Of proteins and lipids:
A molecular dynamics study of membrane-bound Rab5.

Dissertation
zur Erlangung des akademischen Grades

doctor rerum naturalium
(Dr. rer. nat.)

von M.Sc. Eileen Münzberg

geb. am 09.06.1988 in Berlin

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

Promotionskommission: Prof. Dr.-Ing. Andreas Seidel-Morgenstern

Prof. Dr. rer. nat. Dieter Schinzer

Prof. Dr. rer. nat. Stefan Kalkhof

Dr. rer. nat. Matthias Stein

eingereicht am: 06.11.2018

Promotionskolloquium am: 01.02.2019

Only entropy comes easy.
- Anton Chekhov (Russ. dramatist)

Abstract

Cellular trafficking involves spatially distinct organelles which are related by a sophisticated vesicular transport system. Vesicles from the plasma membrane are transported to early endosomes, which serve as sorting stations directing cargo either to the trans-Golgi network, to be degradaded via late endosomes and lysosomes, or back to the plasma membrane for recycling. The membrane-bound small GTPase Rab5 is the master regulator of endocytic cargo transport, and controls, amongst others, vesicular transport between the plasma membrane and the early endosome, early endosome fusion and events in phagocytosis. Due to its prominent role in endocytosis, Rab5 decides about the fate of internalized cargo including cell surface signaling receptors, and is thus indirectly involved in a variety of infectious diseases, neuronal disorders and cancer. As a small GTPase, Rab5 consists of a structurally conserved catalytic GTPase (G) domain, which is able to bind guanosine di- or triphosphate (GDP/GTP) and to hydrolyze GTP with a slow intrinsic rate. Rab5 serves as a molecular switch shuttling between an inactive, GDP-bound and an active, GTP-bound state, the latter recruiting effector proteins to the membrane. The G domain is connected to the membrane via a long C-terminal hypervariable region (HVR). Two covalently attached geranylgeranyl chains at the Rab5 extreme C-terminus enable bilayer anchorage. While Rab5(GTP) is solely membrane-localized, Rab5(GDP) shuttles between the bilayer surface and the cytoplasm in a RabGDI (GDI)-dependent manner.

Until today, Rab5 crystal structures contain only G domain residues and are missing the flexible protein N- and C-termini. The geranylgeranyl post-translational modification represents another challenge with regard to protein synthesis for Rab5 investigation by wet lab experiments. These issues render preparation and experimental analysis of Rab5 extremely labor-intensive, costly and time-consuming. Facing these experimental challenges, Rab5 characterization by computational studies represents a promising alternative strategy.

A comparative modeling approach combined with molecular dynamics simulations was used in this thesis to obtain full-length Rab5(GDP) and Rab5(GTP) structures and to investigate their dynamics and interactions with membranes and protein binding partners. In the complex charged membrane, but not in the uncharged model bilayers, the monomeric Rab5 G domain tilted toward the bilayer surface in a nucleotide-specific manner. The switch regions, which are essential for binding partner recognition, predominately adapted a buried conformation in the GDP-bound state. The GTP-bound state was mainly associated with fully accessible switch regions on the protein side distal to the membrane. The tilting of the G domain is a necessary prerequisite which enables the small GTPase to recruit and bind effectors in very close proximity to the membrane. The G domain re-orientation was found to be mainly driven by long-range electrostatic interactions between basic residues of the HVR and the negatively charged PI(3)P signaling lipid in the early endosome membrane. These interactions were shown to slow down the peptide lateral diffusion and increased the Gibbs free energy associated with the bilayer-to-water transfer of the membrane anchor. Long-term coarse-grained simulations revealed an active recruitment of PI(3)P to the HVR, which serves the formation of spatially distinct Rab5- and PI(3)P-enriched signaling platforms.

The capability of Rab5 to form dimers is controversially debated in literature. Here, dimerization of membrane-bound Rab5 into low-affinity complexes was observed via distinct dimerization interfaces in both nucleotide states. In case of Rab5(GDP), dimerization occurred preferentially via the switch I β-interface, leaving the switch II region largely accessible. In contrast, helical interfaces of the switch II region contributed predominately to dimerization of Rab5(GTP). Based on the protein surface areas involved in dimerization, GDP-bound Rab5 dimers were more stable compared to Rab5(GTP) dimers. Rab5(GDP) dimerization at high protein concentrations may serve to generate

a membrane-associated pool of immediately available Rab5(GDP) independent of GDI binding. In Rab5(GTP) dimers, at least one monomer exhibited an improved switch region accessibility compared to monomeric Rab5 due to a more exposed position farther from the membrane surface. The preference of both Rab5 activation states for different dimerization sites reflects their participation in specific protein-protein interactions. Residues in the switch regions essential for effector recognition were mainly accessible in Rab5(GTP) dimers. GDI is a major binding partner of Rab5(GDP) and was shown to interact with the small GTPase primarily via switch II residues that are hardly involved in dimerization. GDI releases Rab5(GDP) from the bilayer into the cytoplasm, keeps it soluble by accommodating the hydrophobic geranylgeranyl chains in a binding pocket, and recycles it back to donor membranes. Based on the identified interactions in the Rab5:GDI complex at the membrane and in cytoplasm, a scheme of the membrane extraction process was suggested. First, approaching GDI was found to contact the G domain switch regions of membrane-bound, tilted Rab5. Subsequently, GDI re-oriented, thereby establishing contacts to the Rab5 HVR and simultaneously bringing the GDI prenyl binding pocket in close proximity to the membrane surface. Interactions between the Rab5 HVR and GDI turned out to be of great importance for the orientation and binding of the G domain to GDI and thus, for the overall complex stability. Hydrophobic patches were identified which enclose the binding pocket and guided the geranylgeranyl chains into the cavity. In a last step, the tightly bound Rab5:GDI complex diffused from the membrane.

Rab5 dysfunctions due to mutations or pathogens cause a disturbance of endocytic transport with enormous consequences for cellular recycling and degradation processes. This thesis provides a deep understanding of Rab5 dynamics, lipid- and protein-interactions in intact cells. This lays the foundation for better insights into distorted interactions and mechanisms in infected or mutated cells. Thus, residues predicted to be essential for interactions with certain effectors could be specifically mutated to influence signaling. Furthermore, the observed nucleotide-dependent Rab5 monomer orientations provide a potential drug target in order to cure Rab5-related dysfunctions.

Zusammenfassung

Der zelluläre Transport umfasst räumlich getrennte Organellen, die durch ein differenziertes vesikuläres System miteinander verbunden sind. Vesikel von der Plasmamembran werden zunächst zu frühen Endosomen transportiert. Diese dienen als Sortierstationen, die ihre Fracht entweder in das trans-Golgi-Netzwerk, über späte Endosomen in Lysosomen zum Abbau oder zurück an die Plasmamembran zur Wiederverwertung weiterleiten. Die membrangebundene kleine GTPase Rab5 ist der Hauptregulator endozytotischer Vorgänge und kontrolliert, unter anderem, den vesikulären Transport zwischen der Plasmamembran und den frühen Endosomen, die Fusion von frühen Endosomen und Vorgänge in der Phagozytose. Aufgrund seiner Funktion in der Endozytose bestimmt Rab5 die Transportvorgänge aufgenommener Fracht, wie z.B. von Oberflächenrezeptoren, und ist damit indirekt an einer Vielzahl von Infektionskrankheiten, neuronalen Störungen und Tumorerkrankungen beteiligt. Wie alle kleinen GTPasen besteht Rab5 aus einer strukturell konservierten, katalytisch aktiven GTPase (G)-Domäne, welche Guanosindi- oder Guanosintriphosphat (GDP/GTP) bindet und GTP mit einer geringen intrinsischen Rate hydrolysiert. Rab5 dient als molekularer Schalter, der zwischen der inaktiven, GDP-gebundenen und der aktiven, GTP-gebundenen Form wechselt, wobei Letztere Effektorproteine an die Membran rekrutiert. Die G-Domäne ist mit der Membran über eine lange C-terminale hypervariable Region (HVR) verknüpft. Zwei kovalent gebundene Geranylgeranylgruppen am äußeren Rab5 C-Terminus ermöglichen eine Verankerung in der Membran. Während Rab5(GTP) ausschließlich an der Membran lokalisiert ist, pendelt Rab5(GDP) in Abhängigkeit von RabGDI (im Folgenden GDI) zwischen der Membranoberfläche und dem Zytoplasma.

Bis heute beinhalten Rab5 Kristallstrukturen lediglich Aminosäurereste der G-Domäne, während die flexiblen N- und C-Termini fehlen. Die Synthese von Rab5 für experimentelle Untersuchungen wird durch die post-translatorische Modifizierung mit Geranylgeranylgruppen erschwert. Diese Aspekte machen die Vorbereitung und Durchführung von Laborexperimenten sehr aufwendig, kostspielig und zeitintensiv. Angesichts der experimentellen Herausforderungen stellt die Charakterisierung von Rab5 mittels computergestützter Studien eine vielversprechende alternative Strategie dar.

In der vorliegenden Arbeit wurden Homologiemodellierung und Molekulardynamik-Simulationen genutzt, um vollständige Rab5(GDP) und Rab5(GTP) Modelle zu generieren und diese hinsichtlich ihrer Dynamik und Interaktionen mit Membranen und Proteinen zu untersuchen. Im Gegensatz zu den ungeladenen Membranmodellen, orientierte sich die monomere G-Domäne in Abhängigkeit von dem gebundenen Nukleotid in der geladenen Membran hin zur Membranoberfläche. Die für die Erkennung durch Bindungspartner essenziellen Switchregionen nahmen in der GDP-gebundenen Form vorrangig eine abgeschirmte Konformation an. Die GTP-gebundene Form hingegen war hauptsächlich mit vollständig zugänglichen Switchregionen auf der membranabgewandten Seite des Proteins assoziiert. Das Kippen der G-Domäne ist eine notwendige Voraussetzung für die Rekrutierung und das Binden von Effektorproteinen durch die kleine GTPase in großer Nähe zur Membran. Die Umorientierung der G-Domäne wurde hauptsächlich durch elektrostatische Interaktionen zwischen den basischen Resten der HVR und dem negativ geladenen PI(3)P Signallipid der Membran des frühen Endosoms gesteuert. Es zeigte sich, dass diese Interaktionen die laterale Diffusion des Peptides verlangsamten. Die mit dem Transfer des Membranankers aus der Membran assoziierte Gibbs-Energie war in der geladenen Membran deutlich erhöht. Durch den Einsatz von coarse-grained Simulationen wurde eine Rekrutierung von PI(3)P an die HVR beobachtet, welche der Ausbildung von räumlich abgegrenzten, mit Rab5 und PI(3)P angereicherten, Signalplattformen dient.

Die Fähigkeit von Rab5 zur Dimerisierung wird in der Fachliteratur kontrovers diskutiert. In der vorliegenden Arbeit wurde die Dimerisierung von membrangebundenem Rab5(GDP) und Rab5(GTP) in

Komplexe mit geringer Affinität über unterschiedliche Dimerisierungsflächen beobachtet. In Rab5(GDP) fand die Dimerisierung bevorzugt über die Switch I β-Faltblätter statt, während die Switchregion II größtenteils zugänglich blieb. Im Gegensatz dazu, trugen die α-Helices der Switchregion II vorrangig zu der Dimerisierung von Rab5(GTP) bei. Basierend auf den an der Dimerisierung beteiligten Proteinoberflächen waren GDP-gebundene Rab5 Dimere stabiler als GTP-gebundene. Die Dimerisierung von Rab5(GDP) bei hohen Proteinkonzentrationen diente der Anlage eines inaktiven, membranassoziierten Reservoirs von sofort verfügbaren Rab5 Molekülen. Im Fall von Rab5(GTP) Dimeren wies zumindest ein Monomer eine gegenüber dem unkomplexierten Rab5 verbesserte Erreichbarkeit der Switchregionen auf. Dies geschah durch eine exponiertere Lage abseits der Membranoberfläche. Die jeweilige Präferenz eines Rab5 Aktivierungszustandes für unterschiedliche Dimerisierungsflächen spiegelte ihre Rolle in spezifischen Protein-Protein-Interaktionen wider. Aminosäuren in den Switchregionen, welche der Erkennung von Effektorproteinen dienen, waren in den Rab5(GTP) Dimeren weitgehend zugänglich. GDI ist ein wichtiger Interaktionspartner von Rab5(GDP). Es zeigte sich, dass Rab5(GDP)-GDI Interaktionen vorrangig über die Rab5 Switchregion II stattfanden, welche kaum an der Dimerisierung beteiligt waren. GDI extrahiert Rab5(GDP) aus der Membran in das Zytoplasma, stabilisiert den löslichen Komplex durch die Aufnahme der hydrophoben Geranylgeranylgruppen in eine Bindungstasche und transferiert Rab5 zurück auf eingehende Vesikel. Basierend auf den identifizierten Interaktionen im membrangebundenen und löslichen Rab5:GDI Komplex wurde ein Schema des Membranextraktionsmechanismus entworfen. Sich näherndes GDI kontaktierte zunächst die Switchregionen der G-Domain des membrangebundenen Rab5. GDI richtete sich anschließend aus, was mit der Ausbildung von Kontakten zur Rab5 HVR einherging und gleichzeitig die Prenylbindungstasche in der Nähe der Membranoberfläche positionierte. Die Interaktionen zwischen der Rab5 HVR und GDI waren dabei von großer Bedeutung für die Orientierung und das Binden der G-Domäne an GDI und damit für die Stabilität des gesamten Komplexes. Hydrophobe Bereiche entlang der Bindungstasche, welche die Geranylgeranylgruppen in die Tasche leiteten, wurden identifiziert. In einem letzten Schritt diffundierte der mit starker Affinität bindende Rab5:GDI Komplex von der Membranoberfläche weg.

