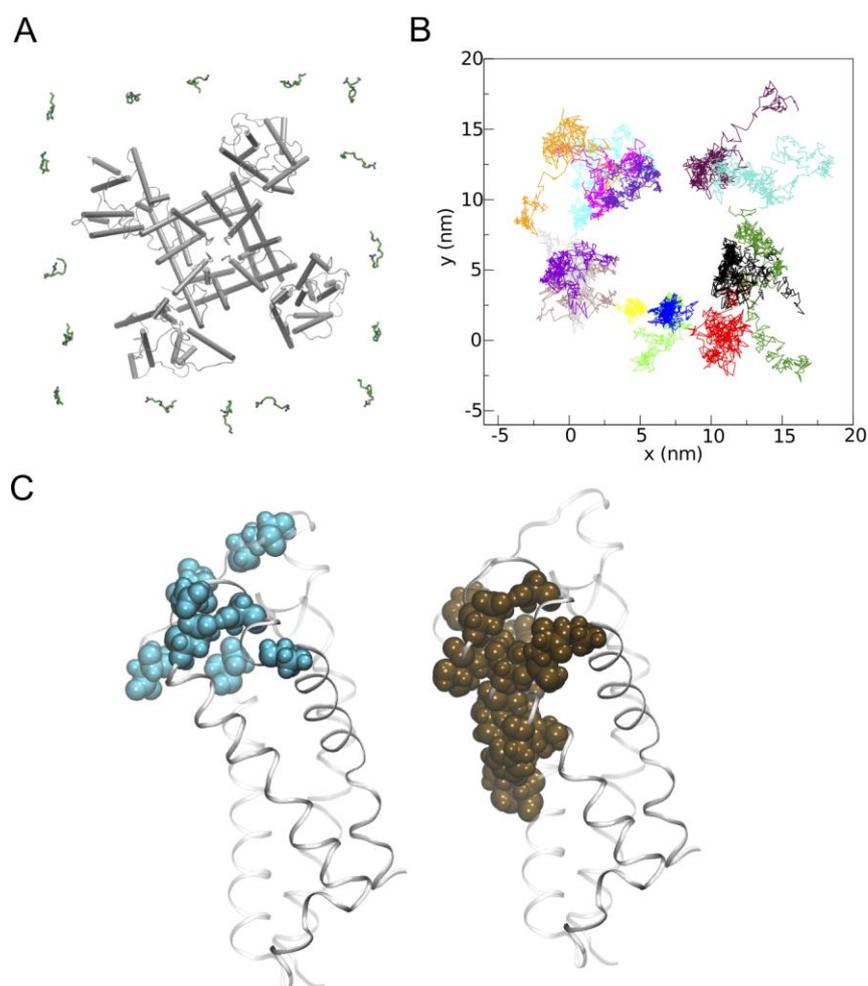


Figure III.3 PUFA- *Shaker* interactions in a 5 μ s MD simulation. (A) Top-view of the *Shaker* tetramer with 16 PUFAs on each leaflet of the membrane. The PUFA carboxyl head group and carbon tail are colored in blue and green, respectively. (B) Lateral displacement (x, y dimensions) of PUFAs in the membrane bilayer over 5 μ s with the protein COM centered in the box. (C) Interacting VSD residues presented separately for the PUFA carboxyl head groups (blue) and carbon tails (brown).

To increase sampling, I opted for a new system configuration starting with PUFAs packed around the VSD of each monomer in a semi-circular fashion (Figure III.1 C) using steered MD simulations. Starting from these initially closely packed positions, extensive sampling would in principle allow differentiation between specific and non-specific binding by monitoring PUFA structural dynamics given a PUFA diffusion constant of $4 \times 10^{-9} \text{ cm}^2/\text{s}$ observed in the simulation. The resulting simulation trajectory showed the PUFAs to form clusters in the vicinity of helices S_3 and S_4 across all VSD subunits rather than pore helices S_5 - S_6 (Figure III.1 C, green mesh surface). To identify interaction patterns, amino acid residues that resided within 3.5 \AA of the PUFAs in the outer lipid leaflet for more than 300 ns were compared between the PUFA carboxyl head groups and carbon tails (Figure III.4). All the reported contacting residues were positioned on the VSD and not on the pore domain. In addition, the interactions were significantly different between the PUFA head group and tail. While the PUFA carboxyl head groups engaged in fewer contacts but with

higher contact frequencies (Figure III.4 A), the tails displayed a wider range of contacts occurring with lower frequencies (Figure III.4 B).

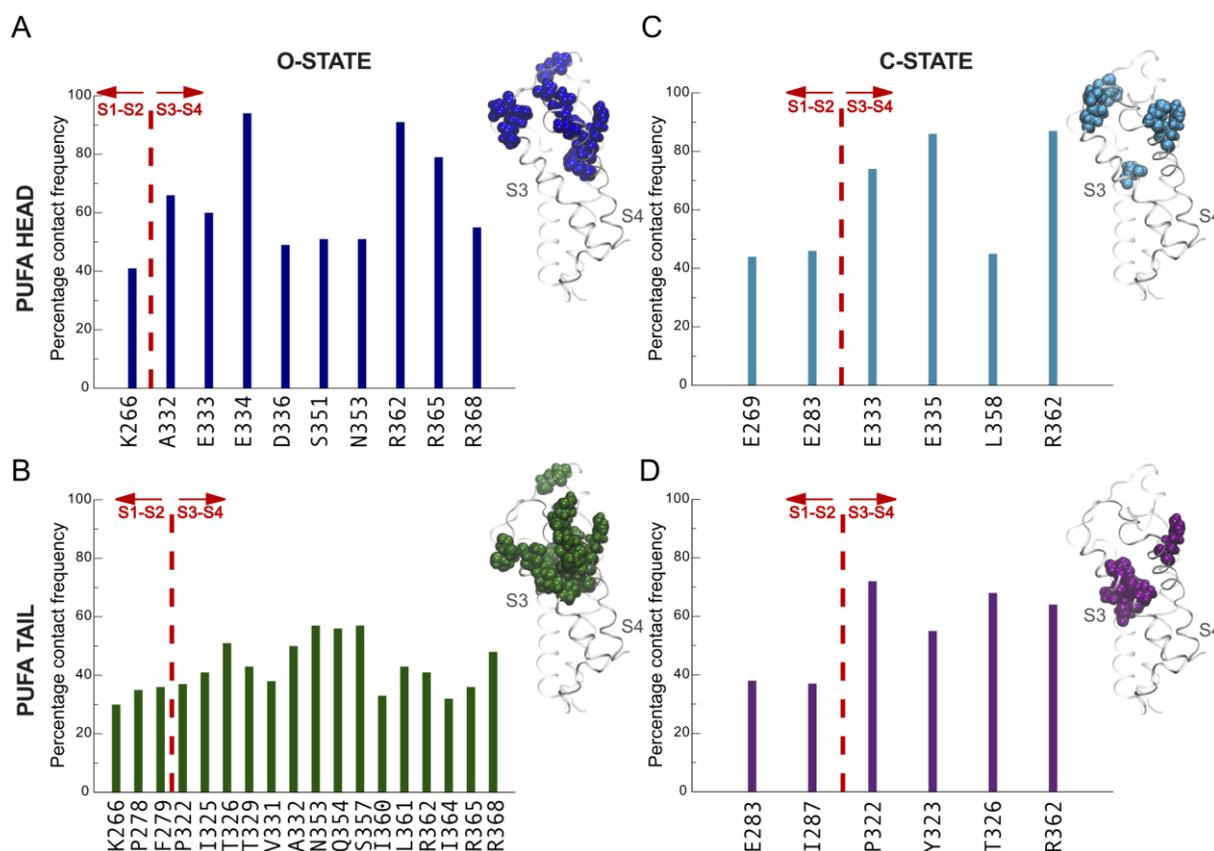


Figure III.4 PUFA contacts to *Shaker* tetramer residues in the open and closed states. The contact frequencies of amino acid residues within 3.5 Å of PUFA carboxyl head groups and carbon tails are displayed for the open (A and B) and closed (C and D) states. The red dotted line differentiates contacts on helices S1 and S2 or helices S3 and S4 of the VSD. Side-view of a VSD and interacting residues are displayed separately for the PUFA carboxyl head groups and tails for each state of the channel (insets).

Protein residues contacting the PUFA tails were predominantly hydrophobic and distributed across the extracellular halves of the S1-S2 loop and helices S3 and S4 with 85 % of the contacting residues localized on the extracellular halves of S3 and S4 (Figure III.4 B). The PUFA tail interactions to the protein displayed lower contact frequencies compared to the head groups, which indicates non-specific lipophilic interactions. In contrast, the majority of the contacts between the PUFA head groups and the protein were located on the top parts of the extracellular halves of S3 and S4, incorporating the S3-S4 linker and were either charged or polar residues displaying high contacting frequencies (Figure III.4 A). Specifically, the S3-S4 linker residues S351 and N353 formed hydrogen bonds with the PUFA head groups. In addition, the negatively charged carboxyl groups of residues E333, E334, and D336 located on the S3-S4 linker interacted with the negatively charged PUFA head groups mediated by Na⁺ ions. Presence of 100 mM NaCl neutralized the simulation system and sodium ions are typically observed in the vicinity of the protein. An additional PUFA-protein electrostatic interaction with relatively high contact frequency is mediated by R362, which is one of the gating charges (R1). Parenthetically, I also observed

interactions between the S₄ gating charges R365 (R₂), and R368 (R₃) and one PUFA head group. These interactions were enabled by a nosediving movement of that particular PUFA molecule into the water-filled extracellular crevice. Because this PUFA-protein interaction appeared only once in our simulations, I will not ascribe significant physiological relevance to the interaction per se, but rather point out that PUFA interactions to the gating charges are structurally possible in a dynamic system.

H.3.3 PUFA interactions with the K_v Shaker channel in the closed state

To investigate the PUFA interactions with the K_v Shaker channel in the closed state. I generated a system consisting of a closed-state model with PUFA molecules in similar starting configurations as for the open-state system and simulated for 1 μ s. The observed packing environment differed in-between PUFAs in the open and closed states of the channel as displayed by the results of contact analyses. In general, significantly fewer PUFA interactions were made in the closed-state simulation with an equal distribution of the number of contacts between the head and tail groups (Figure III.4 C and D). In addition, only half of these contacts were observed in the open-state simulation. To further differentiate the PUFA interaction pattern between the open and closed states of the channel, I measured minimum distances between different segments of the fatty acids and the channel (Figure III.5). The major differences in minimum distances to the channel were found in the head groups with PUFA carboxyl head groups being 0.5 ± 0.2 Å closer in the open state. In contrast, no significant variations between open and closed states were observed in the minimum distances between the channel and different sections of the PUFA tails.

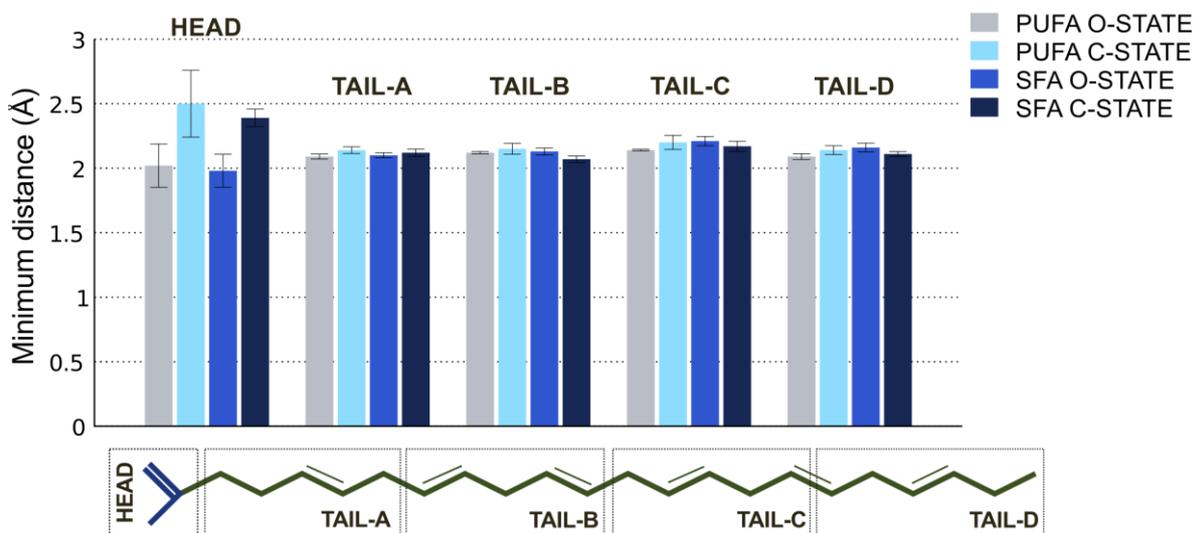


Figure III.5 The average minimum distances between the channel in its open and closed states and the PUFAs and SFAs. The fatty acids were sectioned into five parts consisting of carbons C₂₋₆ (TAIL-A), C₇₋₁₁ (TAIL-B), C₁₂₋₁₆ (TAIL-C), C₁₇₋₂₂ (TAIL-D), and the head group (HEAD). The error bars indicate the standard error of the minimum distances across the four subunits of the channel.

H.3.4 SFA interactions with the K_v Shaker channel in the open and closed state

Because SFAs do not exert the modulatory effects of PUFAs, I explored SFA interactions with the channel in the open and closed states. The minimum SFA-channel distances showed a pattern that was identical to that observed for the PUFAs (Figure III.5). The SFA carboxyl head groups were on average 0.4 ± 0.1 Å closer to the channel in the open state, while no significant differences were detected across the different sections of the tail. However, the numbers and contact frequencies of interactions made by the SFA carboxyl head groups and carbon tails to the protein differed significantly between the open and closed states. SFA interactions for both head and tail regions in the closed state were lower both in number and contact frequency compared to the open state interactions (Figure III.6).