Störungen der Rab5 Funktion aufgrund von Mutationen oder Pathogenen verursachen eine Beeinträchtigung der endozytotischen Transportvorgänge mit enormen Auswirkungen auf die zelluläre Wiederverwertungs- und Abbaumaschinerie. Die vorliegende Arbeit ermöglicht tiefgreifende Einblicke in die Dynamiken von Rab5 und dessen Interaktionen mit Lipiden und Proteinen in intakten Zellen. Dies liefert die Voraussetzungen für das Verständnis gestörter Interaktionen und Mechanismen in infizierten oder mutierten Zellen. Ausgehend von der Vorhersage wichtiger Aminosäurereste für Interaktionen mit bestimmten Effektoren, können diese gezielt mutiert werden, um Signalkaskaden zu beeinflussen. Zudem stellen die beobachteten nukleotidabhängigen Rab5 Monomer Orientierungen einen möglichen Ansatzpunkt für die Entwicklung von Inhibitoren und Arzneimitteln zur Bekämpfung von mit Rab5 assoziierten Funktionsstörungen dar.

Contents

Nomenclature

Roman symbols

Symbol	Description	Unit
A	Helmholtz Energy	J
a	acceleration	$\mathrm{m\,s^{-2}}$
D	diffusion coefficient	$\mathrm{m^2\,s^{-1}}$
d	distance	m
DOF	degree of freedom	-
F	force	N
G	Gibbs free energy	J
i	insertion depth	m
k	spring constant	$\mathrm{N\,m^{-1}}$
l	bond length	m
m	mass	kg
N	total number of atoms	-
n	constant	-
$\vec{n}$	membrane normal vector	-
P	propability distribution	-
p	propability	-
q	atomic charge	C
$\tilde{q}$	charge density	$\mathrm{C\,m^{-1}}$
r	radius	m
S	entropy / order parameter tensor	$\mathrm{J\,K^{-1}}$ / -
s	position vector	m
T	temperature	K
t	time	s
U	potential energy	J
u	velocity	$\mathrm{m\,s^{-1}}$
V	principal component	-
$\mathbf{v}$	eigenstate / eigenvector	-
w	weight	-
y	principal component coefficient	-
Z	partition function	-

Greek symbols

Symbol	Description	Unit
β	thermodynamic beta	$\mathrm{J^{-1}}$
γ	phase shift angle	rad
Δ	measure of the influence on the principal component in dPCA	rad
ε	depth of the Lennard-Jones potential well	-

Symbol	Description	Unit
θ	bond angle	$^\circ$
λ	Lyapunov exponent / eigenvalue	- / -
ξ	reaction coordinate	-
ρ	probability density	-
σ	intermolecular distance	m
Φ	protein backbone torsional angle / phase angle	$^\circ$ / rad
φ	interaction potential	$\mathrm{J\,mol^{-1}}$
ϕ	phase angle	rad
χ	angle in Rab5 dimer	$^\circ$
ψ	protein backbone torsional angle	$^\circ$
ω	protein backbone torsional angle / torsional angle	$^\circ$/rad

Indices

Index	Description
bend	angle bending
bias	biased
CD	carbon-deuterium
COM	center of mass
cyt	cytoplasmic
elec	electrostatic
eq	equilibrium
f	unconstrained degrees of freedom
GG	geranylgeranyl
gyr	gyration
i	atomic number / window
ij	pair of atoms (i and j)
inos	inositol ring
insert	membrane-inserted
lr	long-range
n	number
non-bond	non-bonded
PN	phosphor-nitrogen
sr	short-range
stretch	bond stretching
swI	switch I region
swII	switch II region
tail	lipid tail
tilt	tilting
torsion	angle torsion
unbias	unbiased
us	umbrella sampling
x,y,z	directions
α	α-atom / -helix
β	β-sheet

Constants

Constant	Description	Value
e_c	charge of an electron	$1.6022 \cdot 10^{-19}\,\mathrm{C}$
k_B	Boltzmann constant	$1.3806 \cdot 10^{-23}\,\mathrm{m^2\,kg\,s^{-2}\,K^{-1}}$
ε_0	dielectric constant	$8.8541\,\mathrm{A\,s\,V^{-1}\,m^{-1}}$

Abbreviations

Abbreviation	Description
AD	Alzheimer's Disease
ADP	adenosine diphosphate
APBS	adaptive Poisson-Boltzmann solver
APP	amyloid precursor protein
APP-βCTF	amyloid precursor protein β-cleavage C-terminal fragment
APPL1	adaptor protein containing pleckstrin homology domain, phosphotyrosine binding (PTB) domain, and leucine zipper motif
Arf	ADP ribosylation factor
ATP	adenosine triphosphate
BD	Brownian dynamics
BDNF	brain-derived neutrotrophic factor
BPTI	bovine pancreatic trypsin inhibitor
CCR	C-terminus coordinating region
CCV	clathrin-coated vesicle
COSMO	conductor-like screening model
CPU	central processing unit
DFT	density functional theory
DOPE	discrete optimized protein energy
dPCA	dihedral principal component analysis
DSSP	define secondary structure of proteins
EE	early endosome
EEA1	early endosome antigen 1
EGF	epidermal growth factor
ER	endoplasmic reticulum
GAP	GTPase activating protein
GDF	GDP dissociation inhibitor displacement factor
GDI	GDP dissociation inhibitor
G domain	GTPase domain
GDP	guanosine diphosphate
GEF	guanine nucleotide exchange factor
GG	geranylgeranyl
GGPP	geranylgeranyl pyrophosphate
GGTase	geranylgeranyl transferase
GNP	phosphoaminophosphonic acid-guanylate ester
GridMAT	grid-based membrane analysis tool
GTP	guanosine triphosphate
HPC	high-performance computing
H-Ras	Harvey rat sarcoma
HVR	hypervariable region
ILV	intraluminal vesicle

Abbreviation	Description
INPP5B	inositol polyphosphate-5-phosphatase B
JDA	jump distance analysis
K-Ras	Kirsten rat sarcoma
ld	liquid-disordered
LE	late endosome
LJ	Lennard-Jones
lo	liquid-ordered
MCMM	Monte Carlo multiple minimum
MD	molecular dynamics
MEL	mobile effector loop
MM	molecular mechanics
MP2	Møller-Plesset pertubation
MSD	mean square displacement
mTOR	mechanistic target of rapamycin
NF-κB	nuclear factor κB
NGF	nerve growth factor
NMR	nuclear magnetic resonance
N-Ras	neuroblastoma rat sarcoma
OCRL	oculocerebrorenal syndrome of Lowe protein
PBC	periodic boundary conditions
PC	phosphatidylcholine
PE	phosphatidylethanolamine
PG	phosphatidylglycerol
PI	phosphatidylinositol
PI(3)P	phosphatidylinositol 3-phosphate
P-loop	phosphate-binding loop
PME	particle mesh Ewald
PMF	potential of mean force
POPC	palmitoyl-oleoyl-phosphatidylcholine
POPE	palmitoyl-oleoyl-phosphatidylethanolamine
POPS	palmitoyl-oleoyl-phosphatidylserine
POVME	pocket volume measurer
PS	phosphatidylserine
PSM	palmitoyl-sphingomyelin
QM	quantum mechanics
Rab	Ras gene from rat brain
Raf	rapidly accelerated fibrosarcoma
Ran	Ras-related nuclear
Ras	rat sarcoma
RBP	Rab binding platform
REP	Rab escort protein
RGGT	Rab geranylgeranyl transferase
Rho	Ras homologous
RIN1	Ras and Rab interactor 1
RMSD	root mean square deviation
RMSF	root mean square fluctuation
RN-tre	related to the N-terminus of tre oncogene
RTK	receptor tyrosine kinase
SASA	solvent-accessible surface area
SEEKR	simulation enabled estimation of kinetic rates
SM	sphingomyelin

Abbreviation	**Description**
SMD	steered molecular dynamics
TCL	tool command language
TIP3P	transferable intermolecular potential 3P
US	umbrella sampling
vdW	van der Waals
VMD	visual molecular dynamics
WHAM	weighted histogram analysis method
Yip	Ypt-interacting protein
Ypt	yeast protein two

1 Introduction

1.1 Clinical relevance of Rab5

Rab5 (Ras gene from rat brain) belongs to the Ras (Rat sarcoma) superfamily of small GTPases (guanosine triphosphatases) and is a key regulator of early endosomal membrane trafficking, fusion and sorting.[1–3] As a GTPase, Rab5 cycles between an inactive, guanosine diphosphate (GDP)-loaded and an active, guanosine triphosphate (GTP)-loaded state. The activation cycle is controlled by guanine nucleotide exchange factors (GEFs) which replace GDP by GTP and thus activate Rab5, and GTPase activating proteins (GAPs) which accelerate intrinsic GTP hydrolysis and restore the inactive Rab5 state. Membrane-bound active Rab5 recruits a large spectrum of effector proteins which open up an exceptionally complex signaling machinery downstream of Rab5.[4] Apart from its role in early endosome transport and sorting processes that decide the fate of different cargo, Rab5 activity is connected to signaling related to the endosome-nuclear pathway,[5] receptor tyrosine kinases (RTKs),[6] cell cycle and inflammation.[7] Besides its function in the endo-lysosomal system, Rab5 regulation of phagosome maturation and transport has important implications for pathogen degradation in the immune system.[8] The Rab5 broad involvement in cellular processes makes it prone for bacterial as well as viral attacks, and dysfunctions are often associated with diseases and cancer. The following listing of Rab5-related disorders is not meant to be exhaustive, but gives a rough insight into the clinical relevance of Rab5 (Figure 1.1).

Rab5 activation by GDP exchange to GTP is hindered by *Listeria monocytogenes* bacteria in order to prevent phagosome fusion with lysosomes to escape the degradation pathway via a still unknown mechanism.[9] During *Salmonella* infection, the bacterial virulence factor SopB recruits Rab5 to *Salmonella* containing vacuoles (SCV), which thus promotes the formation of PI(3)P (phosphatidylinositol 3-phosphate) required for SCV maturation.[10] The entry of *Yersinia pseudotuberculosis* into

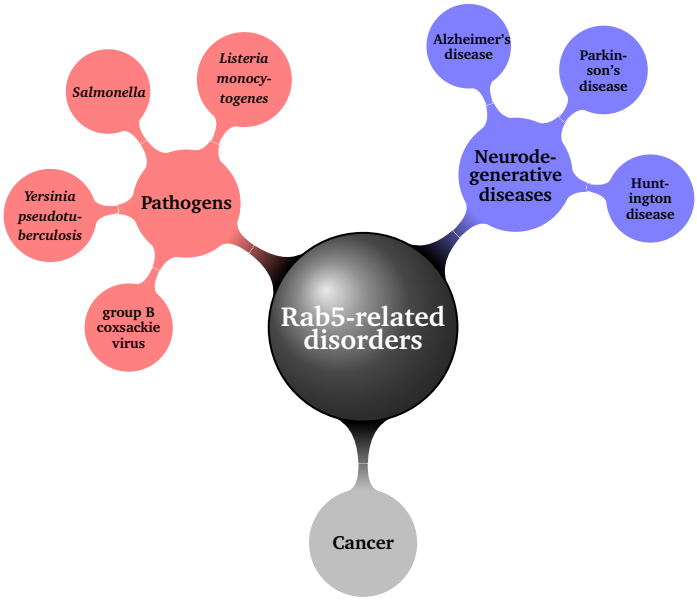


Figure 1.1: Examples for Rab5-related disorders caused by pathogens (e.g. bacteria, viruses), or mutations.

the host cell from PI(4,5)P_2-rich (phosphatidylinositol 4,5-bisphosphate) prevacuoles is only possible after PI(4,5)P_2 dephosphorylation. Recruitment of Rab5-positive vesicles and the effector phosphatases OCRL (oculocerebrorenal syndrome of Lowe protein) and INPP5B (inositol polyphosphate-5-phosphatase B) leads to membrane scission of the prevacuole and release of *Y. pseudotuberculosis*.[11] Infection with the group B coxsackie virus causes an internalization of tight junctions, i.e. protein complexes that regulate the substance flow through polarized epithelial cells. Virus escape into the cytoplasm via the internalized tight junctions depends on correct Rab5 and Rab34 activity.[12]

Rab cellular functions are crucial for neurotransmitter release in neurons, thus Rab5 dysfunctions are associated with several neurodegenerative diseases like Alzheimer's and Parkinson's diseases. Down syndrome patients face an increased risk of developing early-onset Alzheimer's Disease (AD) due to an extra copy of the amyloid precursor protein (APP) gene located on chromosome 21.[13] Typical AD symptoms include the accumulation of β-amyloid plaques and phosphorylated Tau-containing neurofibrillary tangles which leads to synaptic failure, neurotransmitter loss, neuroinflammation and neuronal death.[14,15] The earliest neuronal hallmarks of AD are an increased fusion and enlargement of Rab5-positive early endosomes as well as an upregulation of Rab5 and its effectors that accelerate neuronal endocytosis.[16,17] The Rab5 regulatory protein APPL1 (adaptor protein containing pleckstrin homology domain, phosphotyrosine binding (PTB) domain, and leucine zipper motif) mediates signaling related to several growth factors, e.g. epidermal growth factor (EGF),[18] nerve growth factor (NGF),[19] nuclear factor κB (NF-κB),[20] phosphoinositides[21] and insulin,[22] which is described for early endosome dysfunction in AD.[13] APP is processed by β-secretase to form the β-cleavage C-terminal fragment (βCTF) in the early endosome. In Down syndrome and AD patients, the APP-βCTF concentration is increased and recruits APPL1 to endosomes where it stabilizes active Rab5. Thus, endosome fusion and endocytosis are accelerated which causes the early endosomes to be significantly enlarged. In addition, retrograde axonal transport toward the cell body, which normally takes place and propagates growth and differentiation signals to the nucleus, is inhibited. This causes a premature degradation of neurotrophic signaling before reaching the nucleus and eventually leads to neuronal atrophy in AD as well as in Down syndrome.[13,15]