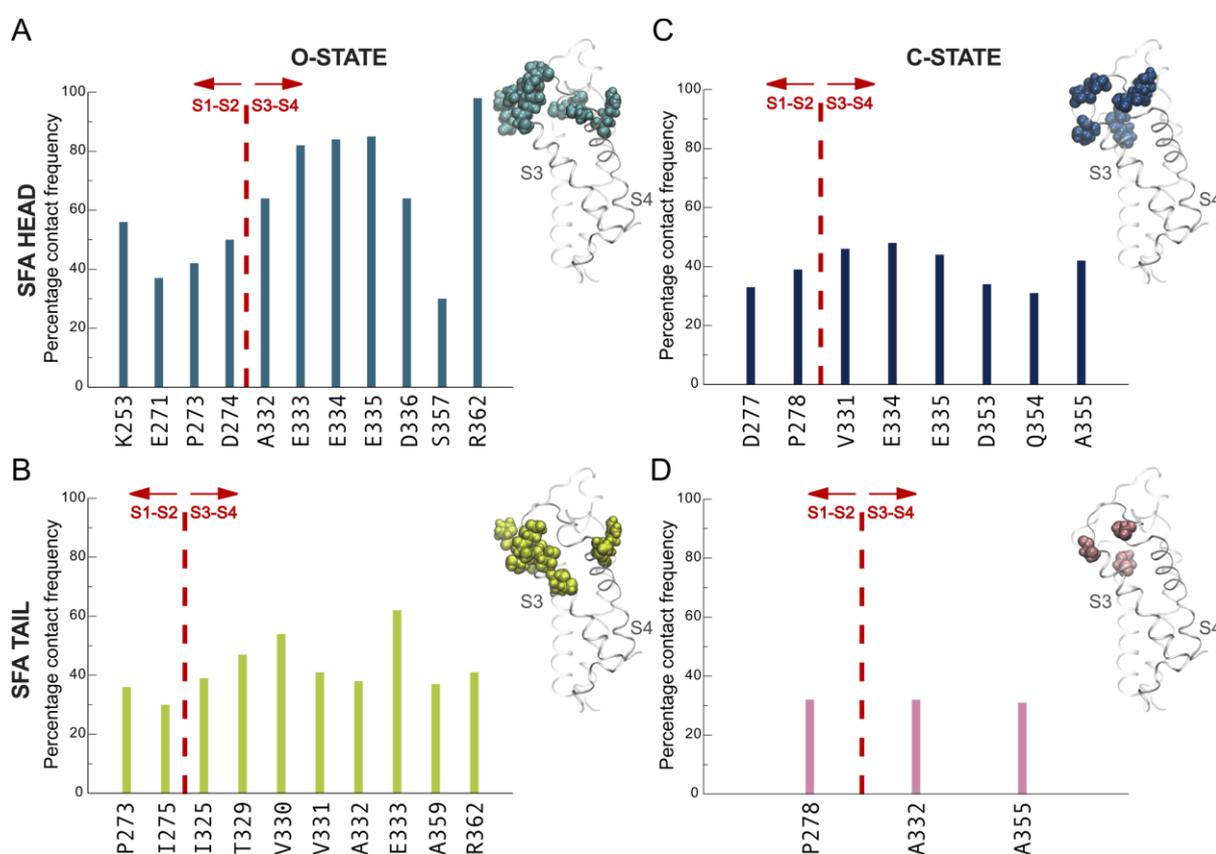


Figure III.6 SFA contacts to the *Shaker* tetramer in the open and closed states. The contact frequencies of amino acid residues within 3.5 Å of SFA carboxyl head groups and carbon tails are displayed for the open (A and B) and closed (C and D) states of the channel. The red dotted line differentiates between helices S1 and S2 or helices S3 and S4 of the VSD. Side-view of a VSD and interacting residues are displayed separately for the PUFA carboxyl head groups and tails for each state of the channel (insets).

Interestingly, the specific protein residues involved in PUFA and SFA interactions differed significantly. While almost all protein residues contacting PUFA head groups in the open state were located on helices S₃-S₄, the contacting residues were shifted to helices S₁-S₂ for the SFAs (Figure III.4 A and III.6 A) and the PUFA tail interactions in the open state were more numerous than for SFAs (Figure III.4 B and III.6 B). In addition, about half of the contacts between the protein and the SFA head and tail groups were completely unique. Finally, the specific contact patterns seem independent of the overall dynamics since there were no significant differences in residency times between PUFAs and SFAs in neither open nor closed states (Figure III.7). However, while the distribution of the POPC residency times indicated more dynamics compared to PUFAs, this difference was reduced in the vicinity of four VSD residues shown experimentally to be involved in PUFA modulation of the *Shaker* channel [129]. In this position, three POPC molecules accompanied one single PUFA and reported 1 μ s residency times.

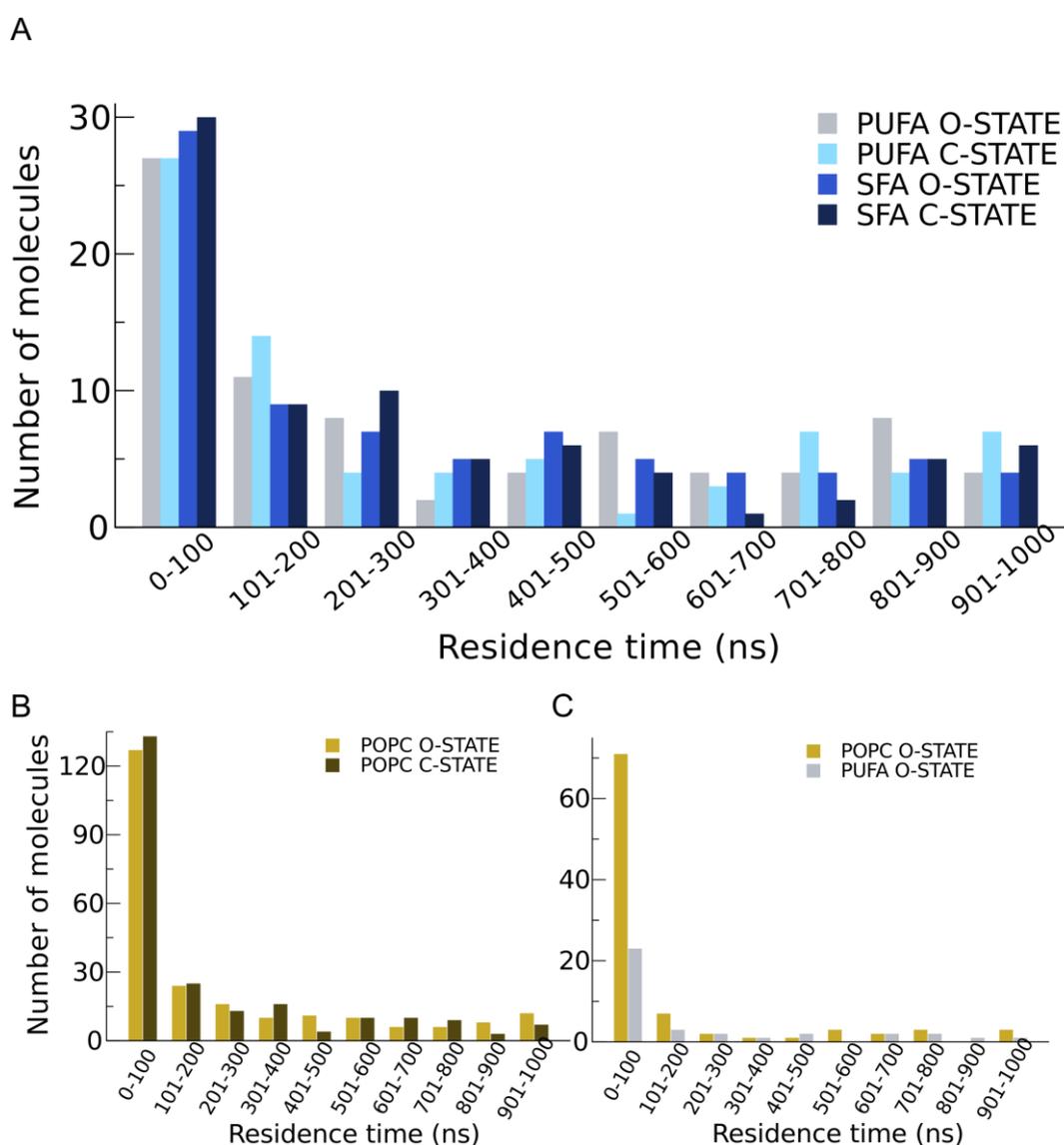


Figure III.7 Number of PUFA/SFA/POPC molecules within 2 Å of the channel. Residence time for the number of PUFA, SFA (A) and POPC molecules (B) in the open and closed

states of the channel. (C) Residence time for the number of PUFA and POPC molecules in the open state of the channel in vicinity of four VSD residues shown experimentally to be involved in mediating the PUFA effect.

H.3.5 Electrostatic interactions characterize the PUFA interaction sites

In the open state PUFA simulation, PUFAs were observed to cluster in the lipophilic pocket of the VSD at the extracellular end of the channel proximal to the S₃-S₄ linker with the hydrophobic tail tucked between helices S₃ and S₄ (Figure III.8 A). To characterize the identified PUFA interaction sites, I turned to reproduction of experimental data. Mutating residues I325 (S₃), T329 (S₃), A359 (S₄), and I360 (S₄) to cysteine and subsequent modification using positively charged MTSEA⁺ affect the PUFA-induced shift of the *Shaker* K_v channel [129]. Therefore, these four residues were proposed to directly modulate function by interactions to PUFA molecules. I introduced positively charged MTSEA⁺ cysteine labels at positions I325, T329, A359, and I360 in the open state and simulated for 1 μ s. The minimum distances between the MTSEA⁺ mutated residues (I325, T329, A359, I360) and the PUFA head groups showed significant interactions varying between 2-4 Å (Figure III.8 B). To further verify this result, I included a negative control also originating from experimental data; mutating the L366 residue did not induce PUFA-mediated shifts in *Shaker* activation [129]. Indeed, the minimum distance between the PUFA head group and L366 remained larger than 10 Å during 500 ns of simulation, which is in stark contrast to the PUFA-sensitive mutations (Figure III.8 B).

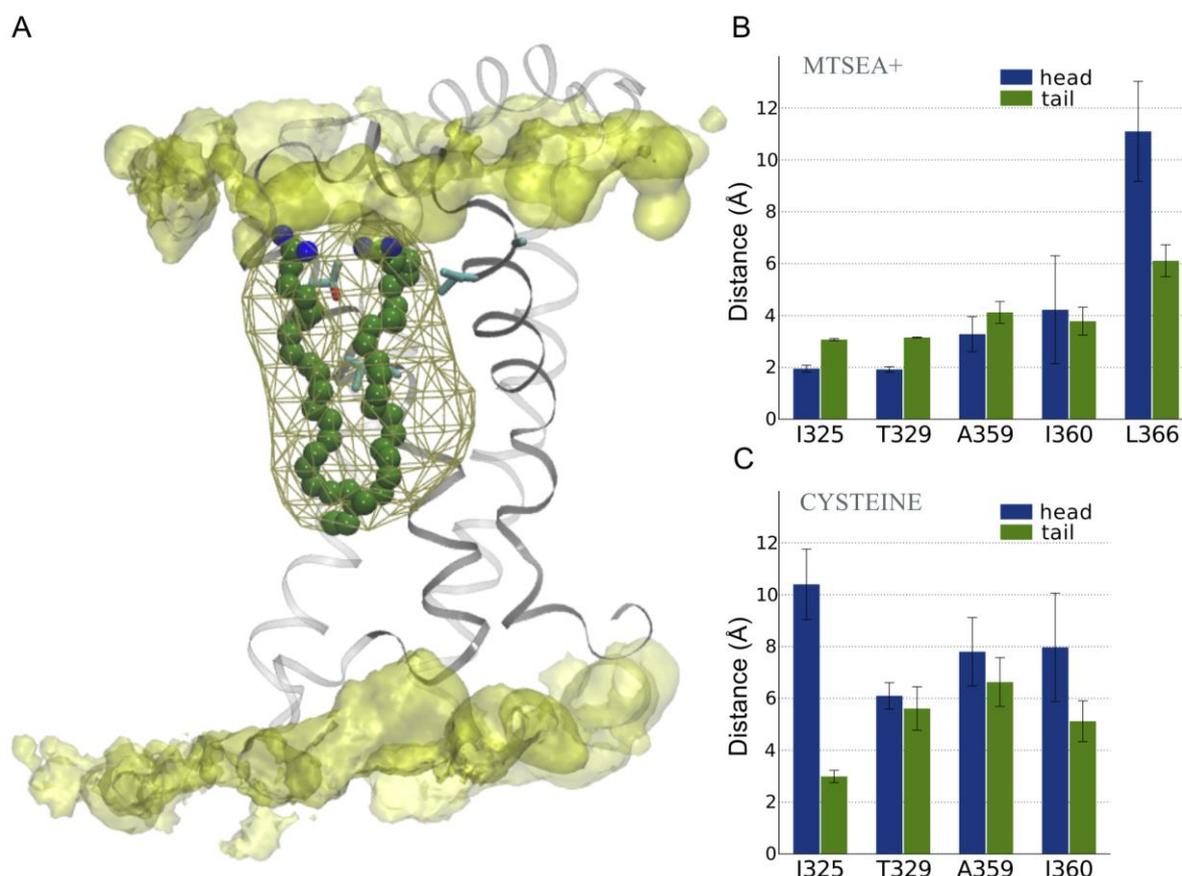


Figure III.8 Characterization of PUFA interaction sites by introduction of MTSEA⁺ and cysteine probes. (A) Side-view of PUFAs interacting with the VSD and a PUFA iso-density surface at 14 % occupancy depicted in brown mesh. (B) Average minimum distances between MTSEA⁺-modified residues and PUFA carboxyl head groups (blue) and carbon tails (green). (C) Average minimum distances between the cysteine-mutated residues and PUFA carboxyl head groups (blue) and PUFA carbon tails (green). The error bars display the standard errors of the VSD-PUFA minimum distances.

In addition, to differentiate between the role of the PUFA head groups and tails and their specific interactions with the channel, I also monitored distances between the carbon tails and the mutated residues. In these analyses, the carbon tails were restricted to carbons C₁₂₋₂₂ in the PUFA carbon chain near the terminal methyl group to enable better discrimination between PUFA head and tail. In contrast to PUFA head groups, the PUFA carbon tails maintained a ~1 Å larger minimum distance for all mutations except the I360 mutation which has a slightly lower average distance although a larger standard error in the head group minimum distance may account for this divergence (Figure III.8 B). A similar tendency was observed for the control mutation L366, where the PUFA tails displayed a significantly lower minimum distance to the mutated residue. In summary, the MTSEA⁺ simulation systems validated the modes of PUFA-protein interactions observed in the WT system where the PUFA head group interacted specifically to the S₃-S₄ linker while the tails maintained less defined protein interactions.

Considering the evident influence on the PUFA-channel interaction caused by MTSEA⁺-modified residues, I introduced an additional test to probe whether the effects

remained after the removal of the added charge. At 500 ns, the MTSEA⁺ modified residues were mutated back into cysteines and were simulated for an additional 500 ns. Upon charge removal, a distinct change in the average distances between the cysteine side chains and the PUFA head groups was observed (Figure III.8 C), with PUFAs on average displaying a minimum distance of 6 - 10 Å across the four mutated systems. Similarly, the distances between the PUFA carbon tails and the cysteines were significantly altered, except for the I325 mutation (Figure III.8 C).

H.4 Discussion

The general effects of PUFAs on both channel activation and inactivation has been documented for K_v channels [103,105], Na_v channels [104,107,237,238], and Ca_v channels [107,239], but exactly where these unsaturated fatty acids bind remains less clear. It is also not ascertained whether it is the ion pore or the VSD that are the targets of action. For instance, reduction of K_{v1.1} channel currents were assigned to PUFA-protein interactions involving hydrophobic residues lining the cavity of the ion pore in the open state [240]. In contrast, a mutational study proposed a PUFA-interaction site to be located on the VSD, specifically the lipid-facing surface of the extracellular side of TM helices S₃ and S₄ [129].