In Parkinson's disease (PD), misfolded α-synuclein accumulates in Lewy bodies in degenerated dopamine neurons. Similar to AD, a defective axonal transport is suspected to cause PD, in which increased levels of α-synuclein inhibit BDNF (brain-derived neurotrophic factor)-induced trophic signaling and distort the retrograde axonal transport. In a PD mouse model, α-synuclein increases the levels of active Rab5 and Rab7 followed by endosomal dysfunction.[23]

In Huntington disease (HD), the protein huntingtin is mutated which results in the accumulation of expanded polyglutamine stretches leading to the typical hallmarks including mitochondrial dysfunction, impaired synaptic transport and cell signaling.[24] Rab5 overexpression was shown to attenuate polyglutamine toxicity, probably due to Rab5 function in early autophagosome formation.[25]

Rab5 function is also linked to cancer cell motility and tumor dissemination via its role in actin remodeling and integrin traffic. Integrins are cell surface heterodimeric receptors which anchor cells to proteins on the extracellular matrix (ECM). Endocytosis and internalization of integrins involves an aggregation in Rab5 positive early endosomes. At this point, Rab5 regulates integrin traffic and in conjunction with caspase-8, a key protein in apoptosis, it promotes tumor motility and metastasis.[26] In addition, Rab5 recruits the Rac1 GEF Tiam1 to early endosomes, being a prerequisite for local actin remodeling which is required for cancer cell migration.[27,28]

The need for Rab5-specific therapies is easily traceable due to its great involvement in various disorders highlighted above. However, exploiting Rab5 as a drug target is not straightforward. The term 'druggability' describes the likelihood of a protein target to be functionally modified by a small molecule drug.[29,30] The closely related Ras protein is an example for a small GTPase which is frequently mutated in human cancers and thus, a popular target in pharmaceutical drug design. With the aim of developing effective Ras inhibitors, researchers focused on five key strategies during the last decades.[31] Those approaches include 1) the direct inhibition of Ras, 2) the inhibition of Ras association with membranes, 3) interference with Ras downstream effector signaling, 4) harnessing

genes which interact with mutated Ras genes to completely abolish function, and 5) the discovery of drugs to exploit Ras-induced changes in cell metabolism. With a profound understanding of Rab5 structure and interactions, these strategies are also applicable to the development of Rab5 inhibitors. However, it remains debatable whether small modifiers are able to interact with Rab5 in a way that preserves physiological function while inhibiting harmful effects. This is indeed a great challenge due to the prominent role of Rab5 in endocytosis which affects almost all parts of cellular trafficking and signaling and thus, decides about cell survival and death.

1.2 Molecular dynamics: State-of-the-art and its applications in medicinal chemistry

The question how structure, dynamics and interactions of molecules affect their biological function is paramount for biological and medicinal research. Kinetics and thermodynamics of macromolecules in their respective cellular environment are examined by experimentalists in the fields of molecular biology, biochemistry, biophysics, medicinal chemistry, organic chemistry and many more. Computational modeling and simulations represent important tools to assist these investigations by the interpretation of experimental data, the design of future experimental setups and the prediction of molecular properties.[32]

Molecular dynamics (MD) is a prominent simulation technique that treats atoms as spherical particles interacting through potential functions. The motion of the particles follows Newton's laws. The integration of the equations of motion allows a prediction of a system's time evolution and thus, its kinetic properties. Alder and Wainwright performed first MD simulations on hard spheres in the late 1950's.[33,34] The first MD study of a biologically relevant system was presented in 1977 by McCammon and coworkers who investigated the bovine pancreatic trypsin inhibitor (BPTI).[35] In the early days of MD, biological molecule models were rather simple due to the fact that hardly any high-resolution experimental data were available and that mathematical algorithms, implementation and the computing power were limited. Partly, this is an issue even today, as experimental data often suffer from a too low resolution to be able to deduce molecular details from them. Thus, MD simulations aim at the prediction of molecular properties that cannot, or only insufficiently, be observed in experiment. On the other hand, it is a striking paradox that due to the lack of sufficient experimental data, the outcome of MD simulations cannot easily be validated.[36]

Today's force fields for biomolecular simulations include physical model-based potentials and empirical parameters which were fitted to experimental or quantum chemical data. Classical force fields treat the atomic nuclei explicitly, whereas electronic charges are implicitly taken into account by fixed partial charges. More sophisticated force fields are the polarizable force fields which consider charge densities that are able to change their distribution in response to the charges in close proximity.[37] *Ab initio* quantum mechanics (QM) calculations are widely used to improve force fields. One recent example is the training of force field potentials related to non-covalent interactions on small molecules using Møller-Plesset perturbation (MP2) theory with neural networks.[38] The improvement of force field parameters is a constant interplay between MD simulations and experiments. Nuclear magnetic resonance (NMR) studies of small peptides provide benchmark data that led to an optimization of the backbone potential in protein force fields and significant improvements in accuracy.[32] Small peptides of 10 to 40 residues length are good benchmark systems as they, on the one hand, exhibit relevant structural transitions and, on the other hand, are small enough for MD simulations to converge, which means that all peptide conformations have been sampled. Convergence of MD simulations is a challenge in computational biology that is referred to as the 'sampling problem'. The aim of biomolecular MD simulations is to identify relevant system conformations and their relative populations. Deciding which conformational states are relevant is only possible if all conformations are sampled and if sampling is comprehensive enough to allow a profound statistical analysis. It is non-trivial to know how much sampling is sufficient since motions may occur on time

scales that are beyond the simulated time scales.[39] Increasing the simulation time is a straightforward way to improve sampling, at least if the system has not been caught in a local energy minimum structure. When increasing the simulation time, computational costs come into play. In 2011, a one-microsecond simulation of a small biomolecular system of approximately 25,000 atoms took months of computation on 24 processors.[40] Dependent on the system size, MD simulations on a microsecond scale with the simultaneous use of thousands of processors take several months to complete even today.[32] Since 2002, parallelization plays an important role as it allows to use multicore processors and high-performance computing (HPC) clusters. More and more, graphics processing units (GPUs) are used to meet the growing computational demands of molecular modeling due to their excellent parallelization properties, large memory capacities and peak performances over 500 billion floating point operations per second.[41,42] In most cases, highly data-parallel calculations are 10 to 20 times faster on GPU compared to CPU (central processing unit) cores.

Another strategy is pursued by the research group of D.E. Shaw. They use specialized hardware with processors exclusively built for extremely computational demanding MD simulations.[43] The supercomputer called 'Anton' perfectly matches the requirements of MD, i.e. calculation of the computationally demanding force field interactions in certain integrated circuits, the consideration of periodic boundary conditions in a specialized communication network and an architecture-specific parallelization algorithm. In 2010, a millisecond-length full-atomistic MD simulation of a globular protein was performed on 'Anton' for the very first time.[43] Today, even larger systems with more than one million atoms are subjected to several microsecond simulations and completed within a day using the successor model 'Anton 2'.[44]

An interesting approach to acquire resources for costly computations is the folding@home project that uses idle resources of personal computers from volunteers around the globe to perform MD simulations.[45] Taken together the distributed simulation snippets generate a multiple milliseconds long MD simulation which is processed by a Markov state model to construct a transition network of all sampled conformational states.[46]

As mentioned beforehand, MD simulations ought to complement experimental studies and are intended to predict molecule properties which are not or hardly accessible by experiment. With the recent improvements in hardware, software and algorithms, MD became increasingly popular especially in the field of drug discovery and design. From its classical applications, e.g. the prediction of protein folding or protein-protein complex interactions, MD today assists elucidating the binding recognition process as well as binding-induced conformational changes important in pharmaceutical mechanisms.[47] Druggable sites, i.e. binding sites for drug-like molecules to evoke a biological effect on the target, on protein surfaces have been identified using classical MD simulations.[48] Later on, druggability and the maximal binding affinities of allosteric proteins were predicted based on MD and free energy calculations.[49] Recently, a cryptic pocket was identified on the highly dynamic dopamine D3 receptor using MD simulations combined with Markov state models.[50] Furthermore, MD is used for virtual screening, i.e. to identify hit molecules and investigate their binding modes within a target cavity. In 2018, an extensive study docked a large number of small molecules on an allosteric binding site of kinesin Eg5 and ranked the molecules according to their docking scores.[51] Subsequently, the best ranked structures were subjected to MD simulations to analyse conformational rearrangements and complex stability. Protein-protein interactions are also viewed as druggable targets as those interactions induce signaling that may be related to diseases. However, protein-protein interactions are difficult to modulate by drugs due to their complexity and their involvement in interaction networks which make it hard to develop specific therapeutics.[47] MD simulations identified hydrophobic hot-spot residues on protein surfaces by ligand-mapping and thus, represent a promising tool for the design of peptidic modulators of protein-protein interactions.[52]

Multiscale approaches are common to use in drug design as they allow a combination of atomic scale chemistry, i.e. QM calculations, with molecular level biological function, e.g. MD or Brownian dynamics (BD). The relevance of multiscale methods received confirmation by awarding the 2013 Nobel Prize in Chemistry to Martin Karplus, Michael Levitt and Arieh Warshel for the development of

multiscale models for complex chemical systems. The multiscale method SEEKR (simulation enabled estimation of kinetic rates) combines MD and rigid-body BD to determine association and dissociation rates in drug binding and protein-ligand complex binding free energies.[53,54] Hybrid QM/MM treats a small region, e.g. an enzyme's active site, by quantum mechanics and the surrounding structure, e.g. the rest of the protein, solvent or membrane, with empirical molecular mechanics. Applications of QM/MM calculations are the design of covalent inhibitors based on studies revealing specificity, reactivity and interactions with transition states.[55] Furthermore, QM/MM studies have been able to reveal details of the membrane-bound cytochrome P450 enzyme reaction mechanism involved in drug metabolism of the pharmaceutical compounds ibuprofen, diclofenac and warfarin.[56,57]

Apart from docking and drug-receptor binding studies, the investigation of phenomena involved in drug delivery, e.g. prediction of formulation properties, represent an important application of molecular simulations.[58] Early studies in this field focused on the orientation of a drug in a biomembrane[59] or on molecular surface properties of single molecules to predict drug absorption.[60] A more recent study modeled glyceride lipid formulations highlighting the molecular interactions between drugs and excipients.[61]

From all these examples mentioned above (Figure 1.2), it becomes clear that molecular simulations have a great potential as virtual screening tools to identify drug candidates and drug carriers. Today, findings from MD are able to assist the development of pharmacophores and docking constraints during early stages of drug candidate screening, the identification of hidden binding modes and allosteric sites, and the optimization of a lead compound molecular structure.

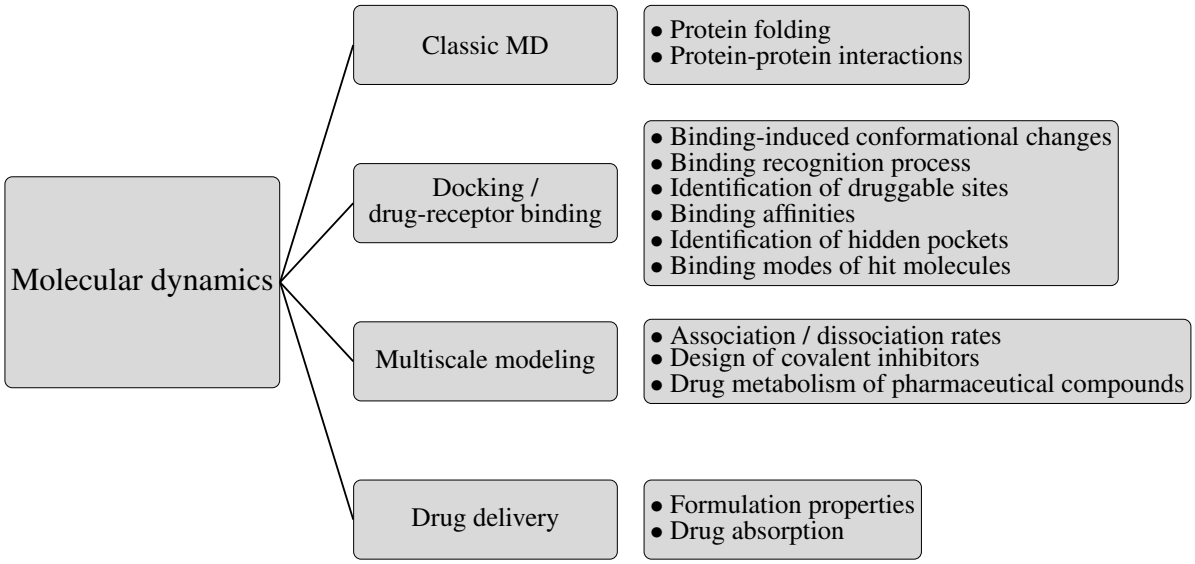


Figure 1.2: Representation of MD-based research areas with a focus on its applications in pharmaceutical drug design.

1.3 Motivation of the study and thesis outline

The structure of a protein determines its function. In order to understand how Rab5 works and to predict how it behaves in its distinct cellular context, it is important to know the three-dimensional (3D) structure of the protein. This knowledge provides the basis for the exploitation of Rab5 as a drug target and a successful treatment of Rab5-related disorders. Widely used structure determining experimental techniques are protein X-ray crystallography, nuclear magnetic resonance (NMR) spectroscopy and cryo-electron microscopy (cryo-EM).[62] Each of them has its specific advantages and limitations. X-ray crystallography depends on the formation of protein crystals which have particular,

rather un-physiological properties. Most of the protein structures in protein data bank archives were determined using X-ray crystallography. The crystallization of membrane proteins is challenging since they are hardly soluble in aqueous solutions which makes the application of customary protein crystallization methods less suitable. The interested reader is recommended to consult an overview on protein crystallization in general[63] and crystallization of membrane proteins in particular.[64] Protein structures in solution are determined by NMR, however, this technique is only applicable to small proteins. Both, X-ray diffraction and NMR rely on highly purified protein samples. Larger protein assembly structures (>200 kDa) are able to be resolved by cryo-EM, which is in turn limited to rigid and homogeneous complexes and resolves individual protein structures only in particular cases.[65] The three 'standard' methods mentioned above are complemented by small-angle X-ray scattering (SAXS) and cross-linking/mass spectrometry (CLMS). Being a low-resolution technique (12 Å-20 Å), SAXS is predominately applied for protein quaternary structure analysis and not suited for tertiary structure determination or fold recognition.[66] CLMS uses molecular probes which introduce covalent links between nearby amino acid residues, that are read by mass spectrometry. This method provides 3D structural information on single proteins and insight into protein-protein interactions in protein complexes. For more information on CLMS in protein structure determination, the reader is referred to the articles of Schneider and Rappsilber.[62,67] Apart from CLMS, all structure determining techniques suffer from the drawback that they study proteins under artificial conditions deviating significantly from the proteins' physiological cellular environments with important implications for structure-function relationships.

Due to experimental inaccuracies, technical challenges, high costs and, of course, time restrictions, the experimental determination of protein structures is not always possible. Computational methods exist that predict the structure of an unknown target protein on the basis of known template structures with a similar sequence (comparative or homology modeling).[68] If no related protein structures are available, *ab initio* modeling[69] is based on the application of physical and biological principles to build an unknown protein structure.

In order to obtain insights into membrane protein structure and dynamics, computer simulations play an increasingly important role. First MD simulations of the intrinsic membrane proteins gramicidin A channel[70] and bacteriorhodopsin[71] in an explicit lipid bilayer were performed in the 1990's. In contrast to peripheral membrane proteins, which are anchored to the membrane via specific hydrocarbon modifications, intrinsic proteins are in large part embedded in the membrane thereby significantly influencing the bilayer structure.

Rab5 is a peripheral membrane protein which exerts its role in endocytosis and membrane trafficking by recruiting effector proteins to the bilayer. Performance of its enzymatic function and binding to regulatory and effector proteins is realized by the catalytically active G domain, which is structurally conserved among different Rab proteins. First Rab5 structure entries in the protein data bank date back to the year 2004.[72,73] However, until today, these crystal structures include only the protein G domain. Rab5 is anchored to the membrane via two lipid chain modifications which are located at a long and flexible C-terminal coil structure, termed the hypervariable region (HVR). The highly flexible N- and C-terminal regions are missing in all available experimental structures. The missing parts comprise 14 residues at the protein N-terminus and 30 residues at the C-terminus. Rab5 experimental studies rely on the availability of lipid-modified, full-length protein, whose biochemical synthesis is very difficult and time-consuming. The interested reader is referred to a review about synthesis strategies of lipidated Ras and Rab proteins.[74] One example for the challenges related to small GTPase synthesis is the covalent attachment of the prenyl chains. These post-translational modifications are characteristic for eukaryotic organisms and thus, not producible using common prokaryotic cultures, e.g. *E. coli*. Cultivation of eukaryotic expression systems is more demanding in terms of time, costs, technical equipment and knowledge. Furthermore, prenylation must not take place at any arbitrary position but is highly site-specific. Due to the challenges related to lipidated GTPase synthesis, the preparation and performance of experimental investigations of full-length Rab5 are difficult, costly and time expensive. Thus, computational MD studies of Rab5, like the simulations

presented here, represent a promising alternative in order to elucidate the protein structure and dynamics at atomistic resolution. From a biological point of view, a full-length model of Rab5 is of utmost importance. The HVR is expected to not only play a passive role in providing a holder for the post-translationally attached lipid chains for membrane anchoring, but is itself involved in protein-membrane interactions. Those interactions are supposed to contribute to nucleotide state-dependent protein dynamics at the bilayer surface, to guide correct membrane targeting and to pave the way for G domain interactions with the membrane, regulatory and effector proteins. In doing so, the HVR significantly influences Rab5 function. Knowing the HVR structure would improve the understanding of Rab5 membrane association and protein-protein binding assisted by HVR interactions. This knowledge provides the basis for the development of Rab5-specific therapeutics and the exploitation of Rab5 as a drug target.

Due to the great flexibility of the HVR, the need for a full-length Rab5 structure could, to date, not be met by X-ray diffraction experiments. Thus, in the present study, a full-length Rab5 model will be derived by computational modeling and subjected to full-atomistic MD simulations. These simulations will elucidate Rab5 dynamics at model bilayers of different compositions and investigate Rab5 complex formation with other proteins. The goals and achievements of this thesis are summarized in the following list.

1. Modeling of human full-length Rab5(GDP) and Rab5(GTP) based on the crystal structures of the Rab5 G domains deposited in the protein data bank[72,73]
2. Investigation of the Rab5 dynamics and orientations at different model bilayers by MD simulations
    a) Determination and validation of force field parameters for the Rab5 post-translationally attached lipid modification
    b) Confirmation of the great flexibility of a truncated membrane-bound Rab5 HVR (residues 206 - 215)
    c) Determination of the potential of mean force (PMF) of the truncated Rab5 HVR bilayer-to-water transfer
    d) Detection of signaling lipid clustering in close proximity to the Rab5 HVR$^{206-215}$ using coarse-grained MD
    e) Discovery of activation state-dependent full-length Rab5 orientations at the membrane surface in a charged bilayer model mimicking the early endosome membrane
    f) Identification of these orientations as interaction sites for nucleotide-specific targeting
    g) Comparison of mono-prenylated full-length Rab5 orientations with the doubly prenylated Rab5 protein
3. Probing the formation of membrane-bound homotypic Rab5 dimers by full-atomistic MD
    a) Identification of Rab5 dimerization interfaces
    b) Discovery of minor activation state-dependent structural differences in the Rab5 dimers
4. Elucidation of human Rab5(GDP):RabGDI complex formation by full-atomistic MD
    a) Identification of Rab5(GDP):RabGDI interaction sites
    b) Determination of the PMF of Rab5:GDI complexation and Rab5 bilayer-to-water transfer
    c) Proposal of a GDI-mediated Rab5 membrane extraction mechanism

The present thesis is organized as follows. The chapter 'Biological background and modeling' gives an overview about the biological systems which were investigated in the thesis and the relevant computational approaches. The biological section characterizes the Rab5 protein within the large family of small GTPases, outlines its important cellular role and illustrates the composition and structure of

biological membranes. The computational part describes the physical basics of unbiased and biased MD. In 'Methods and technical details', the general workflow of the performed MD simulations is represented and the simulation's technical parameters using the software NAMD are explained. The main chapters 'Protein-lipid interactions of membrane-bound Rab5', 'The dimerization of membrane-associated Rab5 proteins', and 'Rab5(GDP) membrane extraction by RabGDI' present the results of the MD simulations concerning the three investigated systems, namely membrane-bound Rab5, the Rab5 dimeric complex as well as the Rab5:RabGDI complex. The last two chapters summarize the results, draw conclusions and contain an outlook regarding further simulations and the application of the findings for therapeutic purposes.

2 Biological background and modeling

The present thesis investigates the peripheral membrane protein Rab5 by molecular dynamics simulations and therefore combines aspects of computer-based modeling and simulation as well as topics from cell biology and theoretical chemistry. Thus, this chapter shall introduce basic concepts of the respective field to both computer scientists as well as biologists. First, the investigated protein system is categorized within the context of small GTPases and its cellular function is described. The second part of this chapter covers biological membranes and their composition. Finally, the mathematical principles of molecular dynamics simulations are explained and the techniques used in this thesis are presented.

2.1 Peripheral Rab5 protein and introduction into membranes

2.1.1 Classification of Rab5 within the superfamily of small GTPases

The genetic code of an organism is translated into the expression and modification of proteins during development or on the basis of external stimuli. The protein primary structure corresponds to the amino acid sequence. Intramolecular interactions between the amino acids backbone atoms, primarily hydrogen bonding, form a well-defined protein secondary structure. Helices, e.g. alpha (α)-helices, and beta (β)-sheets represent the most prominent protein secondary structure elements. Interactions between individual amino acid side chains define the overall protein 3D structure, the so called tertiary structure. Some proteins consist of multiple subunits, whose arrangement establishes the protein quaternary structure.

The focus of the present work is on the small GTPase Rab5. The Ras superfamily of membrane-associated small GTPases comprises a large number of monomeric GTP-binding (G) proteins which share a structurally conserved catalytic domain, the so-called G domain. According to their sequence homology, small GTPases are classified into six prominent functional subgroups (Figure 2.1).[75] The Ras superfamily is named after the 21 kDa Ras oncogene which was first described in 1979.[76] Ras proteins regulate cellular processes such as cell proliferation and differentiation, apoptosis and signal transduction.[77] For more detailed information about Ras proteins readers may refer to the review by Reuther *et al.*[78] The largest branch of the Ras superfamily is the Rab protein family. Rab proteins are responsible for intracellular vesicle trafficking and transport between endocytic and secretory pathways. Their specific localization to distinct membrane compartments establishes membrane identity and ensures correct targeting of vesicles.[3,79] The function of Rab proteins is more thoroughly described in section 2.1.2 of this thesis. Members of the Rho (Ras homologous) family are involved in the regulation of cytoskeletal organization, cell migration, cell cycle progression and cell polarity.[80] Arf (ADP ribosylation factor) proteins regulate early processes in membrane traffic like vesicle budding through the recruitment of coating complexes to the membrane as well as phospholipid metabolism.[81] The Ran (Ras-related nuclear) GTPase is a key regulator in the transport between cytoplasmic and nuclear compartments.[82] The Rag subfamily has been identified only recently and differs from the other small GTPases by a particular long C-terminal region, the ability to form heterodimers and the absence of a membrane-targeting motif. In the lysosome Rag proteins activate mTOR (mechanistic target of rapamycin) complex 1 upon amino acid stimulation, which senses nutrient availability and couples it to cell growth.[83,84]

Apart from the G domain which is conserved among all small GTPases, these proteins exhibit flexible

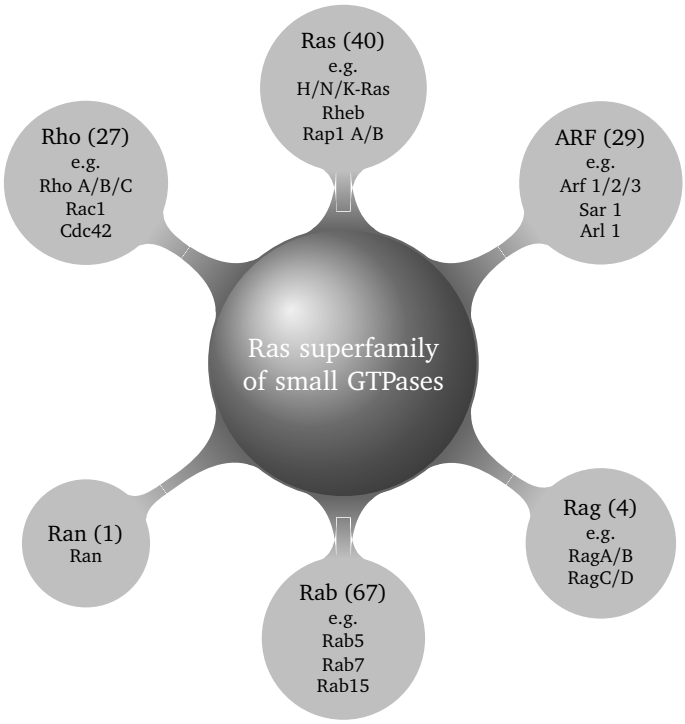


Figure 2.1: Definition of subgroups of the Ras superfamily of small GTPases with currently identified number of members in brackets (modified from[75]).

N- and C-terminal regions which are specific for the individual proteins. In these terminal parts, members from the Ras, Rab, Rho and Arf subfamilies contain sequence motifs for the covalent attachment of hydrophobic chains, so called post-translational modifications, which enable membrane anchorage. Membrane attachment is realized via various types of hydrophobic modifications, such as isoprenylation, palmitoylation, or myristoylation, or via an accumulation of positively charged amino acids.[85] The amino acid sequence with highlighted functional domains is exemplarily shown for the Rab5 small GTPase in Figure 2.2.