Atomistic MD simulations are designed to monitor structural dynamics of complex environments, such as a membrane protein inserted into a lipid environment and solvated by water and ions. Therefore, the MD simulation approach is highly suitable for identifying and monitoring PUFA-protein interactions on the molecular level, which would be difficult to determine experimentally. Because a PUFA-binding site has never been characterized in atomistic detail, I set out to identify and characterize a potential PUFA binding site. The observed PUFA interaction site was located on the lipid-facing side of TM helices S₃ and S₄ on the extracellular side of the *Shaker* K_v channel in its open state, in agreement with previous experimental findings [129]. In addition, I observed a significant difference in the structural dynamics and protein interaction patterns of the PUFA carboxylic head group and the lipophilic carbon tail. The PUFA head groups favored fewer and highly specific polar/charged residue interactions along the S₃-S₄ extracellular part, in contrast to the carbon tails that interacted non-specifically with a wider range of S₃-S₄ residues buried in the hydrophobic core in the lipid bilayer. Furthermore, by monitoring the structural dynamics displayed by SFAs and PUFAs, I ascertained an increased conformational flexibility in the polyunsaturated carbon tails compared to saturated carbon chains. Because the saturation level of fatty acids are highly correlated to modulation of channel function [128,238,241,242], the structural dynamics of fatty acids in the lipid bilayer are likely important to obtain optimal interactions to the channel protein. The results indicate that the flexibility of the unsaturated carbon tail allows the fatty acid to visit several binding modes that enable a strong specific interaction between the carboxyl head group and the channel, which would otherwise be inaccessible to a more rigid saturated carbon tail. A similar conclusion was drawn from recent experimental studies of the *Shaker* channel [128]. To test this predicted interaction I performed simulations in the presence of SFA molecules. Indeed, the saturated carbon tails of the SFAs displayed significantly less interactions to the protein compared to the PUFA tails in the open state of the channel. In the closed state, both PUFAs and SFAs formed different interaction patterns compared to the open state

simulations. A reduced number of interactions, in particular in the carbon tail interactions, contrasted the open and closed state interactions.

The increasing flexibility with unsaturation level of fatty acids is well documented. For instance, NMR studies have shown DHA to undergo rapid conformational transitions with short correlation times and exceptionally low deuterium order parameters [243,244]. In addition, quantum mechanical calculations have shown that polyunsaturated chains sample greater conformational space around the *cis* double bonds with more rapid reorientations near the methyl end of the chain [233]. The increased conformational flexibility in PUFAs has been explained by lowered torsional energy barriers for the rotatable bonds in these carbon chains [245]. In addition, several rhodopsin studies show PUFA-specific modulation [245,246]. Therefore, tail flexibility enabling PUFA electrostatic head group interactions may be a general mechanism.

The interactions between the PUFA head groups and the VSD in the open state were concentrated to a few residues displaying high contact frequencies. While all interactions did not occur across all four subunits, sampling interactions across subunits for extended microsecond simulations times enabled identification of the major interaction residues. Interestingly, one of these residues was R362, which is the first (R1) of the gating charge residues. Because involvement of gating charges has also been observed in *Shaker* K_v experiments [129,247], it is indeed possible that PUFAs affect channel function by binding to our proposed binding site and reach the R1 gating charge from this position. In a recent study, mutations of residues M356 and A359 into arginines increased the K^+ channel's sensitivity to PUFAs considerably [247]. These residues, which are positioned on the S4 helix, also show up in our contact analyses but with lower contact frequencies, which might reflect limited sampling. In a study by Xu *et al.*, it was established that higher hydrophobicity in a ten-residue segment in the extracellular part of the S3 helix in the paddle chimera helps to stabilize the open state of the channel [248]. Out of these ten residues, residues I325, T326, T329, V331, and A332 were identified as close contacts to PUFAs in our simulations of the open state channel, indicative of the role PUFAs play in stabilizing the open state of the channel.

Further comparison of the observed PUFA-protein interaction pattern to that of other lipid modulators such as the phosphatidylinositol-4,5-bisphosphate (PIP₂) lipids hints at significant complexity. Recently, PIP₂ lipids were found to stabilize the closed state of $K_{v7.1}$ channels by interacting with the lower residues of S4. However, in an open state PIP₂ lipids migrated to the pore domain (PD) to form salt bridges with the S6 terminus and in this way mediated coupling between the VSD and the PD [249]. PIP₂ lipids were also observed to migrate from the S4-S5 linker to the S2-S3 linker in KCNQ₂ K_v channels to control deactivation kinetics [250]. Thus, MD simulations have identified putative structural features underlying the function of modulatory agents that appear fundamentally different. Future experimental efforts are needed to verify the interaction patterns presented in our study. Finally, the simulations studies provide a structural framework for future studies aimed at determining the free energy associated with K_v channel activation and deactivation in the presence of modulatory agents.

Chapter I

Study IV

1.1 The Regulatory Role of Polyunsaturated Fatty Acids on the Free-Energy Landscape of K_V Shaker Channel Deactivation

High-resolution X-ray structures of K_V channels, in particular that of the chimera channel $K_{V2.1}/K_{V1.2}$ [82], have been instrumental in the progress of gaining an improved understanding of the structure and function of voltage-gated ion channels. In moving towards a more refined picture of the mechanisms of voltage sensing and channel activation recent structural models together with MD simulations [221,235,251,252], as well functional characterizations [253,254] have been major contributors. There are three different functional and conformational states of the voltage sensor that have been proposed by experimental studies [255]; a polarized membrane causing the resting closed state (C), an active open state (O) due to depolarization of the membrane, and an open-inactivated channel that results from prolonged depolarization. Free energy barriers are what separate these three states [256], and through applying an external potential the relative free energy of each state changes, causing the channel to transition into a different state [257]. One should bear in mind that in effect all experimentally solved structures of the voltage sensor are of the open state conformation of the channel and all models of every other conformational state are theoretical models that require experimental validation.

Channel activation involves the independent outward movement of the four S_4 helices, of which three models have been proposed that explain the exact movement of this

helix and its relation to the other helices in the VSD. The *sliding helix* model [94,95] suggests that the positively charged (basic) residues on S₄ make contact with counter negatively charged (acidic) residues on neighboring parts of the VSD by a substantial translational and rotational movement. The *transporter-like* model posits a large rotational but lacking in a significant translational movement of S₄ in transferring charges from the intracellular to the extracellular matrix [96,97]. The third model that followed the publication of the crystal structure of KvAP [97] is known as the *paddle* model, which proposes the conjoint movement of S₄ and S₃, known as the voltage sensor paddle, during gating which reach a transmembrane position in the activated state.

With S₄ moving down in sequential steps in the gating process, the existence of metastable intermediate states has been suggested by extended MD simulations that are also supported by experimental data [220,221]. In each state, one charged residue on S₄ passes a barrier, with four separate energy barriers to overcome. The activation and deactivation barriers are possibly generated by a hydrophobic zone in the VSD that each basic residue on S₄ crosses during the gating process [258]. The well-conserved aromatic residue, F290, located in the middle of the S₂ helix, was first suggested by Long *et al.* to play the role of the hydrophobic plug (Figure IV.1). Shown to be important for gating, F290 has been observed to interact with all four gating charges on S₄ [259,260].

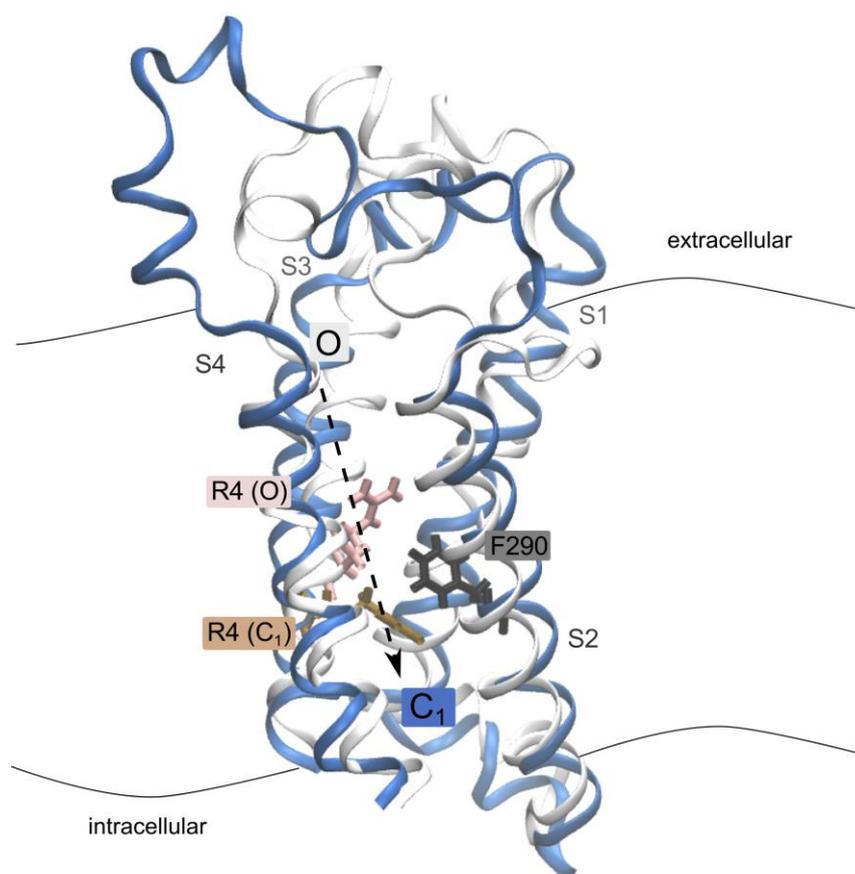


Figure IV.1 Voltage-sensor domain comparison. The voltage-sensor domain of the *Shaker* K_V channel in its open (white) state is superimposed on its putative intermediary C₁ state

(blue), depicting the translation of R₄ along S₄ past the hydrophobic plug F₂₉₀ located on helix S₂.

The idea of secondary structure changes in S₄ during different stages of gating has been critical in explaining stabilization during gating. Supported by crystal structures and experiments including histidine scans and MD simulations, a 10-residue stretch on S₄, downstream of R₃, adopts a 3_{10} -helix conformation in the resting and active states, but transitions into an α -helix when entering into the open-inactivated state [82,255,261-266]. By incorporating a 3_{10} -helix segment, with three residues per turn, S₄ becomes more elongated and tightly wound, and as a consequence the arginine side chains align on the same side on the helix and stabilize the VSD by their interactions with the basic residues on S₂ and S₃ during gating and inhibit large distortions in the VSD [94,95,255,264,267].

Previous gating measurement and mutational studies have shown that PUFAs mainly affect the coordinated opening step of a K_V channel and have less impact on the earlier stages of transition [128,129]. By affecting the final S₄ transition, which is closely linked to the opening of the channel, the larger PUFA effect can be explained by the outward movement of the gating charges moving them closer to the putative PUFA binding site on the VSD [129].

In this study, I compute the energetics of the free-energy landscape of the *Shaker* K_V channel in lipid bilayers free from as well as enriched with PUFAs, using an enhanced sampling approach, namely accelerated-weight histogram (AWH). By choosing a reaction coordinate along the vertical translation of S₄ towards its down state in the deactivation pathway, we investigate the free energy differences in passing the first energy barrier of deactivation, i.e. going from the open (O) state to the closed-1 (C₁) state in a K_V channel that remains either affected or unaffected by PUFAs.

1.2 Methods

1.2.1 *Shaker* channel in PUFA and PUFA-free systems

The simulation model of the tetrameric K_V *Shaker* channel was adopted from the previously modeled (see Study III) open channel configuration of the channel based on the high-resolution X-ray structure of the K_V2.1 paddle-K_V1.2 chimera channel (PDB ID 2R9R) [82]. The *Shaker* system in this study was taken from the fully equilibrated 1 μ s simulation of wild-type *Shaker* inserted into a POPC membrane. Likewise, the PUFA-*Shaker* setup was taken from the MTSEA⁺ modified *Shaker* channel at residue I325 embedded in a PUFA enriched membrane bilayer simulated for 1 μ s. Experimental data have shown how the voltage-dependence of the *Shaker* channel shifts by the introduction of a positively charged MTSEA⁺ probe at this site [129], results which were corroborated in the previous study (Study III). For this reason, a mutated *Shaker* channel was selected for the PUFA-*Shaker* system as this mutation has been observed to directly modulate function by interactions to PUFA molecules. The simulation systems were set up as described previously in Study III and carried out by Gromacs [170,171] using the CHARMM36 force field [231] with 5 fs time steps.

1.2.2 Accelerated weight histogram calculations

In a study by Henrion *et al.* [235], several intermediate closed states (C_1 , C_2 , C_3) between the open (O) and resting states of the *Shaker* channel were modeled based on a set of experimental constraints, mainly the formation of metal-ion bridges between specific residues alongside helices S_3 and S_4 in the VSD allowed the generation of Rosetta models which were verified by molecular simulations. The structural information from the generated intermediate closed-1 state (C_1) permitted investigation of free energy differences between activation and the first step of deactivation (O to C_1) in the *Shaker* K_V channel in the absence and presence of PUFA in the membrane bilayer. The accelerated weight histogram (AWH) method was used to sample the O- C_1 transition by applying a harmonic potential or umbrellas to the center-of-mass of a pull group on helix S_4 and the center-of-mass of a reference group made up of helices S_1 - S_3 [145]. The pull group incorporated the backbone residues 365-375 on S_4 and the charged side chains of R2-R3-R4-K5, which were given a relative weight of 2.7. AWH is an extended ensemble sampling technique, which adaptively biases the simulation to promote exploration of the free energy landscape by using a probability weight histogram that allows for efficient free energy updates. With a pull direction along an updated vector defined by the backbone residues of helices S_1 and S_3 , the AWH simulations started from the open state (O-state) pulling down towards the C_1 state (C_1 -state). As one of perks of the AWH method is repeated passage over the reaction coordinate, transitions from the O to C_1 state allowed for exploration of multiple pathways. Diffusion constant along reaction coordinate was set to $1e-6$ (nm^2/ps), and a maximum and a minimum reaction coordinate were fixed with an interval of 8 Å. The AWH simulation for the *Shaker* and the PUFA-*Shaker* systems was run for 5 μs each.

1.3 Results

1.3.1 Sampling approach

With a number of different ways of computing free energy, the main complication lies in extracting accurate free energies from molecular simulations. The thorough exploration of free energy landscapes requires sampling large conformational changes and slow transitions. I, therefore, have chosen the accelerated weight histogram (AWH) method to calculate free energy along a reaction coordinate [145]. A technique that by using a probability weight histogram enables efficient free energy updates, AWH enhances sampling by adaptively biasing the simulation to promote exploration of the free energy landscape. A proper choice of reaction coordinate increases the scope of exploration of the configuration space and consequently resulting in a more accurate estimate of the free-energy profile along the chosen reaction coordinate. In the event of a poorly chosen reaction coordinate, detection is facilitated by the AWH method, as the weight histogram would in such cases deviate significantly from the target distribution.

I have calculated the energetics of the O to C_1 deactivation landscape of the *Shaker* K_V channel in its atomistic description inserted in both PUFA-free and PUFA-enriched lipid

bilayers, using the AWH method. In selecting a suitable reaction coordinate to enable an enhanced sampling of the deactivation pathway, a natural choice is one associated with the deactivation pathway itself, namely, the vertical displacement of helix S_4 . In this framework, the reaction coordinate is determined by the spatial translation of the gating charge R_4 located on S_4 in transitioning from the O state to the C_1 state of the channel, progressing down the deactivation pathway along a vector defined by helices S_1 and S_3 of the VSD (Figure IV.1). The free energies associated with the presence and absence of PUFAs on the *Shaker* K_V channel were investigated in 5 μ s long AWH samplings. However, as transitions along the deactivation pathway slowed down significantly and entire O - C_1 transitions barely took place beyond 1 μ s into the simulations, I decided to focus on this initial part of the sampling data.