The G domain of small GTPases consists of several conserved secondary structure elements, namely a six-stranded β-sheet and five α-helices. It binds the nucleotides GDP and GTP with high affinity via specific residues in the guanine binding loop and the phosphate-binding loop (P-loop) (Figure 2.2). Upon exchange of the bound nucleotide, significant conformational changes can be observed in the two so-called 'switch regions', characterizing either the 'active', GTP-bound state or the 'inactive', GDP-bound state. The first switch region (switch I) contacts the third phosphate group (i.e. the

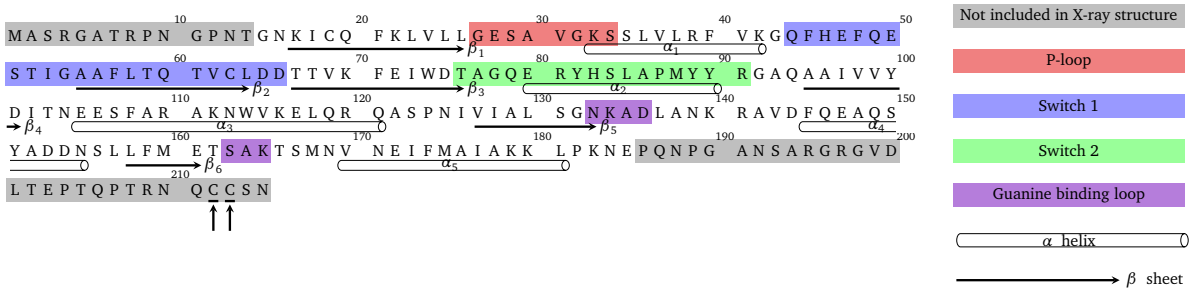


Figure 2.2: Amino acid sequence of the Rab5 protein and important functional regions. From the 215 residues only the catalytic G domain structure from Gly^{15}-Glu^{185} is experimentally resolved by X-ray. The post-translationally attached geranylgeranyl chains are located at Cys^{212} and Cys^{213} (arrows).

γ-phosphate) in GTP as well as an associated magnesium ion, and is involved in GTPase binding to effector proteins, i.e. proteins that selectively bind the GTPase, regulate its activity or exert a specific biological function. The second switch region (switch II) also interacts with the GTP γ-phosphate and with regulatory proteins involved in nucleotide exchange and GTP hydrolysis. When GTP is bound, the switch region conformation is stabilized by interactions with the nucleotide, whereas in the GDP-bound state, the structure of the switch regions is rather disordered and flexible.[86]

2.1.2 The cellular functions of Rab GTPases

The subfamily of Rab GTPases comprises more than 60 members in humans that are regulators of membrane trafficking and are found at the cytoplasmic site of cellular membranes.[87] Intracellular vesicular trafficking is essential for living cells as it includes i) the transport of newly synthesized proteins and lipids via the endoplasmic reticulum (ER) and Golgi network to the plasma membrane (exocytic pathway), and ii) the uptake and transport of macromolecules from the plasma membrane via endosomes to degradation in the lysosomes (endocytic pathway).[90] Rab proteins are key

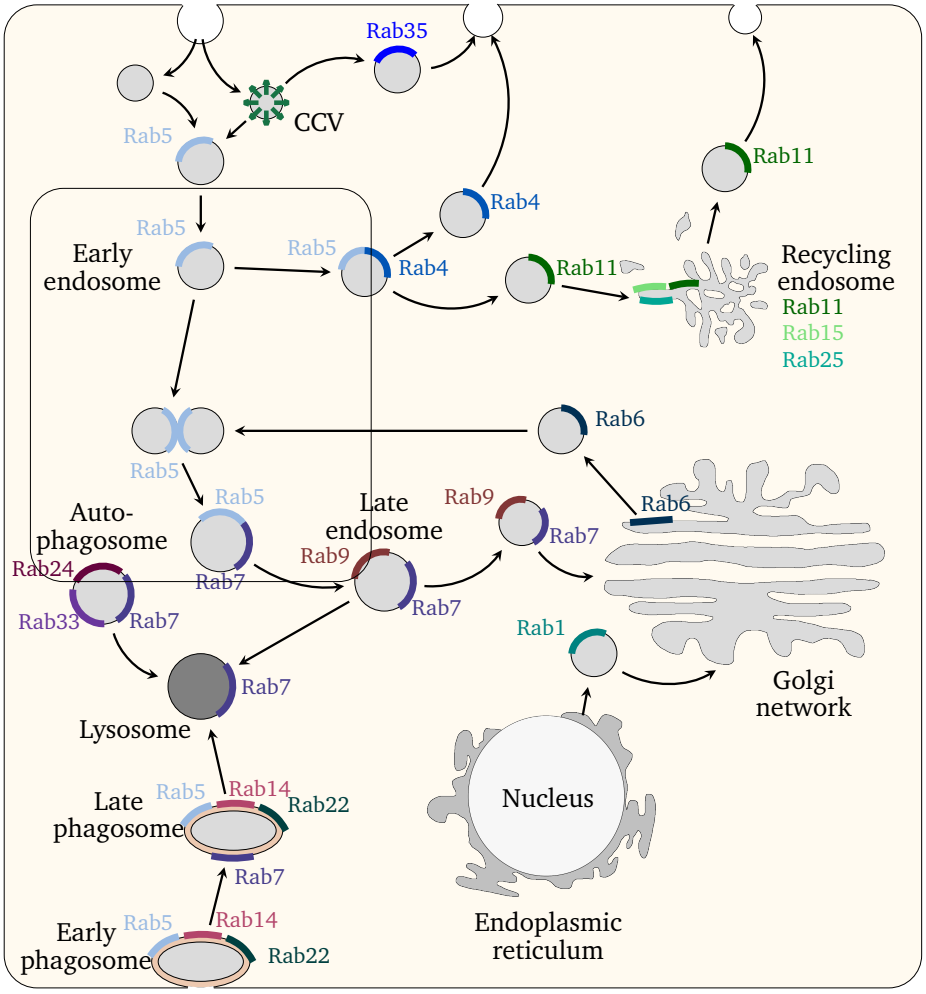


Figure 2.3: Localization and signaling of selected Rab GTPases within a eukaryotic cell. Rab proteins are mainly involved in intracellular vesicular transport and establish membrane identity due to their selective targeting to membrane compartments. They are key regulators of caveolar and clathrin-coated vesicle (CCV)-mediated endocytosis, phagocytosis, endosome maturation, anterograde and retrograde transport between the endoplasmic reticulum (ER) and the Golgi apparatus and recycling of cargo back to the plasma membrane.[88,89]

regulators of these pathways and selectively associated with specific membrane compartments.[3] Different Rab proteins are involved in caveolar and clathrin-coated vesicle (CCV)-mediated endocytosis, phagocytosis and macropinocytosis (i.e. the uptake of solute molecules, antigens and nutrients),[91] in regulated exocytosis and secretion as well as in anterograde and retrograde transport (Figure 2.3). The anterograde transport pathway affects newly synthesized proteins which are translocated and modified in the ER, transported to the Golgi apparatus for further modification and distributed to their target compartments. Proteins that are located and active in the ER are carried back to the ER via the retrograde transport pathway. An excellent overview of the functions of different Rab GTPases is given by Schwartz *et al.*[92]

Based on a comparison of sequences, structures and molecular interaction fields, human Rab proteins could be grouped into six subclusters.[93] Proteins in one subcluster are expected to be very similar in their primary structure and are likely to have similar electrostatic potentials and biological functions. Accordingly, proteins in the 'Rab3' subcluster, containing the isoforms Rab3a-d and Rab27a/b, perform similar functions in regulated exocytosis and the transport of secretory vesicles. The 'Rab5' subcluster consists of Rab proteins associated with the early endosome, e.g. Rab5a-c, Rab4a/b, and Rab14. Members of the 'Rab6' subcluster like Rab6a-c, Rab7a or Rab41 are involved in transport to and from the Golgi network. Similarly, Rab1a/b, Rab9a/b, and Rab11a/b form the 'Rab11' subcluster which is regulating steps in Golgi apparatus transport, e.g. from the Golgi apparatus to recycling endosomes and from the endoplasmic reticulum to the Golgi network. The 'Rab37' subcluster comprises Rab23, Rab26, and Rab37, which are involved in various intracellular transport processes like phagosome-lysosome fusion and regulated exocytosis.[89,92] The isoforms Rab40a-c are clustered in the 'Rab40' subcluster; however, the exact cellular function of Rab40 is still not fully understood.

Rab proteins are peripheral membrane proteins and targeted to specific membrane compartments in order to fulfill their diverse functions in membrane trafficking. Membrane anchoring is realized via post-translational modifications, i.e. geranylgeranylation in the case of Rab GTPases (Figure 2.4).

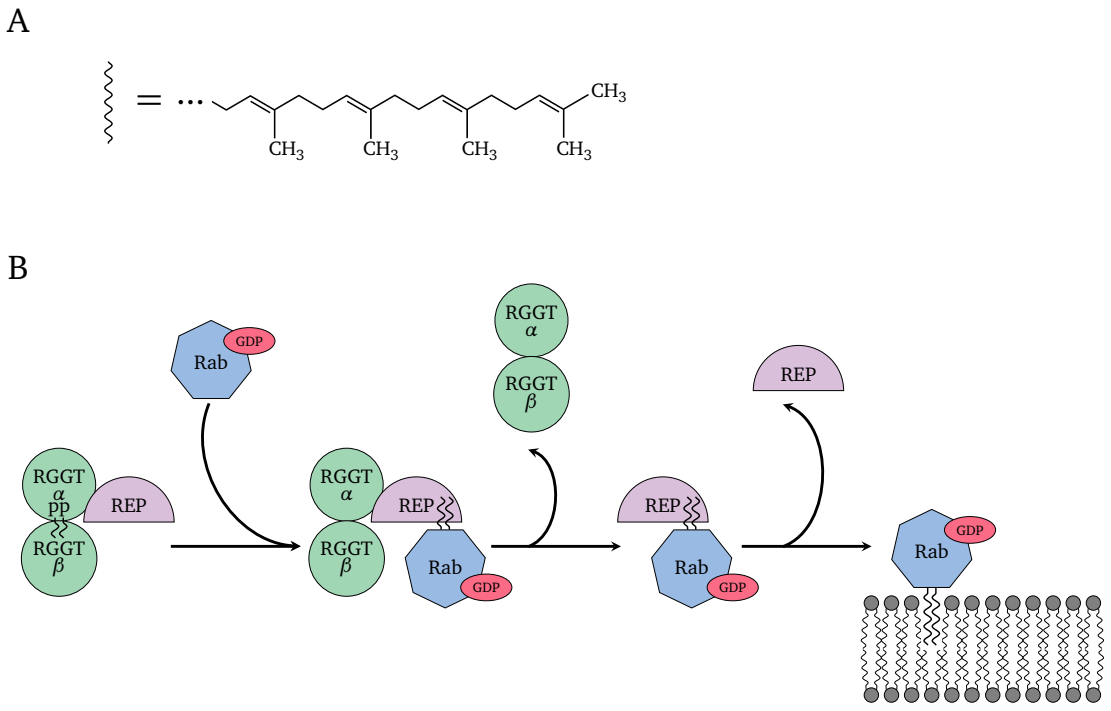


Figure 2.4: A) Structure of a geranylgeranyl chain (20 carbon atoms) that is covalently attached to a cysteine amino acid residue via thioether linkage in Rab proteins. B) A preformed complex of Rab geranylgeranyl transferase (RGGT) and associated Rab escort protein (REP) catalyzes the geranylgeranylation of inactive Rab proteins. Afterwards, lipidated Rab is escorted to the membrane and inserted into the bilayer.

Geranylgeranylation of one or two cysteine residues at the C-terminus of the Rab protein is catalyzed by the Rab geranylgeranyl transferase (RGGT) in concert with the associated Rab escort protein (REP).[94] Baron *et al.* suggested a preferential pathway of Rab geranylgeranylation *in vivo*. First, the heterodimeric RGGT associates with REP and this preformed complex binds unprenylated Rab and transfers the geranylgeranyl moieties from a geranylgeranyl pyrophosphate (GGPP) to the Rab cysteine residues. Pyrophosphate is recycled by the GGPP synthase to form GGPP which is used in a subsequent transfer step.[95] GGPP stabilizes the preformed RGGT:REP complex which dissociates after GGPP transfer to Rab, thereby releasing the geranylgeranyl (GG)-Rab:REP complex, which is delivered to a target membrane. The GG-Rab:REP complex features a high functional similarity to a related complex formed between GG-Rab and RabGDI (GDP dissociation inhibitor). RabGDI releases prenylated inactive Rab proteins from membranes and keeps them soluble in the cytoplasm. The analysis of the molecular interactions between Rab and RabGDI is part of the present thesis and is discussed in detail in chapter 6. An overview about the regulation of GG-Rab proteins is provided in Figure 2.5. RabGDI selectively extracts inactive Rab(GDP) from membranes, transports it between membranes and thus creates a recycling pathway of Rab proteins back to donor membranes. Due to the high affinity of the small GTPases for GDP (i.e. GDP dissociation rates in the order of $10^{-5}\,s^{-1}$ to $10^{-4}\,s^{-1}$)[96,97] , the spontaneous dissociation of GDP occurs within hours or even days.[98] Therefore, the release of GDP and exchange for GTP is realized by regulatory proteins called GEFs which facilitate GDP dissociation. Thereby, GEFs and nucleotides allosterically compete for Rab binding.[99] In

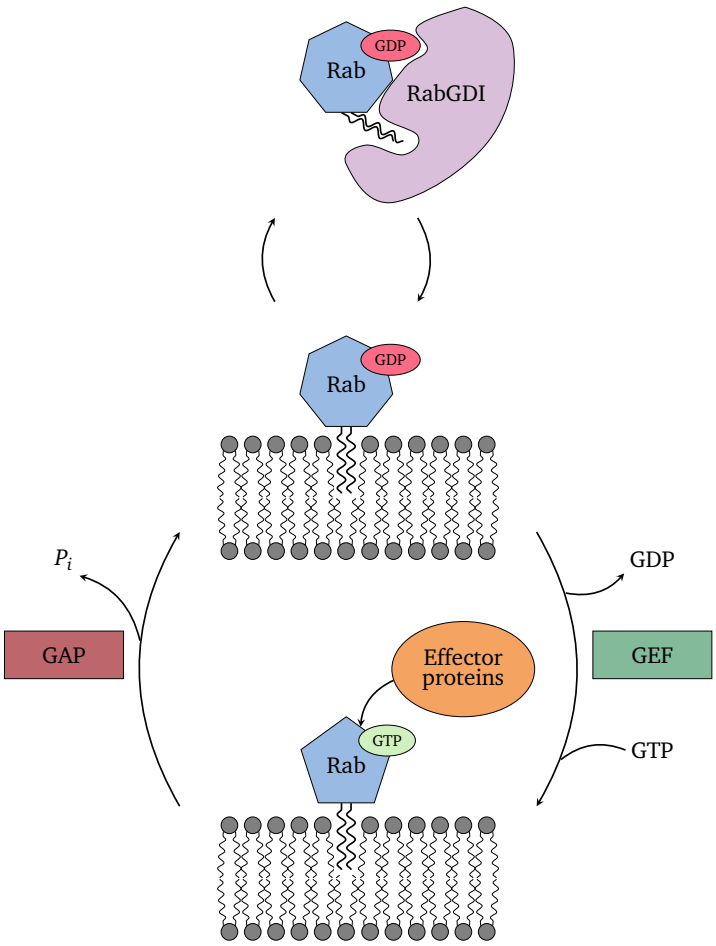


Figure 2.5: Rab proteins shuttle between inactive, GDP-bound and active, GTP-bound states which affect their cellular localization. Transport between different vesicular membranes is realized by RabGDI which exclusively binds Rab(GDP). GEFs catalyze the exchange of GDP to GTP and thereby the activation of Rab proteins. Rab(GTP) is solely membrane-associated and able to recruit a large number of effector proteins. GTP hydrolysis and recovery of Rab(GDP) is achieved by GAPs.[8]

a multistep process, a GEF binds Rab(GDP) to form an intermediate complex, enables dissociation of GDP by stabilizing the nucleotide-free GTPase, and dissociates from the complex as GTP binds to the Rab protein. Unlike Rab(GDP), active Rab(GTP) is exclusively membrane-associated and able to interact with effector proteins. The GTPase intrinsic phosphatase activity is very slow. Thus, in order to rapidly restore the inactive state, GTP hydrolysis is accelerated by GAPs. These proteins provide an arginine residue to stabilize partial negative charges of the transition state and a glutamine residue to activate a water molecule to split off the γ-phosphate by nucleophilic attack.[100]

2.1.3 Structure of biological membranes

Biological membranes comprise a lipid bilayer, membrane proteins and carbohydrate molecules. Lipids are amphiphilic molecules, i.e. they consist of a hydrophilic and a hydrophobic region, which are able to self-associate in order to minimize the interaction of their hydrophobic parts with water molecules. Biological membranes allow the spatial separation of cellular components from the external aqueous environment and, in more complex organisms, the compartmentalization of functionally different constituents into discrete organelles. This enables cellular processes and chemical reactions to take place in highly specialized, spatially separate compartments with increased efficiency. Since transport of solutes like small molecules and ions between different compartments has to be ensured, biological membranes contain embedded membrane proteins, e.g. channel proteins, which allow a selective passage of charged and bulky molecules through the bilayer. The inclusion of molecules into a cell and their intracellular transport is realized via vesicles budding from the membrane surface.[101] Our conception of a biological membrane developed over the past decades.[102] By the end of the 19$^{\text{th}}$ century, cells were described to have an end layer called fatty oil that allows lipophilic molecules to pass easily, whereas the transport of ions and hydrophilic molecules was restricted. In 1925, Gorter and Grendel investigated erythrocyte membranes and found that lipids are arranged in a bilayer.[103] Later, the existence of proteins coating the lipid layer was detected, leading to a membrane model described by Danielli and Davson in 1935.[104] With the development of NMR methods, it was possible to observe the motion of lipid molecules within the bilayer. Consequently, in 1975, the membrane was rather seen as a dynamic fluid with lipids asymmetrically distributed among the two leaflets.[105] Shortly before, in 1972, Singer and Nicolson investigated membrane proteins and discovered that these proteins could be distinguished as integral and peripheral membrane proteins.[106] They postulated their famous fluid mosaic model of membranes which is widely accepted to date. Membrane proteins are embedded into the fluid lipid bilayer in order to minimize the water contacts of their hydrophobic surfaces. Lateral diffusion and rotation of the lipids and proteins is random, but restricted by the cytoskeleton or high local concentrations. The idea that lipid molecules may not be randomly distributed within the bilayer evolved only in the 1990s. The 'lipid-raft' theory deals with large (up to hundreds of nm in diameter) cholesterol- and sphingolipid-rich lipid domains moving within the bilayer. These rafts serve as platforms for binding proteins and are involved in cell signaling and membrane trafficking.[107]

Fahy *et al.* classified lipids into eight different categories based on their hydrophobic and hydrophilic substructures: fatty acyls, glycerolipids, glycerophospholipids, sphingolipids, sterol lipids, prenol lipids, saccharolipids, and polyketides.[108] The structures of representatives from four of these categories are shown in Figure 2.6. The present thesis focuses on glycerolipids, glycerophospholipids, sphingolipids, and sterols as they are the most important components of eukaryotic cell membranes. However, the reader interested in learning more about lipid molecules may refer to the mentioned studies.[101,108] Glycerophospholipids, in short termed phospholipids, consist of a polar head group, a phosphate group and two fatty acid chains of varying length that are connected to the phosphate via a glycerol backbone. The polar head group of glycerophospholipids is formed by a choline (phosphatidylcholine, PC), which is most frequently found in biological membranes, a serine (phosphatidylserine, PS), an ethanolamine (phosphatidylethanolamine, PE), a glycerol (phosphatidylglycerol, PG), or a myo-inositol (phosphatidylinositol, PI). The fatty acid chains are 14 to 24 carbon atoms in length, normally with one unsaturated and one saturated leg. Glycerophospholipids constitute the

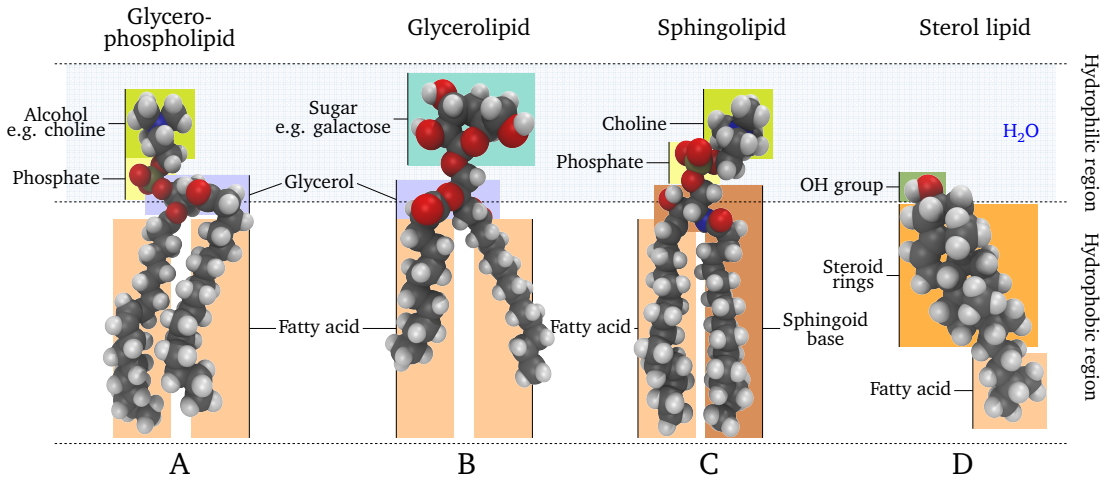


Figure 2.6: Biological membranes are composed of lipid molecules of different types, e.g. A) glycerophospholipids (e.g. with phosphatidylcholine head group), B) glycerolipids (e.g. monogalactosyl diglyceride from plant cells), C) sphingolipids (e.g. sphingomyelin), and D) sterols (e.g. cholesterol).

major class of lipids in eukaryotic cells.

Similar to glycerophospholipids, glycerolipids contain two fatty acid chains and a glycerol backbone, but no phosphate group. In some cases the polar head group consists of a sugar molecule, e.g. galactose, glucose, or N-acetylgalactosamine, forming the subclass of glyceroglycans. Glyceroglycans are mainly found in plant cells and bacteria as part of photosynthetic membranes.

The hydrophobic backbone of sphingolipids is formed by an 18-20 carbon atom long-chain sphingoid base providing an amide-linkage to an acyl chain. In contrast to glycerophospholipids, the 3-position of the sphingoid base is bound to a free hydroxyl group that contributes to the polar properties of the lipid head group. At the primary hydroxyl of the sphingoid base a head group is attached. This head group may be e.g. a hydrogen to form the simplest sphingolipid ceramide, or a phosphocholine to yield sphingomyelin (a phosphosphingolipid).[109] If the sphingosine backbone is connected to a sugar molecule, similar to the glyceroglycans, a glycosphingolipid is formed which is mainly present in animal cells. Glycosphingolipids and glyceroglycans with their bulky sugar head groups are examples of lipids being asymmetrically distributed among both membrane layers and exclusively located in the non-cytosolic monolayer delimiting the cell from the outer environment.[110,111]

Sterol lipids make up another important category of membrane lipids. A sterol's hydrophilic head group consists of a hydroxyl group connected to the A-ring of the four-cyclic structure and a short hydrocarbon chain. In mammalian cell membranes cholesterol is the predominating sterol. It preferentially interacts with sphingolipids to shield the non-polar sterol from water molecules.[101,111]

The plasma membrane consists of more than one hundred different lipid types, whereby the exact lipid composition depends on external factors, like environmental determinants, and internal factors as the type of organism and cell as well as the cell cycle. In addition, the lipid compositions of the inner and outer bilayer leaflets are different, i.e. the layers are asymmetric. An average mammalian plasma membrane is supposed to be dominated by PC, SM and gangliosides in its outer leaflet, whereas PE, PS, and various charged lipids are preferentially located in the inner leaflet. Lipid fatty acids with unsaturated hydrocarbon chains are preferentially found in the inner leaflet. The amount of sterols, especially cholesterol in eukaryotic plasma membranes, varies between 20 % and 50 %.[112]

The present study focuses on the early endosome membrane which is the site of action where Rab5 exerts its biological function. The early endosome membrane composition resembles that of the plasma membrane, but contains approximately 5 % of the phosphatidylinositol 3-phosphate (PI(3)P) signaling lipid as a special characteristic. The enrichment of PI(3)P in early endosomes is related to Rab5 activity and examined in the following section.

2.1.4 Rab5 as a marker and designer of the early endosome

Rab5 is located in the early endosome (EE) membrane. In endocytosis, EEs serve as accumulation compartments, sorting cargo either for degradation in the lysosome, transport to the trans-Golgi network or for recycling. The degradative process, i.e. the transport of cargo to lysosomes, is accompanied by the maturation of EEs to late endosomes (LEs), which eventually fuse with lysosomes. During endosome maturation, endosomal components are removed whereas lysosomal components are added via continuous exchange with the trans-Golgi network. Several changes occur during the endosome maturation process, amongst others: i) Rab5, the marker for EEs, is replaced by Rab7, ii) intraluminal vesicles (ILVs) are formed through inward budding of the endosome membrane, iii) the pH decreases from pH>6 to pH 6.0-4.9 and ion concentrations change, iv) the signaling lipid PI(3)P, which significantly contributes to the identity of EEs, is converted to PI(3,5)P_2 on LEs, v) the ability to fuse with EEs and to form recycling endosomes is absent in LEs.[113] Rab5 is primarily associated with EE trafficking. Its various functions involve i) the budding of caveolar and clathrin-coated vesicles (CCV) from the plasma membrane, ii) the transport of such vesicles to EEs, iii) EE fusion, iv) the regulation of phagocytic transport and maturation of phagosomes, and v) control of macropinocytosis (Figure 2.7).[88]

In most cases, Rab5(GTP) action is linked to the presence of the signaling lipid PI(3)P (Figure 2.8A), since a large number of Rab5 effector proteins contain a so called FYVE zinc finger domain responsible for PI(3)P binding. Therefore, it is not surprising that PI(3)P production is related to Rab5 activity and that PI(3)P is enriched in EEs in close proximity to Rab5 domains.[114] Phosphoinositides are formed by multiple phosphorylation and dephosphorylation steps on their inositol ring at positions 3 ,4, and/or 5 by kinases and phosphatases. The synthesis of PI(3)P in the cell is realized via two Rab5-dependent mechanisms (Figure 2.8B, C). PI(3)P is a key marker for EEs, where it is directly formed from phosphatidylinositol (PtdIns) by phosphorylation by the class III PI 3-kinase Vps34. In a second mechanism, PI(3)P is synthesized via an enzymatic cascade starting with growth factor or cytokine stimulation at the plasma membrane. In a first step, the class I PI 3-kinase PI3Kβ phosphorylates PI(4)P or PI(4,5)P_2 to form PI(3,4)P_2 or PI(3,4,5)P_3, respectively. In order to syn-

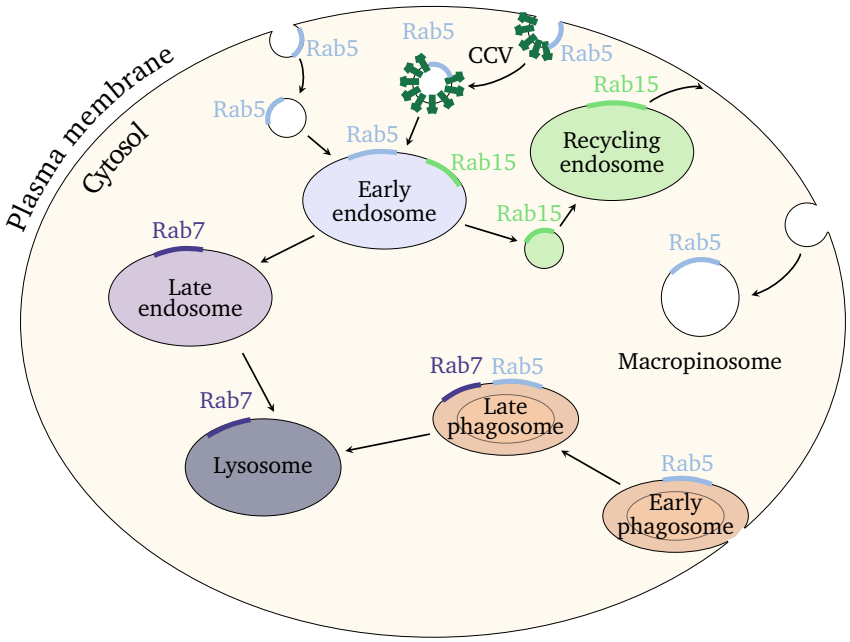


Figure 2.7: Rab5 is the key regulator of early endosomal trafficking. It is involved in endocytosis via CCV and caveolar vesicles, in the fusion of early endosomes and phagosomes as well as in control of macropinocytosis. Rab5 action is linked to other Rab GTPases; e.g. with Rab7 in the process of endosome maturation or Rab15 in recycling of cargo.

thesize PI(3)P, PI(3,4,5)P_3 is successively dephosphorylated by 5-phosphatases (e.g. OCRL1 or type II inositol polyphosphate-5-phosphatase (INPP5B)) and 4-phosphatases.