1.3.2 PUFA affects the free-energy landscape of deactivation

Sampling over the course of 1 μ s allowed the calculation of free energies of deactivation for the *Shaker* K_V channel in membrane bilayers both free of and enriched with PUFAs. The resulting free-energy surface reveals striking differences in favorable states along the deactivation pathway (Figure IV.2). The average free energy, calculated across the four subunits of the channel, reveals a lower free energy in the open state of the K_V channel in the PUFA enriched system with a difference of ~ 8 kcal/mol. In contrast, at the intermediary C_1 state of the channel, with an energy level ~ 10 kcal/mol lower, the K_V channel in the PUFA free membrane is at a more stable state. Noteworthy, is the comparison of the energy landscape before and after the transitioning across the hydrophobic barrier. Prior to the transition across this barrier, the PUFA enriched system maintains a much lower free energy, however, once the barrier has been crossed and the transition to the metastable C_1 state is approaching the PUFA enriched system enters into a higher free-energy state.

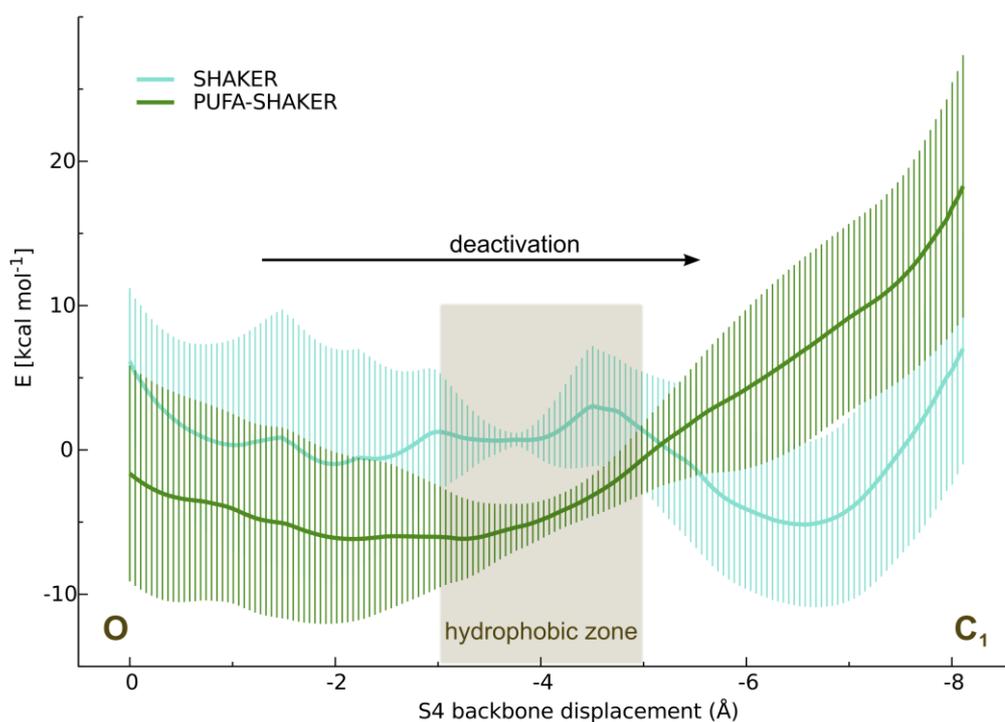


Figure IV.2 Free-energy profile of the O to C₁ transition. Free-energy profiles along S₄ backbone displacement for the *Shaker* K_V channel in PUFA free (blue) and PUFA enriched (green) lipid bilayers. The error bars display the standard error of free energy across the four channel subunits. The shaded area depicts the energy barrier characterized by a hydrophobic zone along S₂ and S₄.

The large standard errors and the shape of the free energy profile lacking a barrier distinguishing the activated state from the intermediate C₁ state by the absence of two stable minima suggests that a much more thorough sampling of the deactivation pathway may be required. This could increase the possibility that a finer description of the configurational space can potentially give rise to a smoother free-energy landscape with a visible barrier separating the two thermodynamically stable states. Deterred by overcoming the hydrophobic barrier within the scope of current sampling, I deduce that the obtained free-energy profile has not converged and requires further work.

1.3.3 PUFA stabilizes the open state configuration

The distortion of the channel during the transitions can be assessed by calculating the root mean-square deviation (RMSD) of the entire VSD and segments S₁-S₃. The tendency of the PUFA enriched system to stabilize the open state of the *Shaker* K_V channel as observed by the free energy landscape, is also reflected in RMSD measurements of the VSD of the channel as illustrated in Figure IV.3. This is particularly evident in the RMSD of the initial O to C₁ transition. With the S₁-S₃ domain of the K_V channel in the PUFA enriched membrane bilayer deviating considerably less from the open state crystal structure, an RMSD between 1.5 – 2.0 Å is observed. In contrast, the S₁-S₃ domain of the PUFA free system displays higher perturbation and initially has an RMSD of roughly 2.0 – 2.5 Å. However, following

the initial O to C₁ transition in which S₄ has translated down once, both the PUFA free and PUFA enriched systems converge to comparable RMSD values, fluctuating around an RMSD of ~ 2 Å. A similar pattern is observed in the RMSD for the entire VSD of the channel in both systems in which the same behavior is observed except for slightly higher RMSD values, which is due to the inclusion of S₄ that contributes further by the pull-down motion along the deactivation pathway (Figure IV.3, inset).

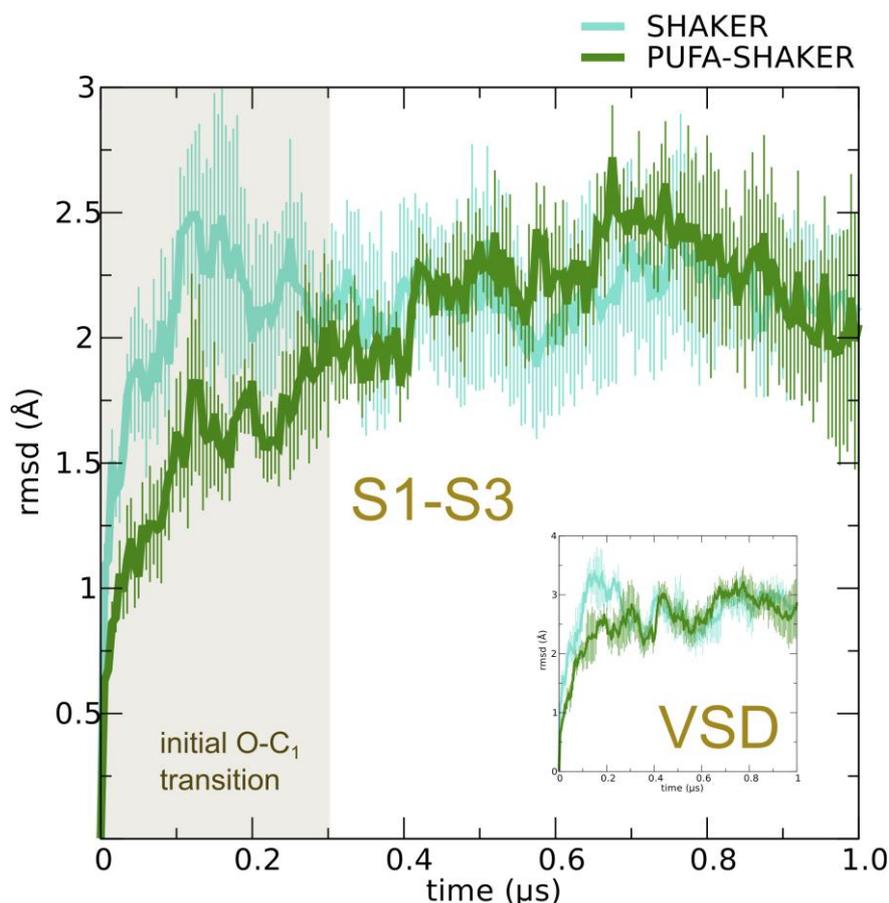


Figure IV.3 Root-mean square deviation (RMSD) of the backbone atoms of segments S1-S3 during 1 μ s of AWH sampling. S₄ is translated in a downward motion driving the channel from an open state to an intermediate C₁ state in PUFA free (blue) versus PUFA enriched (green) systems. RMSD was calculated relative to the initial conformation after a least squares fitting of S1-S3, with error bars indicating the standard errors across the four subunits of the channel. (Inset) RMSD for the entire VSD compared in both systems.

1.3.4 3₁₀-helix integrity

To further investigate the structural dynamics of S₄ during translation into a down state, I performed secondary structure evolution calculations to obtain a clearer picture of structural element formations of the S₄ helix when the channel transitions into the C₁ state. As illustrated in Figure IV.4 A, in both systems free of and enriched with PUFAs, the S₄ residues between R1 and K5, largely adopt a 3₁₀-helix formation during the initial transition. It is beyond this first transition that a differing trend separates the two systems. Beyond the

initial transition frame, there is a spontaneous change of secondary structure in S₄, with the 3_{10} -helix region losing its stability and converting to an α -helix in the PUFA enriched system.

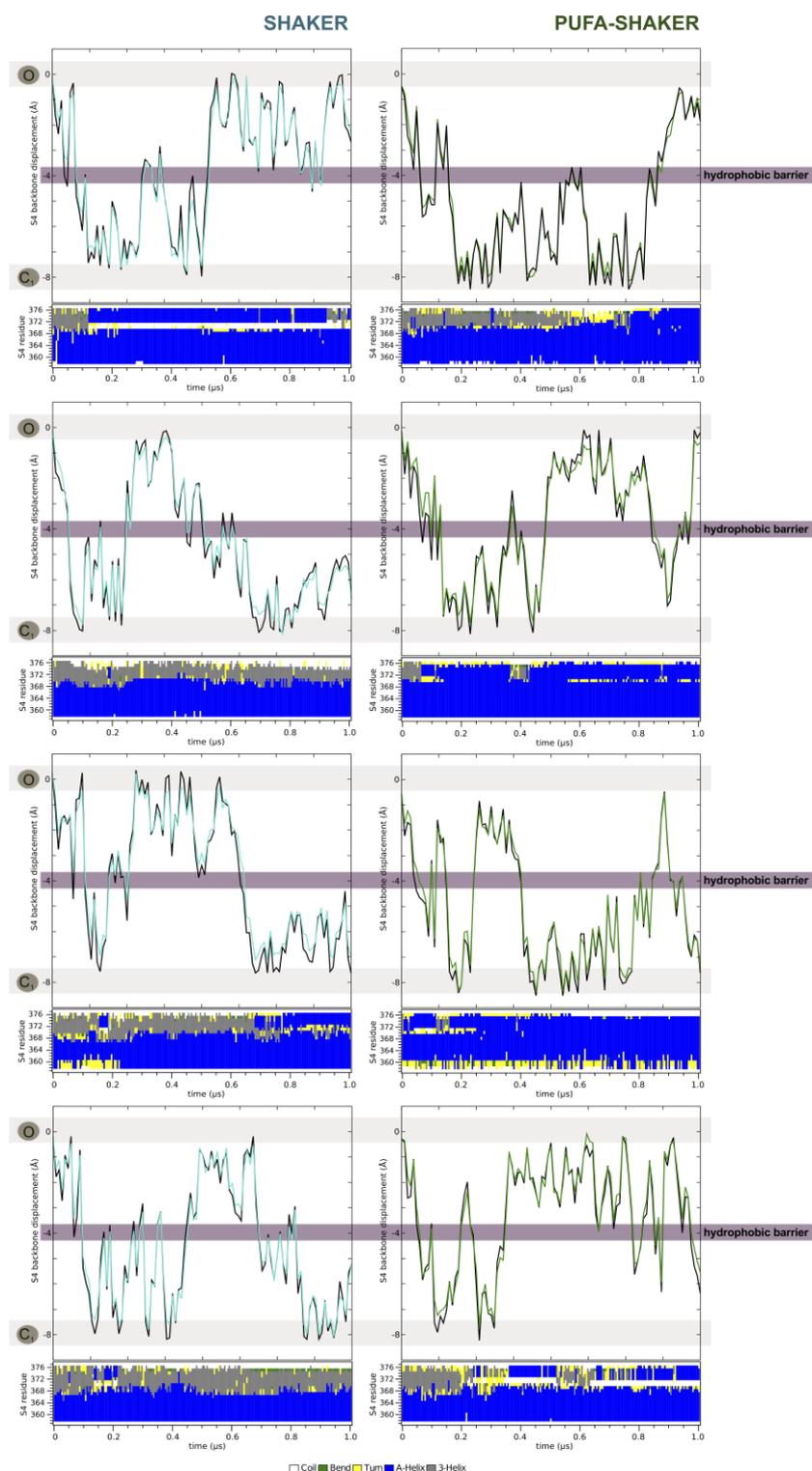


Figure IV.4 Secondary structure evolution of S₄ segment. 3_{10} -helix formation in the R1-K5 region evolves as S₄ translates from the O state down towards the C₁ state along the

reaction coordinate in the PUFA free (left panel) and PUFA enriched (right panel) systems, displayed individually for each subunit of the tetrameric channel.

In general, the 3_{10} -helix propensity for the R1-K5 segment remains higher for the PUFA free system as opposed to the PUFA enriched system (Figure IV.5). Although this can be interpreted as support for stabilization of a 3_{10} conformation as S_4 translates down in a PUFA free system, it is worth noting that this could be due to totally arbitrary events as the derived free-energy landscape for both PUFA free and enriched systems lacked proper energy barriers and could not be totally relied upon.

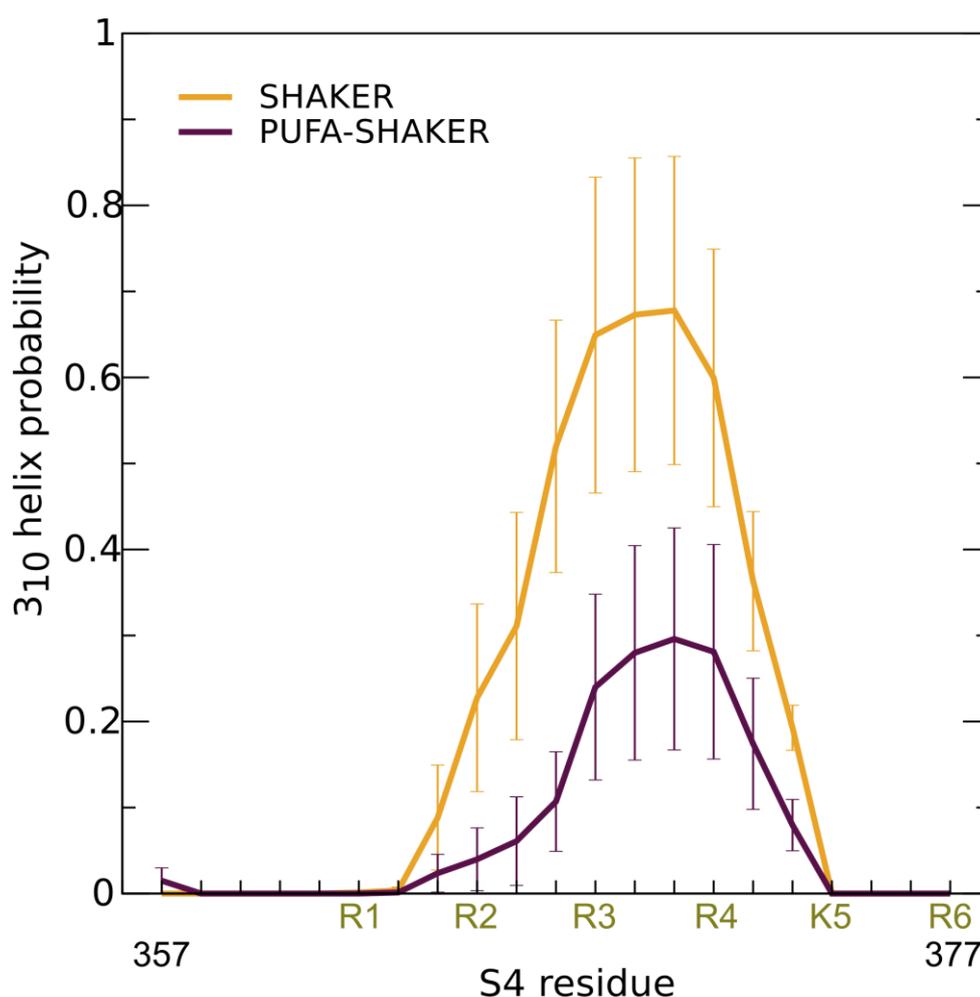


Figure IV.5 3_{10} -helix propensity. Probability distribution of 3_{10} -helix formation for a 21-residue segment on S_4 in systems free of (orange) and enriched with (maroon) PUFAs.

1.4 Discussion

In recent years, literature investigating lipid-protein interactions have increasingly been highlighting the fundamental importance of lipids on voltage sensor function. Both experimental and simulation studies (see Study III) have generated substantial evidence

towards PUFAs modulating K_V channel function by interacting with a specific binding site situated at the protein-lipid interface of the VSD in proximity of the extracellular halves of the S_3 and S_4 helices, thereby stabilizing the open active state of the channel [129]. In particular, gating measurement studies have identified the last step of transition towards an open state in the gating pathway to be dominated by the PUFA effect [128,129]. Naturally, a relevant and interesting question, which follows, is the estimation of the free energy difference that arises from the effects of PUFAs in transitioning one step down the deactivation pathway, from the open state to the C_1 state.

These results, despite its' limitations, has shed light on apparent free energy differences during the $O - C_1$ transition. The PUFAs' close interactions with the K_V channel results in a lower free energy at the open state stabilizing this conformation of the channel. As the PUFAs bring the channel into a thermodynamically more favorable state, their absence on the other hand, results in a relatively higher free energy enabling the transition to the C_1 state more readily. Once past the hydrophobic barrier, the absence of the effect of PUFAs on the channel leads to lower free energy compared to an energetically more unstable C_1 conformational state of the channel that is interacting with PUFAs. Although it was not possible to pull slowly enough to obtain repeated $O - C_1$ transitions to derive statistically accurate work profiles, results from our initial simulations indicate that further sampling of the deactivation pathway is required. As previously mentioned, the large standard errors in addition to the non-existence of clear barriers separating the two energetically stable states, the open versus the C_1 intermediate, are indicative of sampling that has not converged and demands additional exploration of the free energy landscape along the deactivation pathway. In fact, in order to ensure continued and consistent PUFA-channel interaction the introduction of a single mutation to the channel may very well have had its effects on the statistical sampling and contributed to inaccuracies on the free energy landscape of the VSD and altered the intrinsic equilibrium between the states.

The conformational change of an α -helix into a 3_{10} -helix of the S_4 segment that carries the gating charges has been linked as a component of the inactivation barrier [255,256,264,266]. The 3_{10} -helix adoption has been suggested to lower the energetic cost of the intermediate barriers, thus stabilizing the α -helix after crossing the hydrophobic barrier requires higher energies than that of the 3_{10} -helix. This explains the higher free energy associated with the PUFA enriched system after passing the hydrophobic barrier in this study, in which S_4 fails to keep its 3_{10} helical conformation beyond the initial transition to the C_1 state.

Concluding Remarks and Future Perspectives

Every question answered in biology corresponds to the correct placement of a piece of a jigsaw puzzle. With every piece, as one part of the complete picture is revealed, hints are given as to the remaining questions. A theoretical approach in solving the jigsaw puzzle has been tremendously instrumental. However, the application of computational methods in exploring the dynamics of biological systems alongside its' strengths carries limitations as well. The strengths and limitations have naturally been reflected in the different studies carried out as part of this thesis.

The strengths of computational methods particularly manifest themselves when wet-lab experiments are unable to capture certain states of structures or the relevant dynamics associated with protein function. The structural characterization of membrane proteins are for instance challenging, as the requirement of a membrane environment for purification and crystallization is tricky to fulfill or solving the structure of a highly dynamic and potentially disordered N-terminal extension of water-soluble proteins as witnessed in this thesis. Capturing a dynamical process such as the gating mechanism of ion channels or the structural effects of post-translational modifications of proteins is not easily achieved by experimental procedures. Modeling techniques as well as simulations have enabled the study of these processes. We have obtained a clearer picture of the structural elements of I κ B α and the effects of phosphorylation on the dynamic properties of this transcription factor inhibitor. Simulations have shed light on the dynamic nature of interactions of PUFAs with the *Shaker* K_V channel and aided in understanding the underlying mechanisms.

The limitations of computational methods are equally many. Limited computational power leading to poor covering of biologically relevant time scales and thorough exploration of the entire configurational space with non-optimal sampling methods is one of the major drawbacks. Improved force fields with parameter validations are also necessary for accurate simulations. Inefficient and insufficient sampling was especially evident during the free energy calculations of the first step of deactivation of the *Shaker* K_V channel, and additional sampling are absolutely required to obtain a more accurate free energy landscape.

Bibliography

1. Lehninger AL, Nelson DL, Cox MM (2008) *Lehninger principles of biochemistry*. New York: W.H. Freeman.
2. What Are Proteins? Central Dogma of Molecular Biology.
3. Kendrew JC, Bodo G, Dintzis HM, Parrish RG, Wyckoff H, et al. (1958) A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature* 181: 662-666.
4. Perutz MF, Rossmann MG, Cullis AF, Muirhead H, Will G, et al. (1960) Structure of haemoglobin: a three-dimensional Fourier synthesis at 5.5-Å resolution, obtained by X-ray analysis. *Nature* 185: 416-422.
5. Ramachandran GN, Ramakrishnan C, Sasisekharan V (1963) Stereochemistry of polypeptide chain configurations. *J Mol Biol* 7: 95-99.
6. Ramachandran GN, Sasisekharan V (1968) Conformation of polypeptides and proteins. *Adv Protein Chem* 23: 283-438.
7. Pauling L, Corey RB (1950) Two Hydrogen-Bonded Spiral Configurations of the Polypeptide Chain. *Journal of the American Chemical Society* 72: 5349-5349.
8. Pauling L, Corey RB (1951) The pleated sheet, a new layer configuration of polypeptide chains. *Proc Natl Acad Sci U S A* 37: 251-256.
9. Pauling L, Corey RB, Branson HR (1951) The structure of proteins: Two hydrogen-bonded helical configurations of the polypeptide chain. *Proceedings of the National Academy of Sciences* 37: 205-211.
10. Protein Structure I.
11. Yang AS, Honig B (1995) Free energy determinants of secondary structure formation: I. alpha-Helices. *J Mol Biol* 252: 351-365.
12. Ermolenko DN, Richardson JM, Makhatadze GI (2003) Noncharged amino acid residues at the solvent-exposed positions in the middle and at the C terminus of the alpha-helix have the same helical propensity. *Protein Science* 12: 1169-1176.
13. Vieira-Pires RS, Morais-Cabral JH (2010) 3(10) helices in channels and other membrane proteins. *Journal of General Physiology* 136: 585-592.
14. Farzadfard F, Gharaei N, Pezeshk H, Marashi SA (2008) Beta-sheet capping: signals that initiate and terminate beta-sheet formation. *J Struct Biol* 161: 101-110.
15. Rose GD (1978) Prediction of chain turns in globular proteins on a hydrophobic basis. *Nature* 272: 586-590.

16. Fersht AR, Matouschek A, Serrano L (1992) The folding of an enzyme. I. Theory of protein engineering analysis of stability and pathway of protein folding. *J Mol Biol* 224: 771-782.
17. Chothia C, Janin J (1975) Principles of protein-protein recognition. *Nature* 256: 705-708.
18. Groves MR, Barford D (1999) Topological characteristics of helical repeat proteins. *Curr Opin Struct Biol* 9: 383-389.
19. Kobe B, Kajava AV (2000) When protein folding is simplified to protein coiling: the continuum of solenoid protein structures. *Trends Biochem Sci* 25: 509-515.
20. Main ER, Jackson SE, Regan L (2003) The folding and design of repeat proteins: reaching a consensus. *Curr Opin Struct Biol* 13: 482-489.
21. Breeden L, Nasmyth K (1987) Similarity between cell-cycle genes of budding yeast and fission yeast and the Notch gene of *Drosophila*. *Nature* 329: 651-654.
22. Lux SE, John KM, Bennett V (1990) Analysis of cDNA for human erythrocyte ankyrin indicates a repeated structure with homology to tissue-differentiation and cell-cycle control proteins. *Nature* 344: 36-42.
23. Marcotte EM, Pellegrini M, Yeates TO, Eisenberg D (1999) A census of protein repeats. *Journal of Molecular Biology* 293: 151-160.
24. Mosavi LK, Cammett TJ, Desrosiers DC, Peng ZY (2004) The ankyrin repeat as molecular architecture for protein recognition. *Protein Science* 13: 1435-1448.
25. Sedgwick SG, Smerdon SJ (1999) The ankyrin repeat: a diversity of interactions on a common structural framework. *Trends Biochem Sci* 24: 311-316.
26. Sen R, Baltimore D (1986) Inducibility of kappa immunoglobulin enhancer-binding protein NF-kappa B by a posttranslational mechanism. *Cell* 47: 921-928.
27. Sen R, Baltimore D (1986) Multiple nuclear factors interact with the immunoglobulin enhancer sequences. *Cell* 46: 705-716.
28. Baldwin AS, Jr. (1996) The NF-kappa B and I kappa B proteins: new discoveries and insights. *Annual Review of Immunology* 14: 649-683.
29. Bonizzi G, Karin M (2004) The two NF-kappaB activation pathways and their role in innate and adaptive immunity. *Trends Immunol* 25: 280-288.
30. Ghosh S, May MJ, Kopp EB (1998) NF-kappa B and Rel proteins: evolutionarily conserved mediators of immune responses. *Annual Review of Immunology* 16: 225-260.
31. Miyamoto S, Verma IM (1995) Rel/NF-kappa B/I kappa B story. *Adv Cancer Res* 66: 255-292.
32. Attar RM, Caamano J, Carrasco D, Iotsova V, Ishikawa H, et al. (1997) Genetic approaches to study Rel/NF-kappa B/I kappa B function in mice. *Semin Cancer Biol* 8: 93-101.

33. Whiteside ST, Israel A (1997) I kappa B proteins: structure, function and regulation. *Semin Cancer Biol* 8: 75-82.
34. Vallabhapurapu S, Karin M (2009) Regulation and function of NF-kappaB transcription factors in the immune system. *Annual Review of Immunology* 27: 693-733.
35. Karin M, Ben-Neriah Y (2000) Phosphorylation meets ubiquitination: The control of NF-kappa B activity. *Annual Review of Immunology* 18: 621-+.
36. Arenzana-Seisdedos F, Turpin P, Rodriguez M, Thomas D, Hay RT, et al. (1997) Nuclear localization of I kappa B alpha promotes active transport of NF-kappa B from the nucleus to the cytoplasm. *J Cell Sci* 110 (Pt 3): 369-378.
37. Hoffmann A, Levchenko A, Scott ML, Baltimore D (2002) The IkappaB-NF-kappaB signaling module: temporal control and selective gene activation. *Science* 298: 1241-1245.
38. Chen L, Fischle W, Verdin E, Greene WC (2001) Duration of nuclear NF-kappaB action regulated by reversible acetylation. *Science* 293: 1653-1657.
39. DiDonato JA, Hayakawa M, Rothwarf DM, Zandi E, Karin M (1997) A cytokine-responsive IkappaB kinase that activates the transcription factor NF-kappaB. *Nature* 388: 548-554.
40. Mercurio F, Zhu H, Murray BW, Shevchenko A, Bennett BL, et al. (1997) IKK-1 and IKK-2: cytokine-activated IkappaB kinases essential for NF-kappaB activation. *Science* 278: 860-866.
41. Woronicz JD, Gao X, Cao Z, Rothe M, Goeddel DV (1997) IkappaB kinase-beta: NF-kappaB activation and complex formation with IkappaB kinase-alpha and NIK. *Science* 278: 866-869.
42. Yamaoka S, Courtois G, Bessia C, Whiteside ST, Weil R, et al. (1998) Complementation cloning of NEMO, a component of the IkappaB kinase complex essential for NF-kappaB activation. *Cell* 93: 1231-1240.
43. Mercurio F, Murray BW, Shevchenko A, Bennett BL, Young DB, et al. (1999) IkappaB kinase (IKK)-associated protein 1, a common component of the heterogeneous IKK complex. *Mol Cell Biol* 19: 1526-1538.
44. Huynh QK, Boddupalli H, Rouw SA, Koboldt CM, Hall T, et al. (2000) Characterization of the recombinant IKK1/IKK2 heterodimer. Mechanisms regulating kinase activity. *Journal of Biological Chemistry* 275: 25883-25891.
45. Beg AA, Sha WC, Bronson RT, Ghosh S, Baltimore D (1995) Embryonic lethality and liver degeneration in mice lacking the RelA component of NF-kappa B. *Nature* 376: 167-170.
46. Delhase M, Hayakawa M, Chen Y, Karin M (1999) Positive and negative regulation of IkappaB kinase activity through IKKbeta subunit phosphorylation. *Science* 284: 309-313.

47. Brown K, Gerstberger S, Carlson L, Franzoso G, Siebenlist U (1995) Control of I kappa B-alpha proteolysis by site-specific, signal-induced phosphorylation. *Science* 267: 1485-1488.
48. Karin M (1999) The beginning of the end: IkappaB kinase (IKK) and NF-kappaB activation. *Journal of Biological Chemistry* 274: 27339-27342.
49. Brasier AR (2006) The NF-kappaB regulatory network. *Cardiovasc Toxicol* 6: 111-130.
50. Brasier AR, Jamaluddin M, Casola A, Duan W, Shen Q, et al. (1998) A promoter recruitment mechanism for tumor necrosis factor-alpha-induced interleukin-8 transcription in type II pulmonary epithelial cells. Dependence on nuclear abundance of Rel A, NF-kappaB1, and c-Rel transcription factors. *Journal of Biological Chemistry* 273: 3551-3561.
51. Kaltschmidt B, Kaltschmidt C (2009) NF-kappaB in the nervous system. *Cold Spring Harb Perspect Biol* 1: a001271.
52. Karin M (2006) Nuclear factor-kappaB in cancer development and progression. *Nature* 441: 431-436.
53. Masters SL, Simon A, Aksentijevich I, Kastner DL (2009) Horror autoinflammaticus: the molecular pathophysiology of autoinflammatory disease (*). *Annu Rev Immunol* 27: 621-668.
54. Henkel T, Machleidt T, Alkalay I, Kronke M, Ben-Neriah Y, et al. (1993) Rapid proteolysis of I kappa B-alpha is necessary for activation of transcription factor NF-kappa B. *Nature* 365: 182-185.
55. Deshaies RJ (1999) SCF and Cullin/Ring H2-based ubiquitin ligases. *Annu Rev Cell Dev Biol* 15: 435-467.
56. Wu G, Xu G, Schulman BA, Jeffrey PD, Harper JW, et al. (2003) Structure of a beta-TrCP1-Skp1-beta-catenin complex: destruction motif binding and lysine specificity of the SCF(beta-TrCP1) ubiquitin ligase. *Mol Cell* 11: 1445-1456.
57. Yaron A, Hatzubai A, Davis M, Lavon I, Amit S, et al. (1998) Identification of the receptor component of the IkappaBalpha-ubiquitin ligase. *Nature* 396: 590-594.
58. Dyson HJ, Komives EA (2012) Role of disorder in I kappa B-NF kappa B interaction. *IUBMB Life* 64: 499-505.
59. Cuff JA, Clamp ME, Siddiqui AS, Finlay M, Barton GJ (1998) JPred: a consensus secondary structure prediction server. *Bioinformatics* 14: 892-893.
60. Palopoli L, Rombo SE, Terracina G, Tradigo G, Veltri P (2009) Improving protein secondary structure predictions by prediction fusion. *Information Fusion* 10: 217-232.
61. Huxford T, Huang DB, Malek S, Ghosh G (1998) The crystal structure of the IkappaBalpha/NF-kappaB complex reveals mechanisms of NF-kappaB inactivation. *Cell* 95: 759-770.