Since recruitment of effector proteins to the membrane is only possible in the active Rab5 state, the small GTPase has to be activated first. Such activation by GEFs is often realized in a positive feedback loop in which the GEF is recruited together with an effector protein (Figure 2.9). In the case of Rab5, the effector binding to Rab5(GTP) is Rabaptin-5 which is complexed with the GEF Rabex-5. Rabex-5 alone was shown to be unable to bind Rab5. However, complex formation with Rabaptin-5 exposes the Rab5 binding site and enables GEF activity.[115] Subsequently, Rabex-5 activates further Rab5 proteins resulting in a local accumulation of Rab5(GTP).

On the EE, where Rab5 is responsible for endosome fusion and maturation, the accumulation of Rab5(GTP) is linked to an enhanced synthesis of PI(3)P in order to form signaling platforms, which serve as docking sites for effector proteins. Rab5(GTP) recruits Vps34 to locally produce PI(3)P. Many Rab5 effector proteins are so called tethering factors that enable docking and fusion of transport vesicles with acceptor membranes. Examples are the early endosome antigen 1 (EEA1), Rabenosyn-5 and Rabankyrin-5 which have specific FYVE domains that bind PI(3)P lipids. Since these effectors detect the concomitant presence of Rab5(GTP) and PI(3)P their recruitment to membranes is highly specific.[3,88]

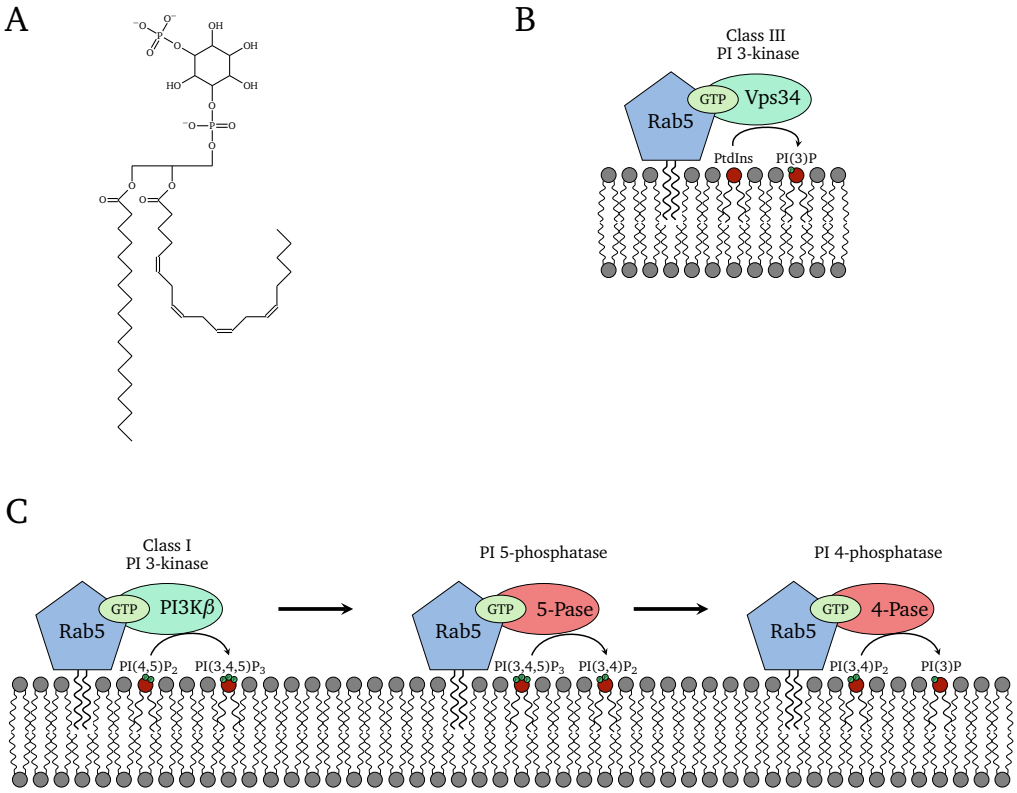


Figure 2.8: A) Structure of the signaling lipid PI(3)P. Rab5 is involved in the production of PI(3)P via two mechanisms. B) On EEs, the PI 3-kinase Vps34, an effector of Rab5, directly phosphorylates phosphatidylinositol (PtdIns) to PI(3)P. C) PI(3)P is formed by interaction of kinases and different phosphatases in a reaction cascade. Growth factor or cytokine stimulation at the plasma membrane results in the phosphorylation of PI(4,5)P_2 to yield PI(3,4,5)P_3, which is subsequently dephosphorylated by PI 5- and PI 4-phosphatases.[114]

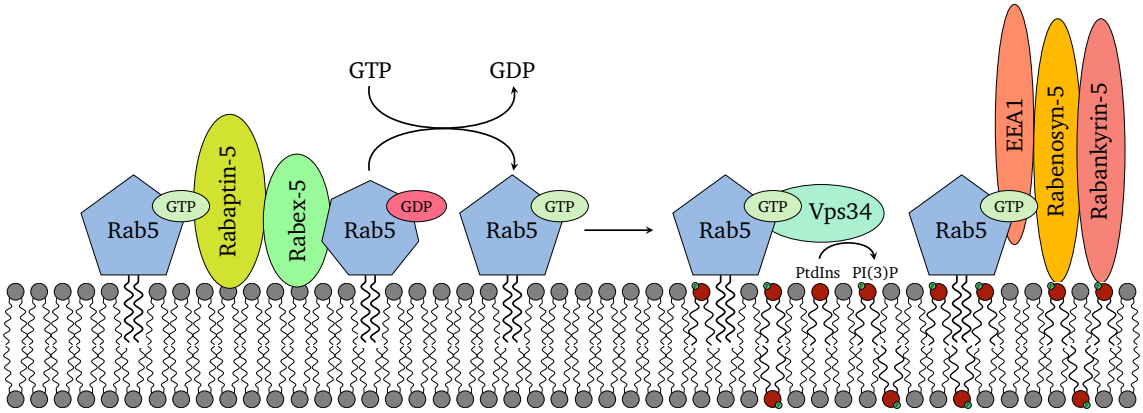


Figure 2.9: Positive feedback loops of Rab5 activation and effector protein recruitment provoke the formation of Rab5-enriched signaling platforms on the EE membrane that are a prerequisite for endosome fusion. Rab5(GTP) recruits the effector protein Rabaptin-5 in complex with Rabex-5, the latter being a GEF that leads to subsequent activation of further Rab5 proteins. Another effector, the phosphatidylinositol 3-kinase (PI(3)K) Vps34 produces the signaling lipid phosphatidylinositol 3-phosphate (PI(3)P) which results in the local enrichment of PI(3)P and Rab5. These domains in the membrane serve as docking platforms for early endosome antigen 1 (EEA1) which forms a complex effector network with Rabenosyn-5 and Rabankyrin-5, which cooperate to regulate membrane tethering and fusion.[3]

2.2 Molecular dynamics simulations

2.2.1 A molecule's potential energy and force fields

In molecular mechanics (MM), a molecule is modeled as a collection of hard spheres, representing the atoms, which are connected by springs, illustrating the bonds. The motion of the nuclei is described by classical Newtonian mechanics. A molecule's potential energy depends on its three-dimensional geometry. The deviation of a structural parameter, e.g. the bond length, from its equilibrium state results in an increase of energy. The mathematical expressions and parameters required to calculate the potential energy are called force fields. The expression 'force field' originates from the fact that the negative first derivative of the potential energy, U, gives the force acting on the i=1,2,...,N particles in the system, with s_i being the position vectors.[116,117]

$$F_i(s_1, s_2, \ldots, s_N) = -\frac{d}{ds_i} U(s_1, s_2, \ldots, s_N) \tag{2.1}$$

Usually, the potential energy of a molecule is described as the sum of the following terms, describing its intra- and intermolecular energy contributions:

$$U = \sum_{\text{bonds}} U_{\text{stretch}} + \sum_{\text{angles}} U_{\text{bend}} + \sum_{\text{dihedrals}} U_{\text{torsion}} + \sum_{\text{pairs}} U_{\text{non-bond}} \tag{2.2}$$

The first three components characterize bonded interactions between atoms, namely the bond length U_{stretch}, the bond angle U_{bend}, and the dihedral angle U_{torsion}. The fourth component describes non-bonded interactions, $U_{\text{non-bond}}$, between all pairs of atoms. With the mathematical terms for the

different energy contributions inserted in equation (2.2), the potential energy is calculated as follows.

$$U = \sum_{\text{bonds}} k_{\text{stretch}} \left(l - l_{\text{eq}}\right)^2 + \sum_{\text{angles}} k_{\text{bend}} \left(\theta - \theta_{\text{eq}}\right)^2 + \sum_{\text{dihedrals}} k_{\text{torsion}} \left[1 + \cos\left(n\omega - \gamma\right)\right] + \left(\sum_{\substack{\text{nonpolar} \\ \text{i,j pairs}}} 4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6} \right] + \sum_{\substack{\text{electrostatic} \\ \text{i,j pairs}}} \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \right)_{\text{non-bond}} \tag{2.3}$$

A schematic representation of the terms of interaction is given in Figure 2.10. The increase in energy associated with a deviation of the actual bonded parameter, l or θ, from its equilibrium value, l_{eq} or θ_{eq} is modeled by a harmonic potential. The constants k_{stretch} and k_{bend} represent one-half of the force constant of the bond or angle bending (i.e. spring), respectively. The torsional angle, ω, refers to the angle between two planes made up by the atoms A, B, C and B, C, D. The torsional term is described by a cosine function which takes the repetitive geometry of the bond rotation every 360° into account. The appearance of the cosine function is modified by the parameters k_{torsion} varying the amplitude of the curve, n influencing the periodicity, and γ shifting the curve along the rotation axis. Non-bonded interactions are calculated between all pairs of atoms, i and j, which are separated by at least two atoms. The most important contributions to the non-bonded interactions are nonpolar interactions between neutral particles modeled by a Lennard-Jones (LJ) potential, and electrostatic interactions between charged particles modeled by a Coulomb potential. The LJ potential is characterized by a very steep energy increase for interatomic distances r_{ij} smaller than the atomic van der Waals (vdW) radii. These repulsive forces are included in the first positive term $(1/r_{ij}^{12})$ which

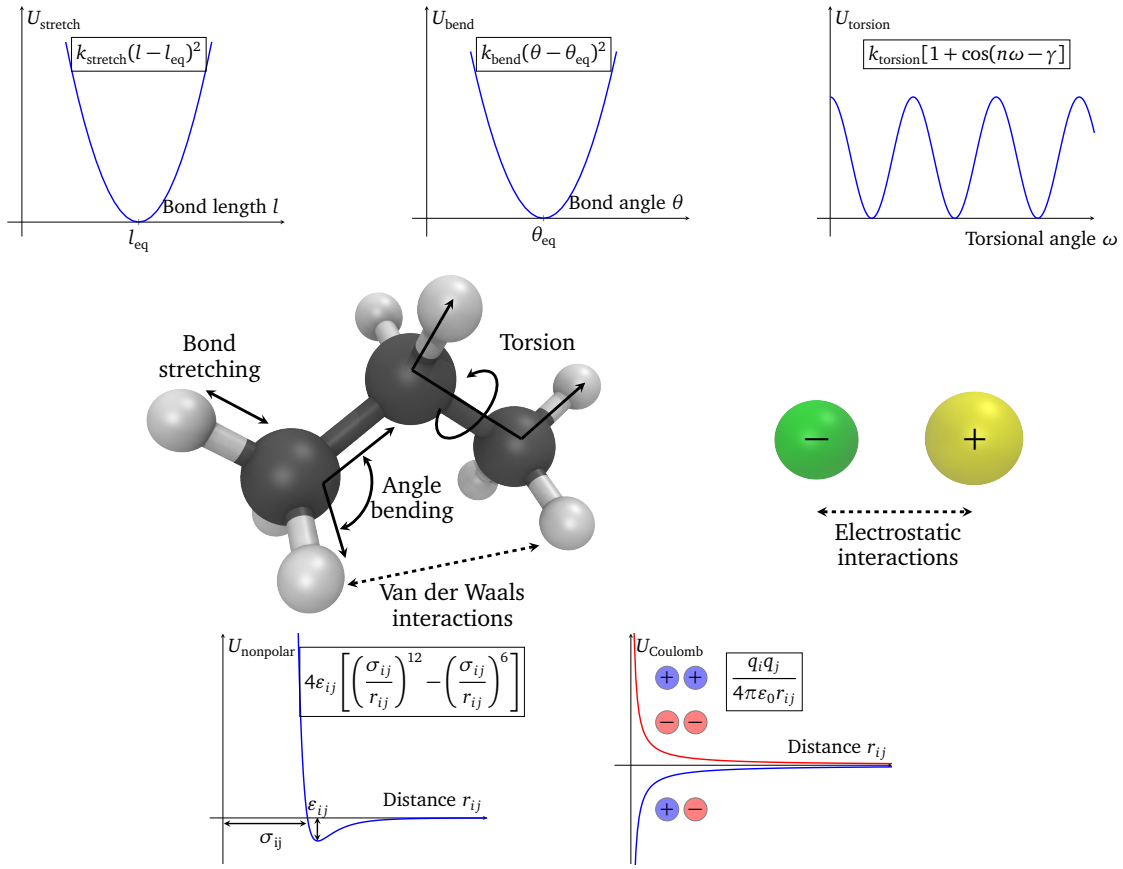


Figure 2.10: Potential energy terms in a force field. Bond stretching, bond angle bending, and torsional angle variations make up the bonded interactions in a molecule (upper panel). The non-bonded interactions comprise attraction and repulsion between nonpolar atoms and electrostatic interactions between charged particles (lower panel).