62. Jacobs MD, Harrison SC (1998) Structure of an I kappa B alpha/NF-kappa B complex. *Cell* 95: 749-758.
63. Croy CH, Bergqvist S, Huxford T, Ghosh G, Komives EA (2004) Biophysical characterization of the free I kappa B alpha ankyrin repeat domain in solution. *Protein Science* 13: 1767-1777.
64. Truhlar SM, Torpey JW, Komives EA (2006) Regions of IkappaBalpha that are critical for its inhibition of NF-kappaB.DNA interaction fold upon binding to NF-kappaB. *Proc Natl Acad Sci U S A* 103: 18951-18956.
65. Cervantes CF, Markwick PRL, Sue SC, McCammon JA, Dyson HJ, et al. (2009) Functional Dynamics of the Folded Ankyrin Repeats of I kappa B alpha Revealed by Nuclear Magnetic Resonance. *Biochemistry* 48: 8023-8031.
66. Schuster M, Annemann M, Plaza-Sirvent C, Schmitz I (2013) Atypical I kappa B proteins - nuclear modulators of NF-kappa B signaling. *Cell Communication and Signaling* 11.
67. Ferreira DU, Cervantes CF, Truhlar SM, Cho SS, Wolynes PG, et al. (2007) Stabilizing IkappaBalpha by "consensus" design. *Journal of Molecular Biology* 365: 1201-1216.
68. Pons J, Evrard-Todeschi N, Bertho G, Gharbi-Benarous J, Sonois V, et al. (2007) Structural studies on 24P-I kappa B alpha peptide derived from a human I kappa B-alpha protein related to the inhibition of the activity of the transcription factor NF-kappa B. *Biochemistry* 46: 2958-2972.
69. Luckey M (2008) *Membrane structural biology : with biochemical and biophysical foundations*. Cambridge ; New York: Cambridge University Press. xi, 332 p. p.
70. Hodgkin AL, Huxley AF (1952) A quantitative description of membrane current and its application to conduction and excitation in nerve. *J Physiol* 117: 500-544.
71. Purves D (2012) *Neuroscience*. Sunderland, Mass.: Sinauer Associates.
72. Barnett MW, Larkman PM (2007) The action potential. *Pract Neurol* 7: 192-197.
73. Harmar AJ, Hills RA, Rosser EM, Jones M, Buneman OP, et al. (2009) IUPHAR-DB: the IUPHAR database of G protein-coupled receptors and ion channels. *Nucleic Acids Res* 37: D680-685.
74. Wulff H, Castle NA, Pardo LA (2009) Voltage-gated potassium channels as therapeutic targets. *Nat Rev Drug Discov* 8: 982-1001.
75. Gutman GA, Chandy KG, Grissmer S, Lazdunski M, McKinnon D, et al. (2005) International Union of Pharmacology. LIII. Nomenclature and molecular relationships of voltage-gated potassium channels. *Pharmacol Rev* 57: 473-508.
76. Papazian DM, Schwarz TL, Tempel BL, Jan YN, Jan LY (1987) Cloning of genomic and complementary DNA from Shaker, a putative potassium channel gene from *Drosophila*. *Science* 237: 749-753.

77. Doyle DA, Morais Cabral J, Pfuetzner RA, Kuo A, Gulbis JM, et al. (1998) The structure of the potassium channel: molecular basis of K⁺ conduction and selectivity. *Science* 280: 69-77.
78. Heginbotham L, Lu Z, Abramson T, MacKinnon R (1994) Mutations in the K⁺ channel signature sequence. *Biophys J* 66: 1061-1067.
79. Gulbis JM, Zhou M, Mann S, MacKinnon R (2000) Structure of the cytoplasmic beta subunit-T1 assembly of voltage-dependent K⁺ channels. *Science* 289: 123-127.
80. Kreuzsch A, Pfaffinger PJ, Stevens CF, Choe S (1998) Crystal structure of the tetramerization domain of the Shaker potassium channel. *Nature* 392: 945-948.
81. Shen NV, Pfaffinger PJ (1995) Molecular recognition and assembly sequences involved in the subfamily-specific assembly of voltage-gated K⁺ channel subunit proteins. *Neuron* 14: 625-633.
82. Long SB, Tao X, Campbell EB, MacKinnon R (2007) Atomic structure of a voltage-dependent K⁺ channel in a lipid membrane-like environment. *Nature* 450: 376-382.
83. Aggarwal SK, MacKinnon R (1996) Contribution of the S4 segment to gating charge in the Shaker K⁺ channel. *Neuron* 16: 1169-1177.
84. Seoh SA, Sigg D, Papazian DM, Bezanilla F (1996) Voltage-sensing residues in the S2 and S4 segments of the Shaker K⁺ channel. *Neuron* 16: 1159-1167.
85. Jogini V, Roux B (2007) Dynamics of the Kv1.2 voltage-gated K(+) channel in a membrane environment. *Biophysical Journal* 93: 3070-3082.
86. Schmidt D, Jiang QX, MacKinnon R (2006) Phospholipids and the origin of cationic gating charges in voltage sensors. *Nature* 444: 775-779.
87. Shafrir Y, Durell SR, Guy HR (2008) Models of voltage-dependent conformational changes in NaChBac channels. *Biophys J* 95: 3663-3676.
88. Keynes RD, Elinder F (1998) Modelling the activation, opening, inactivation and reopening of the voltage-gated sodium channel. *Proc Biol Sci* 265: 263-270.
89. Ledwell JL, Aldrich RW (1999) Mutations in the S4 region isolate the final voltage-dependent cooperative step in potassium channel activation. *J Gen Physiol* 113: 389-414.
90. Schoppa NE, Sigworth FJ (1998) Activation of Shaker potassium channels. III. An activation gating model for wild-type and V2 mutant channels. *J Gen Physiol* 111: 313-342.
91. Sigg D, Stefani E, Bezanilla F (1994) Gating current noise produced by elementary transitions in Shaker potassium channels. *Science* 264: 578-582.
92. Zagotta WN, Hoshi T, Aldrich RW (1994) Shaker potassium channel gating. III: Evaluation of kinetic models for activation. *J Gen Physiol* 103: 321-362.

93. Borjesson SI, Elinder F (2008) Structure, function, and modification of the voltage sensor in voltage-gated ion channels. *Cell Biochem Biophys* 52: 149-174.
94. Catterall WA (1986) Voltage-dependent gating of sodium channels: correlating structure and function. *Trends in Neurosciences* 9: 7-10.
95. Guy HR, Seetharamulu P (1986) Molecular model of the action potential sodium channel. *Proc Natl Acad Sci U S A* 83: 508-512.
96. Papazian DM, Bezanilla F (1997) How does an ion channel sense voltage? *News in Physiological Sciences* 12: 203-210.
97. Jiang YX, Ruta V, Chen JY, Lee A, MacKinnon R (2003) The principle of gating charge movement in a voltage-dependent K⁺ channel. *Nature* 423: 42-48.
98. Elinder F, Arhem P, Larsson HP (2001) Localization of the extracellular end of the voltage sensor S4 in a potassium channel. *Biophysical Journal* 80: 1802-1809.
99. Suh BC, Hille B (2005) Regulation of ion channels by phosphatidylinositol 4,5-bisphosphate. *Curr Opin Neurobiol* 15: 370-378.
100. Catterall WA, Cestele S, Yarov-Yarovoy V, Yu FH, Konoki K, et al. (2007) Voltage-gated ion channels and gating modifier toxins. *Toxicon* 49: 124-141.
101. Swartz KJ (2007) Tarantula toxins interacting with voltage sensors in potassium channels. *Toxicon* 49: 213-230.
102. Lee HC, Wang JM, Swartz KJ (2003) Interaction between extracellular Hanatoxin and the resting conformation of the voltage-sensor paddle in Kv channels. *Neuron* 40: 527-536.
103. Honore E, Barhanin J, Attali B, Lesage F, Lazdunski M (1994) External Blockade of the Major Cardiac Delayed-Rectifier K⁺ Channel (Kv1.5) by Polyunsaturated Fatty-Acids. *Proceedings of the National Academy of Sciences of the United States of America* 91: 1937-1941.
104. Leifert WR, McMurchie EJ, Saint DA (1999) Inhibition of cardiac sodium currents in adult rat myocytes by n-3 polyunsaturated fatty acids. *Journal of Physiology-London* 520: 671-679.
105. McKay MC, Worley JF (2001) Linoleic acid both enhances activation and blocks Kv1.5 and Kv2.1 channels by two separate mechanisms. *American Journal of Physiology-Cell Physiology* 281: C1277-C1284.
106. Poling JS, Vicini S, Rogawski MA, Salem N, Jr. (1996) Docosahexaenoic acid block of neuronal voltage-gated K⁺ channels: subunit selective antagonism by zinc. *Neuropharmacology* 35: 969-982.
107. Xiao YF, Sigg DC, Leaf A (2005) The antiarrhythmic effect of n-3 polyunsaturated fatty acids: Modulation of cardiac ion channels as a potential mechanism. *Journal of Membrane Biology* 206: 141-154.

108. Xu XP, Erichsen D, Borjesson SI, Dahlin M, Amark P, et al. (2008) Polyunsaturated fatty acids and cerebrospinal fluid from children on the ketogenic diet open a voltage-gated K channel: A putative mechanism of antiseizure action. *Epilepsy Research* 80: 57-66.
109. McNamara JO (1999) Emerging insights into the genesis of epilepsy. *Nature* 399: A15-22.
110. Lefevre F, Aronson N (2000) Ketogenic diet for the treatment of refractory epilepsy in children: A systematic review of efficacy. *Pediatrics* 105.
111. Freeman JM, Vining EPG, Pillas DJ, Pyzik PL, Casey JC, et al. (1998) The efficacy of the ketogenic diet - 1998: A prospective evaluation of intervention in 150 children. *Pediatrics* 102: 1358-1363.
112. Vining EPG, Freeman JM, Ballaban-Gil K, Camfield CS, Camfield PR, et al. (1998) A multicenter study of the efficacy of the ketogenic diet. *Archives of Neurology* 55: 1433-1437.
113. Cunnane SC, Musa K, Ryan MA, Whiting S, Fraser DD (2002) Potential role of polyunsaturates in seizure protection achieved with the ketogenic diet. *Prostaglandins Leukot Essent Fatty Acids* 67: 131-135.
114. Spector AA (2001) Plasma free fatty acid and lipoproteins as sources of polyunsaturated fatty acid for the brain. *J Mol Neurosci* 16: 159-165; discussion 215-121.
115. Fraser DD, Whiting S, Andrew RD, Macdonald EA, Musa-Veloso K, et al. (2003) Elevated polyunsaturated fatty acids in blood serum obtained from children on the ketogenic diet. *Neurology* 60: 1026-1029.
116. Taha AY, Ryan MA, Cunnane SC (2005) Despite transient ketosis, the classic high-fat ketogenic diet induces marked changes in fatty acid metabolism in rats. *Metabolism* 54: 1127-1132.
117. Vreugdenhil M, Bruehl C, Voskuyl RA, Kang JX, Leaf A, et al. (1996) Polyunsaturated fatty acids modulate sodium and calcium currents in CA1 neurons. *Proc Natl Acad Sci U S A* 93: 12559-12563.
118. Xiao Y, Li X (1999) Polyunsaturated fatty acids modify mouse hippocampal neuronal excitability during excitotoxic or convulsant stimulation. *Brain Res* 846: 112-121.
119. Biervert C, Schroeder BC, Kubisch C, Berkovic SF, Propping P, et al. (1998) A potassium channel mutation in neonatal human epilepsy. *Science* 279: 403-406.
120. Browne DL, Gancher ST, Nutt JG, Brunt ERP, Smith EA, et al. (1994) Episodic Ataxia Myokymia Syndrome Is Associated with Point Mutations in the Human Potassium Channel Gene, *Kcna1*. *Nature Genetics* 8: 136-140.
121. Cooper EC (2001) Potassium channels: How genetic studies of epileptic syndromes open paths to new therapeutic targets and drugs. *Epilepsia* 42: 49-54.

122. Eunson LH, Rea R, Zuberi SM, Youroukos S, Panayiotopoulos CP, et al. (2000) Clinical, genetic, and expression studies of mutations in the potassium channel gene KCNA1 reveal new phenotypic variability. *Annals of Neurology* 48: 647-656.
123. Singh NA, Charlier C, Stauffer D, DuPont BR, Leach RJ, et al. (1998) A novel potassium channel gene, KCNQ2, is mutated in an inherited epilepsy of newborns. *Nature Genetics* 18: 25-29.
124. Smart SL, Lopantsev V, Zhang CL, Robbins CA, Wang H, et al. (1998) Deletion of the K(v)1.1 potassium channel causes epilepsy in mice. *Neuron* 20: 809-819.
125. Bendahhou S, Cummins TR, Agnew WS (1997) Mechanism of modulation of the voltage-gated skeletal and cardiac muscle sodium channels by fatty acids. *American Journal of Physiology-Cell Physiology* 272: C592-C600.
126. Hallaq H, Smith TW, Leaf A (1992) Modulation of Dihydropyridine-Sensitive Calcium Channels in Heart-Cells by Fish Oil Fatty-Acids. *Proceedings of the National Academy of Sciences of the United States of America* 89: 1760-1764.
127. Kang JX, Leaf A (1996) Evidence that free polyunsaturated fatty acids modify Na⁺ channels by directly binding to the channel proteins. *Proceedings of the National Academy of Sciences of the United States of America* 93: 3542-3546.
128. Borjesson SI, Hammarstrom S, Elinder F (2008) Lipoelectric modification of ion channel voltage gating by polyunsaturated fatty acids. *Biophysical Journal* 95: 2242-2253.
129. Borjesson SI, Elinder F (2011) An electrostatic potassium channel opener targeting the final voltage sensor transition. *Journal of General Physiology* 137: 563-577.
130. van der Kamp MW, Shaw KE, Woods CJ, Mulholland AJ (2008) Biomolecular simulation and modelling: status, progress and prospects. *J R Soc Interface* 5 Suppl 3: S173-190.
131. Tai K, Fowler P, Mokrab Y, Stansfeld P, Sansom MS (2008) Molecular modeling and simulation studies of ion channel structures, dynamics and mechanisms. *Methods Cell Biol* 90: 233-265.
132. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, et al. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* 25: 3389-3402.
133. Karplus K, Barrett C, Hughey R (1998) Hidden Markov models for detecting remote protein homologies. *Bioinformatics* 14: 846-856.
134. Karchin R, Cline M, Mandel-Gutfreund Y, Karplus K (2003) Hidden Markov models that use predicted local structure for fold recognition: Alphabets of backbone geometry. *Proteins-Structure Function and Bioinformatics* 51: 504-514.
135. Jones DT, Taylor WR, Thornton JM (1992) A New Approach to Protein Fold Recognition. *Nature* 358: 86-89.

136. Sali A, Blundell TL (1993) Comparative Protein Modeling by Satisfaction of Spatial Restraints. *Journal of Molecular Biology* 234: 779-815.
137. Mccammon JA, Gelin BR, Karplus M (1977) Dynamics of Folded Proteins. *Nature* 267: 585-590.
138. Verlet L (1967) Computer Experiments on Classical Fluids .I. Thermodynamical Properties of Lennard-Jones Molecules. *Physical Review* 159: 98-+.
139. Hockney RW (1970) The Potential Calculation and Some Applications. *Methods in Computational Physics* 9: 136-211.
140. Hess B, Kutzner C, van der Spoel D, Lindahl E (2008) GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *Journal of Chemical Theory and Computation* 4: 435-447.
141. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, et al. (2005) Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry* 26: 1781-1802.
142. Brooks BR, Brooks CL, Mackerell AD, Nilsson L, Petrella RJ, et al. (2009) CHARMM: The Biomolecular Simulation Program. *Journal of Computational Chemistry* 30: 1545-1614.
143. Kirkwood J (1935) Statistical mechanics of fluid mixtures. *Journal of Chemical Physics* 3: 300-313.
144. Torrie GM, Valleau JP (1977) Non-Physical Sampling Distributions in Monte-Carlo Free-Energy Estimation - Umbrella Sampling. *Journal of Computational Physics* 23: 187-199.
145. Lindahl V, Lidmar J, Hess B (2014) Accelerated weight histogram method for exploring free energy landscapes. *Journal of Chemical Physics* 141.
146. Ghosh G, Wang VY-F, Huang D-B, Fusco A (2012) NF- κ B regulation: lessons from structures. *Immunological Reviews* 246: 36-58.
147. Ferreiro DU, Komives EA (2010) Molecular mechanisms of system control of NF-kappaB signaling by IkappaBalpha. *Biochemistry* 49: 1560-1567.
148. Huxford T, Hoffmann A, Ghosh G (2011) Understanding the Logic of I kappa B:NF-kappa B Regulation in Structural Terms. In: Karin M, editor. *Nf-Kb in Health and Disease*. pp. 1-24.
149. Jacobs MD, Harrison SC (1998) Structure of an IkappaBalpha/NF-kappaB complex. *Cell* 95: 749-758.
150. Traenckner EB, Baeuerle PA (1995) Appearance of apparently ubiquitin-conjugated I kappa B-alpha during its phosphorylation-induced degradation in intact cells. *Journal of cell science Supplement* 19: 79-84.
151. Huxford T, Malek S, Ghosh G (1999) Structure and mechanism in NF-kappa B/I kappa B signaling. *Cold Spring Harbor symposia on quantitative biology* 64: 533-540.

152. Hatada EN, Nieters A, Wulczyn FG, Naumann M, Meyer R, et al. (1992) The ankyrin repeat domains of the NF-kappa B precursor p105 and the protooncogene bcl-3 act as specific inhibitors of NF-kappa B DNA binding. *Proc Natl Acad Sci U S A* 89: 2489-2493.
153. Jaffray E, Wood KM, Hay RT (1995) Domain organization of I kappa B alpha and sites of interaction with NF-kappa B p65. *Mol Cell Biol* 15: 2166-2172.
154. Sun SC, Elwood J, Greene WC (1996) Both amino- and carboxyl-terminal sequences within I kappa B alpha regulate its inducible degradation. *Molecular and Cellular Biology* 16: 1058-1065.
155. Rost B (2001) Review: Protein secondary structure prediction continues to rise. *Journal of Structural Biology* 134: 204-218.
156. Simossis VA, Heringa J (2004) Optimally segmented consensus secondary structure prediction. SYMPRED server at
157. Rost B, Sander C (1994) CONSERVATION AND PREDICTION OF SOLVENT ACCESSIBILITY IN PROTEIN FAMILIES. *Proteins-Structure Function and Genetics* 20: 216-226.
158. Pollastri G, Przybylski D, Rost B, Baldi P (2002) Improving the prediction of protein secondary structure in three and eight classes using recurrent neural networks and profiles. *Proteins* 47: 228-235.
159. Lin K, Simossis VA, Taylor WR, Heringa J (2005) A simple and fast secondary structure prediction method using hidden neural networks. *Bioinformatics* 21: 152-159.
160. Jones DT (1999) Protein secondary structure prediction based on position-specific scoring matrices. *Journal of Molecular Biology* 292: 195-202.
161. Lemman JK, Mueller R, Karakas M, Woetzel N, Meiler J (2013) Simultaneous prediction of protein secondary structure and transmembrane spans. *Proteins: Structure, Function, and Bioinformatics* 81: 1127-1140.
162. Petersen B, Petersen TN, Andersen P, Nielsen M, Lundegaard C (2009) A generic method for assignment of reliability scores applied to solvent accessibility predictions. *BMC Struct Biol* 9: 51.
163. Pollastri G, McLysaght A (2005) Porter: a new, accurate server for protein secondary structure prediction. *Bioinformatics* 21: 1719-1720.
164. Rost B, Yachdav G, Liu J (2004) The PredictProtein server. *Nucleic Acids Research* 32: W321-326.
165. Cheng J, Randall AZ, Sweredoski MJ, Baldi P (2005) SCRATCH: a protein structure and structural feature prediction server. *Nucleic Acids Research* 33: W72-W76.
166. Lobley A, Sadowski MI, Jones DT (2009) pGenTHREADER and pDomTHREADER: new methods for improved protein fold recognition and superfamily discrimination. *Bioinformatics* 25: 1761-1767.

167. Sillitoe I, Cuff AL, Dessailly BH, Dawson NL, Furnham N, et al. (2013) New functional families (FunFams) in CATH to improve the mapping of conserved functional sites to 3D structures. *Nucleic Acids Research* 41: D490-498.
168. Jacobson MP, Friesner RA, Xiang Z, Honig B (2002) On the role of the crystal environment in determining protein side-chain conformations. *Journal of Molecular Biology* 320: 597-608.
169. Jacobson MP, Pincus DL, Rapp CS, Day TJ, Honig B, et al. (2004) A hierarchical approach to all-atom protein loop prediction. *Proteins* 55: 351-367.
170. Pronk S, Pall S, Schulz R, Larsson P, Bjelkmar P, et al. (2013) GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* 29: 845-854.
171. Van der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, et al. (2005) Gromacs: Fast, Flexible, and Free. *Journal of Computational Chemistry* 26: 1701-1718.
172. Scott WRP, Hunenberger PH, Tironi IG, Mark AE, Billeter SR, et al. (1999) The GROMOS biomolecular simulation program package. *Journal of Physical Chemistry A* 103: 3596-3607.
173. Hess B, Bekker H, Berendsen HJC, Fraaije JGEM (1997) LINCS: A linear constraint solver for molecular simulations. *Journal of Computational Chemistry* 18: 1463-1472.
174. Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, et al. (1995) A Smooth Particle Mesh Ewald Method. *Journal of Chemical Physics* 103: 8577-8593.
175. Fan H, Mark AE (2004) Refinement of homology-based protein structures by molecular dynamics simulation techniques. *Protein Science : A Publication of the Protein Society* 13: 211-220.
176. Elofsson A, Nilsson L (1993) How Consistent Are Molecular-Dynamics Simulations - Comparing Structure and Dynamics in Reduced and Oxidized Escherichia-Coli Thioredoxin. *Journal of Molecular Biology* 233: 766-780.
177. Mizuguchi K, Blundell T (2000) Analysis of conservation and substitutions of secondary structure elements within protein superfamilies. *Bioinformatics* 16: 1111-1119.
178. Geourjon C, Combet C, Blanchet C, Deléage G (2001) Identification of related proteins with weak sequence identity using secondary structure information. *Protein Science* 10: 788-797.
179. Simons KT, Kooperberg C, Huang E, Baker D (1997) Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and Bayesian scoring functions. *Journal of Molecular Biology* 268: 209-225.
180. Bradley P, Misura KMS, Baker D (2005) Toward high-resolution de novo structure prediction for small proteins. *Science* 309: 1868-1871.

181. Kim DE, Chivian D, Baker D (2004) Protein structure prediction and analysis using the Robetta server. *Nucleic Acids Research* 32: W526-W531.
182. Michaely P, Tomchick DR, Machius M, Anderson RG (2002) Crystal structure of a 12 ANK repeat stack from human ankyrinR. *EMBO J* 21: 6387-6396.
183. Malek S, Huxford T, Ghosh G (1998) I κ B α Functions through Direct Contacts with the Nuclear Localization Signals and the DNA Binding Sequences of NF- κ B. *Journal of Biological Chemistry* 273: 25427-25435.
184. Sue SC, Cervantes C, Komives EA, Dyson HJ (2008) Transfer of flexibility between ankyrin repeats in I kappa B alpha upon formation of the NF-kappa B complex. *Journal of Molecular Biology* 380: 917-931.
185. Huang DB, Huxford T, Chen YQ, Ghosh G (1997) The role of DNA in the mechanism of NFkappaB dimer formation: crystal structures of the dimerization domains of the p50 and p65 subunits. *Structure* 5: 1427-1436.
186. Chen FE, Huang DB, Chen YQ, Ghosh G (1998) Crystal structure of p50/p65 heterodimer of transcription factor NF-kappaB bound to DNA. *Nature* 391: 410-413.
187. Miller S, Janin J, Lesk AM, Chothia C (1987) Interior and Surface of Monomeric Proteins. *Journal of Molecular Biology* 196: 641-656.
188. Brockman JA, Scherer DC, McKinsey TA, Hall SM, Qi X, et al. (1995) Coupling of a signal response domain in I kappa B alpha to multiple pathways for NF-kappa B activation. *Mol Cell Biol* 15: 2809-2818.
189. DiDonato JA, Mercurio F, Karin M (1995) Phosphorylation of I kappa B alpha precedes but is not sufficient for its dissociation from NF-kappa B. *Mol Cell Biol* 15: 1302-1311.
190. DiDonato J, Mercurio F, Rosette C, Wu-Li J, Suyang H, et al. (1996) Mapping of the inducible IkappaB phosphorylation sites that signal its ubiquitination and degradation. *Mol Cell Biol* 16: 1295-1304.
191. Traenckner EB, Pahl HL, Henkel T, Schmidt KN, Wilk S, et al. (1995) Phosphorylation of human I kappa B-alpha on serines 32 and 36 controls I kappa B-alpha proteolysis and NF-kappa B activation in response to diverse stimuli. *EMBO J* 14: 2876-2883.
192. Whiteside ST, Ernst MK, LeBail O, Laurent-Winter C, Rice N, et al. (1995) N- and C-terminal sequences control degradation of MAD3/I kappa B alpha in response to inducers of NF-kappa B activity. *Mol Cell Biol* 15: 5339-5345.
193. Perkins ND (2006) Post-translational modifications regulating the activity and function of the nuclear factor kappa B pathway. *Oncogene* 25: 6717-6730.
194. Scheidereit C (2006) I kappa B kinase complexes: gateways to NF-kappa B activation and transcription. *Oncogene* 25: 6685-6705.
195. Groban ES, Narayanan A, Jacobson MP (2006) Conformational changes in protein loops and helices induced by post-translational phosphorylation. *Plos Computational Biology* 2: 238-250.

196. Schrodinger, LLC (2010) The PyMOL Molecular Graphics System, Version 1.3r1.
197. Parrinello M, Rahman A (1981) Polymorphic Transitions in Single-Crystals - a New Molecular-Dynamics Method. *Journal of Applied Physics* 52: 7182-7190.
198. Bussi G, Donadio D, Parrinello M (2007) Canonical sampling through velocity rescaling. *Journal of Chemical Physics* 126.
199. Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA (2001) Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proceedings of the National Academy of Sciences* 98: 10037-10041.
200. Narayanan A, Jacobson MP (2009) Computational studies of protein regulation by post-translational phosphorylation. *Curr Opin Struct Biol* 19: 156-163.
201. Michel Espinoza-Fonseca L, Kast D, Thomas DD (2007) Molecular Dynamics Simulations Reveal a Disorder-to-Order Transition on Phosphorylation of Smooth Muscle Myosin. *Biophysical Journal* 93: 2083-2090.
202. Nurik CE, Cooper DR, Ramaswamy SS, Jayaraman V Effect of Phosphorylation on Structure of C-Terminal Segment of AMPA Receptor. *Biophysical Journal* 108: 286a-287a.
203. Smart JL, McCammon JA (1999) Phosphorylation stabilizes the N-termini of alpha-helices. *Biopolymers* 49: 225-233.
204. Miranda FF, Teigen K, Thorolfsson M, Svebak RM, Knappskog PM, et al. (2002) Phosphorylation and mutations of Ser(16) in human phenylalanine hydroxylase. Kinetic and structural effects. *J Biol Chem* 277: 40937-40943.
205. Dal Peraro M, Alber F, Carloni P (2001) Ser133 phosphate-KIX interactions in the CREB-CBP complex: an ab initio molecular dynamics study. *Eur Biophys J* 30: 75-81.
206. Groban ES, Narayanan A, Jacobson MP (2006) Conformational changes in protein loops and helices induced by post-translational phosphorylation. *PLoS Comput Biol* 2: e32.
207. Steinmetz MO, Jahnke W, Towbin H, Garcia-Echeverria C, Voshol H, et al. (2001) Phosphorylation disrupts the central helix in Op18/stathmin and suppresses binding to tubulin. *EMBO Rep* 2: 505-510.
208. Yang J, Cron P, Thompson V, Good VM, Hess D, et al. (2002) Molecular mechanism for the regulation of protein kinase B/Akt by hydrophobic motif phosphorylation. *Mol Cell* 9: 1227-1240.
209. Pons J, Evrard-Todeschi N, Bertho G, Gharbi-Benarous J, Sonois V, et al. (2007) Structural studies on 24P-IkappaBalpha peptide derived from a human IkappaB-alpha protein related to the inhibition of the activity of the transcription factor NF-kappaB. *Biochemistry* 46: 2958-2972.
210. Evrard-Todeschi N, Pons J, Gharbi-Benarous J, Bertho G, Benarous R, et al. (2008) Structure of the complex between phosphorylated substrates and the SCF beta-TrCP

- ubiquitin ligase receptor: a combined NMR, molecular modeling, and docking approach. *J Chem Inf Model* 48: 2350-2361.
211. Elcock AH (2004) Molecular simulations of diffusion and association in multimacromolecular systems. *Numerical Computer Methods, Pt D* 383: 166-198.
 212. Fogolari F, Brigo A, Molinari H (2002) The Poisson-Boltzmann equation for biomolecular electrostatics: a tool for structural biology. *Journal of Molecular Recognition* 15: 377-392.
 213. Sheinerman FB, Norel R, Honig B (2000) Electrostatic aspects of protein-protein interactions. *Current Opinion in Structural Biology* 10: 153-159.
 214. Leaf A, Xiao YF, Kang JX, Billman GE (2003) Prevention of sudden cardiac death by n-3 polyunsaturated fatty acids. *Pharmacology & Therapeutics* 98: 355-377.
 215. Boland LM, Drzewiecki MM (2008) Polyunsaturated Fatty Acid Modulation of Voltage-Gated Ion Channels. *Cell Biochemistry and Biophysics* 52: 59-84.
 216. Wilder RM (1921) The effect of ketonemia on course of epilepsy. *Mayo Clin Bull* 2: 307-308.
 217. Borjesson SI, Parkkari T, Hammarstrom S, Elinder F (2010) Electrostatic Tuning of Cellular Excitability. *Biophysical Journal* 98: 396-403.
 218. Long SB, Campbell EB, MacKinnon R (2005) Voltage sensor of kv1.2: Structural basis of electromechanical coupling. *Science* 309: 903-908.
 219. Treptow W, Tarek M (2006) Environment of the gating charges in the Kv1.2 Shaker potassium channel. *Biophysical Journal* 90: L64-L66.
 220. Delemotte L, Tarek M, Klein ML, Amaral C, Treptow W (2011) Intermediate states of the Kv1.2 voltage sensor from atomistic molecular dynamics simulations. *Proceedings of the National Academy of Sciences of the United States of America* 108: 6109-6114.
 221. Jensen MO, Jogini V, Borhani DW, Leffler AE, Dror RO, et al. (2012) Mechanism of voltage gating in potassium channels. *Science* 336: 229-233.
 222. Milescu M, Bosmans F, Lee S, Alabi AA, Il Kim J, et al. (2009) Interactions between lipids and voltage sensor paddles detected with tarantula toxins. *Nature Structural & Molecular Biology* 16: 1080-1085.
 223. Mathur R, Zheng J, Yan YY, Sigworth FJ (1997) Role of the S3-S4 linker in Shaker potassium channel activation. *Journal of General Physiology* 109: 191-199.
 224. Gonzalez C, Rosenman E, Bezanilla F, Alvarez O, Latorre R (2000) Modulation of the Shaker K⁺ channel gating kinetics by the S3-S4 linker. *Journal of General Physiology* 115: 193-207.
 225. Eswar N, Webb B, Marti-Renom MA, Madhusudhan MS, Eramian D, et al. (2007) Comparative protein structure modeling using MODELLER. *Curr Protoc Protein Sci* Chapter 2: Unit 2 9.

226. Melo F, Sanchez R, Sali A (2002) Statistical potentials for fold assessment. *Protein Science* 11: 430-448.
227. Berger O, Edholm O, Jahnig F (1997) Molecular dynamics simulations of a fluid bilayer of dipalmitoylphosphatidylcholine at full hydration, constant pressure, and constant temperature. *Biophysical Journal* 72: 2002-2013.
228. Wolf MG, Hoefling M, Aponte-Santamaria C, Grubmuller H, Groenhof G (2010) g_membed: Efficient Insertion of a Membrane Protein into an Equilibrated Lipid Bilayer with Minimal Perturbation. *Journal of Computational Chemistry* 31: 2169-2174.
229. Lindorff-Larsen K, Piana S, Palmo K, Maragakis P, Klepeis JL, et al. (2010) Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins-Structure Function and Bioinformatics* 78: 1950-1958.
230. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of Simple Potential Functions for Simulating Liquid Water. *Journal of Chemical Physics* 79: 926-935.
231. Best RB, Zhu X, Shim J, Lopes PEM, Mittal J, et al. (2012) Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone phi, psi and Side-Chain chi(1) and chi(2) Dihedral Angles. *Journal of Chemical Theory and Computation* 8: 3257-3273.
232. Yesselman JD, Price DJ, Knight JL, Brooks CL (2012) MATCH: An Atom-Typing Toolset for Molecular Mechanics Force Fields. *Journal of Computational Chemistry* 33: 189-202.
233. Feller SE, Gawrisch K, MacKerell AD (2002) Polyunsaturated fatty acids in lipid bilayers: Intrinsic and environmental contributions to their unique physical properties. *Journal of the American Chemical Society* 124: 318-326.
234. Hess B (2008) P-LINCS: A parallel linear constraint solver for molecular simulation. *Journal of Chemical Theory and Computation* 4: 116-122.
235. Henrion U, Renhorn J, Borjesson SI, Nelson EM, Schwaiger CS, et al. (2012) Tracking a complete voltage-sensor cycle with metal-ion bridges. *Proceedings of the National Academy of Sciences of the United States of America* 109: 8552-8557.
236. Knight CJ, Hub JS (Submitted) MemGen: A general web server for the setup of lipid membrane simulation systems.
237. Xiao YF, Kang JX, Morgan JP, Leaf A (1995) Blocking Effects of Polyunsaturated Fatty-Acids on Na⁺ Channels of Neonatal Rat Ventricular Myocytes. *Proceedings of the National Academy of Sciences of the United States of America* 92: 11000-11004.
238. Hong MP, Kim HI, Shin YK, Lee CS, Park M, et al. (2004) Effects of free fatty acids on sodium currents in rat dorsal root ganglion neurons. *Brain Research* 1008: 81-91.
239. Danthi SJ, Enyeart JA, Enyeart JJ (2005) Modulation of native T-type calcium channels by omega-3 fatty acids. *Biochemical and Biophysical Research Communications* 327: 485-493.

240. Decher N, Streit AK, Rapedius M, Netter MF, Marzian S, et al. (2010) RNA editing modulates the binding of drugs and highly unsaturated fatty acids to the open pore of Kv potassium channels. *Embo Journal* 29: 2101-2113.
241. Liin SI, Silvera Ejneby M, Barro-Soria R, Skarsfeldt MA, Larsson JE, et al. (2015) Polyunsaturated fatty acid analogs act antiarrhythmically on the cardiac IKs channel. *Proc Natl Acad Sci U S A* 112: 5714-5719.
242. Zheng HF, Li XL, Jin ZY, Sun JB, Li ZL, et al. (2005) Effects of unsaturated fatty acids on calcium-activated potassium current in gastric myocytes of guinea pigs. *World Journal of Gastroenterology* 11: 672-675.
243. Eldho NV, Feller SE, Tristram-Nagle S, Polozov IV, Gawrisch K (2003) Polyunsaturated docosahexaenoic vs docosapentaenoic acid - Differences in lipid matrix properties from the loss of one double bond. *Journal of the American Chemical Society* 125: 6409-6421.
244. Soubias O, Gawrisch K (2007) Docosahexaenoyl chains isomerize on the sub-nanosecond time scale. *Journal of the American Chemical Society* 129: 6678-6679.
245. Feller SE (2008) Acyl chain conformations in phospholipid bilayers: a comparative study of docosahexaenoic acid and saturated fatty acids. *Chemistry and Physics of Lipids* 153: 76-80.
246. Feller SE, Gawrisch K (2005) Properties of docosahexaenoic-acid-containing lipids and their influence on the function of rhodopsin. *Current Opinion in Structural Biology* 15: 416-422.
247. Ottosson NE, Liin SI, Elinder F (2014) Drug-induced ion channel opening tuned by the voltage sensor charge profile. *Journal of General Physiology* 143: 173-182.
248. Xu YP, Ramu Y, Shin HG, Yamakaze J, Lu Z (2013) Energetic role of the paddle motif in voltage gating of Shaker K⁺ channels. *Nature Structural & Molecular Biology* 20: 574-+.
249. Kasimova MA, Zaydman MA, Cui JM, Tarek M (2015) PIP₂-dependent coupling is prominent in Kv7.1 due to weakened interactions between S4-S5 and S6. *Scientific Reports* 5.
250. Chen L, Zhang Q, Qiu Y, Li Z, Chen Z, et al. (2015) Migration of PIP₂ lipids on voltage-gated potassium channel surface influences channel deactivation. *Sci Rep* 5: 15079.
251. Pathak MM, Yarov-Yarovoy V, Agarwal G, Roux B, Barth P, et al. (2007) Closing in on the resting state of the Shaker K(+) channel. *Neuron* 56: 124-140.
252. Vargas E, Bezanilla F, Roux B (2011) In search of a consensus model of the resting state of a voltage-sensing domain. *Neuron* 72: 713-720.
253. Campos FV, Chanda B, Roux B, Bezanilla F (2007) Two atomic constraints unambiguously position the S4 segment relative to S1 and S2 segments in the closed

- state of Shaker K channel. *Proceedings of the National Academy of Sciences of the United States of America* 104: 7904-7909.
254. Lin MCA, Hsieh JY, Mock AF, Papazian DM (2011) R1 in the Shaker S4 occupies the gating charge transfer center in the resting state. *Journal of General Physiology* 138: 155-163.
255. Villalba-Galea CA, Sandtner W, Starace DM, Bezanilla F (2008) S4-based voltage sensors have three major conformations. *Proceedings of the National Academy of Sciences of the United States of America* 105: 17600-17607.
256. Schwaiger CS, Bjelkmar P, Hess B, Lindahl E (2011) 3(10)-Helix Conformation Facilitates the Transition of a Voltage Sensor S4 Segment toward the Down State. *Biophysical Journal* 100: 1446-1454.
257. Bezanilla F (2000) The voltage sensor in voltage-dependent ion channels. *Physiological Reviews* 80: 555-592.
258. Schwaiger CS, Borjesson SI, Hess B, Wallner B, Elinder F, et al. (2012) The Free Energy Barrier for Arginine Gating Charge Translation Is Altered by Mutations in the Voltage Sensor Domain. *Plos One* 7.
259. Tao X, Lee A, Limapichat W, Dougherty DA, MacKinnon R (2010) A Gating Charge Transfer Center in Voltage Sensors. *Science* 328: 67-73.
260. Lacroix JJ, Bezanilla F (2011) Control of a final gating charge transition by a hydrophobic residue in the S2 segment of a K⁺ channel voltage sensor. *Proceedings of the National Academy of Sciences of the United States of America* 108: 6444-6449.
261. Kosower EM (1985) A Structural and Dynamic Molecular-Model for the Sodium-Channel of *Electrophorus-Electricus*. *Febs Letters* 182: 234-242.
262. Clayton GM, Altieri S, Heginbotham L, Unger VM, Morais-Cabral JH (2008) Structure of the transmembrane regions of a bacterial cyclic nucleotide-regulated channel. *Proceedings of the National Academy of Sciences of the United States of America* 105: 1511-1515.
263. Shafirir Y, Durell SR, Guy HR (2008) Models of voltage-dependent conformational changes in NaChBac channels. *Biophysical Journal* 95: 3663-3676.
264. Bjelkmar P, Niemela PS, Vattulainen I, Lindahl E (2009) Conformational Changes and Slow Dynamics through Microsecond Polarized Atomistic Molecular Simulation of an Integral Kv1.2 Ion Channel. *Plos Computational Biology* 5.
265. Schow EV, Freitas JA, Gogna K, White SH, Tobias DJ (2010) Down-State Model of the Voltage-Sensing Domain of a Potassium Channel. *Biophysical Journal* 98: 2857-2866.
266. Khalili-Araghi F, Jogini V, Yarov-Yarovoy V, Tajkhorshid E, Roux B, et al. (2010) Calculation of the Gating Charge for the Kv1.2 Voltage-Activated Potassium Channel. *Biophysical Journal* 98: 2189-2198.

267. Papazian DM, Shao XM, Seoh SA, Mock AF, Huang Y, et al. (1995) Electrostatic Interactions of S4 Voltage Sensor in Shaker K⁺ Channel. *Neuron* 14: 1293-1301.

Acknowledgements

This thesis was made possible with the contributions of many individuals in my life. The people I would like to give special thanks to are:

Matthias Stein, my supervisor and mentor who has believed in me and encouraged me since day one. Thank you for giving me the freedom to explore any scientific question I had and offering your full support throughout this period. I have learned, grown, and flourished into a scientist under your guidance. You are a true scientist who keeps an open mind and is not afraid of new ideas.

My collaborators in Germany: Prof. Michael Naumann for introducing me into the wonderful world of $\text{I}\kappa\text{B}\alpha$ and all the fruitful discussions that followed. Dr. Serdar Durdagi for critical reading of the $\text{I}\kappa\text{B}\alpha$ paper and introducing me to protein-protein docking. Prof. Helmut Weiss and Prof. Andreas Seidel-Morgenstern for welcoming me into the faculty and giving me the chance to do real science in a stimulating environment.

The members of the PCG group at the Max Planck, that made life at the institute and Magdeburg much more pleasurable, and in their own ways contributed to many happy memories of Germany.

A special thanks to Erik Lindahl who took me under his wings at SciLifeLab and introduced me to the world of voltage-gated ion channels. I will always remember your encouraging words at the end of every meeting where you would say 'Great job!' no matter what. My collaborators Magnus Andersson for all his help with the PUFA project and for always being optimistic and super efficient. Fredrik Elinder for the invaluable insights he brought to the PUFA project.

The Lindahl lab members who made life with Gromacs a lot faster and easier. The Carlsson lab members for all the fun we had exploring Stockholm and introducing me to protein-ligand docking.

The people I met during the course of my PhD with whom I became good friends: Dawid for being the best office mate in the world and my ally at the Max Planck. Jane and Kasia for all our great talks, trips, and great laughs we shared together. Narges for all our adventures together, for always making me laugh and introducing me to the world of TCK. Viveca and Özge for making days on floor 6 a more enjoyable experience, our movie nights and our scientific discussions.

Felicia for always being there for me no matter where in the world you were. Thank you for being the best friend I could ever ask for and believing in me. Rachel for being the most awesome friend since day one at Gillespie's. Love both of you girls.

Elly and Ila for sharing my DNA and being my partners in crime. This thesis would not have been possible without you girls. You are my confidantes and make my life richer by simply being the remarkable souls that you are.

Mom and Dad, I dedicate this thesis to you. You have been the backbone of my life and your faith in me has been relentless. Thank you for your unconditional love and making me see the importance of education from an early age.

List of Publications

The content of this dissertation contains previously published material in Open Access journals that permit unrestricted use, distribution, and reproduction in any medium.

- I. Yazdi S, Durdagi S, Naumann M and Stein M (2015) Structural modeling of the N-terminal signal-receiving domain of I κ B α . *Front. Mol. Biosci.* 2:32. doi: 10.3389/fmolb.2015.00032.
- II. Yazdi S, Stein M, Elinder F, Andersson M, Lindahl E (2016) The Molecular Basis of Polyunsaturated Fatty Acid Interactions with the *Shaker* Voltage-Gated Potassium Channel. *PLoS Comput Biol* 12(1): e1004704. doi:10.1371/journal.pcbi.1004704.

Curriculum vitae

Name Samira Yazdi
Date of birth May 27, 1981
Nationality Canadian
Address Löjtnantsgatan 11, 11550 Stockholm

Education

2011-2016 Ph.D. The Max Planck Institute for Dynamics of Complex Technical Systems, Germany. Dissertation: The Structural Dynamics of Soluble and Membrane Proteins Explored through Molecular Simulations (Supervisor: Matthias Stein)

2007-2009 M.Sc. in Bioinformatics. Stockholm University, Sweden. Dissertation: Homology modeling and docking help in identifying pharmacophores in the three human galanin receptor subtypes (Supervisor: Arne Elofsson)

2001-2005 B.Sc. in Biology and Statistics. University of Toronto, Canada.

2000 Scottish Highers (High School Diploma). James Gillespie's High School, UK.

Academic experience

Mar 2010 – Jul 2010 Research Assistant at the Molecular Modeling Group, Department of Biosciences and Nutrition at Karolinska Institute, Stockholm

Sep 2009 – Feb 2010 Research Assistant at the Unit of Functional Pharmacology, Department of Neuroscience at Uppsala University, Uppsala

Jun 2009 – Aug 2009 Research Assistant at the Center for Biomembrane Research (CBR) at the Department of Biophysics and Biochemistry at Stockholm University, Stockholm