describes the Pauli repulsion representing the extremely unfavorable overlap of electron orbitals. The second term ($1/r_{ij}^6$) takes the attractive van der Waals or London dispersion interactions into account that occur at medium range when the atoms are approaching each other. In the expression for the LJ potential, ε_{ij} gives the depth of the potential well and σ_{ij} is the distance at which the LJ potential equals zero. Charged particles exert a repulsion or attraction that is usually described by the Coulomb potential. Electrostatic interactions lead to a repulsion of particles with equal charge and an attraction of differently charged particles. Consequently, the electrostatic energy depends on the particle charges, q_i and q_j, the dielectric constant ε_0 that incorporates the attenuation of the interaction due to e.g. a solvent environment, and the interatomic distance r_{ij}. Electrostatic interactions are significantly more long-range compared to the rather short-range LJ potential.

2.2.2 Kinetic energy and molecular dynamics simulations

Molecular dynamics (MD) simulations follow the temporal evolution of the coordinates and momenta of molecules. They provide a detailed picture of individual particle movements which leads to a comprehensive understanding of the molecules' structural and functional properties.[118,119] The calculation of future atomic positions, s_i, and other properties, is achieved by numerically integrating Newton's second law.[116]

$$F_i(s_1, s_2, \ldots, s_N) = m_i a_i = m_i \frac{d^2 s_i}{dt^2} \tag{2.4}$$

Together with equation (2.1), the relation between mass m and acceleration a of an atom i and the potential energy gradient of the associated force field is used to compute the forces, positions s, velocities u, and accelerations a for each atom at any point in time. In MD, the system state at any arbitrary point in future time can be calculated from the previous state, which means that MD is a deterministic procedure. Integration of the equation of motion is done using algorithms, with the Verlet algorithm[120] being the most commonly used one. Calculation of the particle motion is split into a very large number of short time steps δt that are in the order of a few femtoseconds. The time step is chosen small enough that the integration is numerically stable but not too small to require unnecessarily long computation times. For each δt, the forces acting on the particles of known initial position are computed, followed by the accelerations, velocities, and positions in the future time $(t + \delta t)$. Using the new positions, the updated potential energy is recalculated. In order to compute the temporal evolution of a particle, the so called trajectory, the integration algorithms employ a Taylor series approach as an approximation:

$$\begin{aligned}
s(t+\delta t) &= s(t) + u(t)\delta t + \frac{1}{2}a(t)\delta t^2 + \frac{1}{6}b(t)\delta t^3 + \ldots \\
s(t-\delta t) &= s(t) - u(t)\delta t + \frac{1}{2}a(t)\delta t^2 - \frac{1}{6}b(t)\delta t^3 + \ldots \\
s(t+\delta t) &= 2s(t) - s(t-\delta t) + a(t)\delta t^2
\end{aligned} \tag{2.5}$$

In the Verlet algorithm the positions and velocities from the current time step as well as the positions from the previous time step are used to calculate the future properties (equation (2.6)). The velocity is not calculated by a Taylor expansion but estimated by the following equation:

$$u(t) = \frac{s(t+\delta t) - s(t-\delta t)}{2\delta t} \tag{2.6}$$

The initial velocity distribution in an MD simulation is taken from a Maxwell-Boltzmann or Gaussian distribution corresponding to a specified temperature T having an overall zero momentum. An atom's probability to have a velocity u_x in x direction is given by

$$p(u_{ix}) = \left(\frac{m_i}{2\pi k_B T}\right)^{\frac{1}{2}} \exp\left(-\frac{1}{2}\frac{m_i u_{ix}^2}{k_B T}\right) \tag{2.7}$$

with k_B being Boltzmann's constant.

The instantaneous temperature $T(t)$ is related to the ensemble average kinetic energy, which corresponds to the average velocity of all atoms.

$$k_B T(t) = \frac{1}{\mathrm{DOF}} \sum_{i,x} m u_{i,x}^2 \tag{2.8}$$

DOF is the number of unconstrained degrees of freedom.[116] $T(t)$ is calculated from the velocity distribution, thus, during the MD simulation, the temperature is kept constant by velocity rescaling.

2.2.3 Challenges and limitations of MD simulations

MD is based on classical dynamics and the separation of nuclear and electronic motion. The effect of the electrons on the nuclei is implicitly accounted for by empirical potentials included in the force field. Thus, a transfer of electrons, which is required for the formation or breakage of bonds like in chemical reactions, cannot be modeled in MD.[121]

A second limitation of MD is the approximate nature of the force field. Force field parameterization is achieved by fitting the force parameters to experimental data or quantum mechanical results. More thoroughly parameterized force fields give more accurate and reliable results. One has to find a compromise between complexity and computational costs and accuracy. In non-polarizable force fields, a fixed atom-centered and constant point partial charge is assigned to the nucleus in order to treat the electrostatic properties of this atom. Obviously, this does not consider the fact that in reality the charge density (i.e. the electrons) is distributed in a cloud around the nucleus. Consequently, polarizability, that is the externally induced change of a molecule's charge distribution, is not considered in these non-polarizable force fields.[37] Furthermore, transferability of force field parameters between different molecular environments is possible but not straight forward. Thus, the usage of potentials or parameters in simulations different from the initial parameterization conditions should be treated with caution. Another approximation related to the force field is the truncation of long-range non-bonded interactions at a certain cut-off value. The computational work required in each time step in an MD simulation is dominated by the calculation of non-bonded interactions between every pair of atoms. Since the LJ potential rapidly approaches zero with increasing interatomic distances, it is generally reduced to zero at a given cutoff radius. Unfortunately, computational time cannot be saved with cut-off methods alone since they require the calculation of all interatomic distances. The solution is based on the generation of neighbor lists, also called Verlet lists. A list is created for each particle containing all the neighboring particles that are found within a sphere with a radius larger than the non-bonded interactions cut-off distance. The lists are periodically updated or newly generated and used to calculate pairwise interaction energies for atoms within this radius. The handling of electrostatic interactions is more challenging since they decay very slowly with distance. Here, typically, the particle mesh Ewald (PME) method[122,123] is used. The exact electrostatic potential can be calculated based on the Ewald summation introduced in 1921.[124] Following this approach the electrostatic potential U_{elec} is split into a direct, short-range term U_{sr} in real space and a long-range term in a reciprocal sum treated by using a Fourier transform U_{lr}.

$$\begin{aligned} U_{\mathrm{elec}} &= U_{\mathrm{sr}} + U_{\mathrm{lr}} \\ U_{\mathrm{elec}} &= \sum_{i,j} \varphi_{\mathrm{sr}}\left(r_j - r_i\right) + \sum_{k} \tilde{\Phi}_{\mathrm{lr}}(k)\,|\tilde{q}(k)|^2 \end{aligned} \tag{2.9}$$

The first term considers the summation of the short-range interaction potentials φ_{sr} between two particles i,j, and the second term includes the Fourier transforms of the potential $\tilde{\Phi}_{\mathrm{lr}}$ and the charge density $\tilde{q}$. In particle mesh methods, the reciprocal potential is calculated on a mesh.

The size of the investigated molecular system is another challenge in MD simulations.[116] Small systems do not provide a realistic picture; however, when the number of particles is too large, computational costs will become prohibitive. The use of periodic boundary conditions (PBC) prevents particles to be located close to an artificial boundary and to be influenced by surface effects. A unit cell is defined which contains the macromolecule under investigation in its physiological environment (i.e. a membrane, solvent, counter ions, etc.). Identical images of this unit cell are placed in all directions around the cell, with the included particles experiencing the same forces like the ones in the unit cell. When one particle leaves the cell, an identical image particle immediately enters the cell on the opposite side. Thus, each particle in the unit cell interacts not only with particles in 'its' cell but also with the images in the surrounding cells. For the calculation of interaction energies follows that the cut-off value should be at most one half of the box size. This minimum image criterion ensures that a particle does not interact with its own image and only with the closest image of any other particle. When PME is used with PBC, the electrostatic interactions are summed up for the periodically replicated system, which may lead to an unnatural ordering due to periodic artefacts. Analytical corrections to deal with self-interaction and to minimize the problems introduced by the artificial periodicity are standard in today's MD packages.[36]

Apart from size limitations also time restrictions play a role in MD simulations. Relevant protein motions occur on time scales in the range from 10^{-14} s (tens of femtoseconds) for bond vibrations, 10^{-12} s (picoseconds) for hydrogen bonding, 10^{-8} s (tens of nanoseconds) for reorganization of water structure, up to helix-coil breakdown or formation within hundreds of nanoseconds, allosteric transitions within milliseconds to seconds and protein folding within minutes.[125,126] With an integration time step of 2 fs the calculation of 'fast' events would take millions of steps, whereas the simulation of e.g. protein aggregation is not feasible at all. The above mentioned size and time limitations in MD simulations are known as the 'sampling problem'. Simulations need to cover time scales longer than the time scale of the investigated motion and length scales larger than the corresponding amplitude of motion.[126] Typically, latest MD simulations comprise several hundreds of nanoseconds for systems containing up to one million of atoms. The purpose-constructed supercomputer 'Anton' by D.E. Shaw Research is specialized on MD simulations of biomolecular systems and allows simulations of a protein folding over one millisecond.[43] Such simulations provide a huge amount of data. Therefore, a proper statistical evaluation of the MD trajectories is required to distinguish random dynamics from statistically significant conformational changes.

Except for the improvement of computing performance, e.g. faster hardware, better algorithms, or better parallelization, an alternative strategy to reduce computational costs is a less detailed description of the investigated system. One example is coarse-grained MD, which combines several atoms in one bead in order to reduce complexity and allow simulations of larger systems on longer time scales. An introduction into coarse-grained MD is given in section 2.2.5.

Besides the mentioned limitations there are several requirements for the numerical integrator in MD simulations. The time step should be chosen as a compromise between accuracy (δt as small as possible) and efficiency (δt large enough to prevent over-frequent computation of forces). In MD, the total energy of an isolated system is conserved. The discretization of variables during the numerical solution could, principally, destroy the energy conservation. To prevent this, the chosen numerical integration algorithm has to be phase-space volume conserving, i.e. symplectic, and time-reversible.[127] However, since the arithmetic operations in MD are performed with finite accuracy, i.e. MD algorithms are approximate; the computed trajectory will by and by increasingly differ from the actual trajectory. This is known as the Loschmidt paradox of macroscopic irreversibility.[128] Despite the fact that many-body atomic systems follow the equations of motions that should in principle be time reversible, those systems are chaotic with unstable dynamics.[129] That means, simulations that differ only slightly in their initial conditions will diverge exponentially following Lyapunov's

instability theorem.

$$\Delta s(t) \sim e^{\lambda t} \tag{2.10}$$

with $\lambda>0$ being a Lyapunov exponent which defines the deterministic chaotic motion in classical many-particle systems present in e.g. MD simulations.[130] Errors introduced by finite accuracy in the numerical integration steps destroy full time reversibility. The question which arises from this instability is, whether the true dynamics of a chaotic system can be derived from a computed trajectory at all. Yes, they can, if the shadow theorem holds true for sufficiently small time steps δt. The shadow theorem is a hypothesis that states the existence of a true trajectory which stays close to a computer-generated noisy trajectory, i.e. it shadows the noisy trajectory, the so called pseudotrajectory.[131] If one assumes that the system equilibrium properties are insensitive to the trajectory of an individual atom and that the pseudotrajectory would accurately describe a particle with the same initial conditions for a long time from a statistical point of view, then MD simulations are able to provide a reliable picture of reality.

2.2.4 Biased MD simulations and free energy methods

One distinguishes unbiased and biased MD simulations. An unbiased simulation, which is also called an equilibrium simulation, is not based on assumptions on the investigated system that influence the dynamics in a certain direction. The observed trajectory is purely an outcome of the internal system's properties and not 'guided' by external forces. Thus, it is highly unlikely to sample rare events, e.g. states that can be reached only by overcoming an energy barrier. The idea in biased simulations is to apply an external force in order to direct or steer the simulation in a certain direction. Steered MD (SMD) is such a method which introduces a mechanical force along a reaction coordinate and thereby improves the sampling of less populated states.[121] In 'constant pulling velocity' SMD simulations, a group of atoms is attached to a virtual spring that constrains the atoms to move with a constant velocity in the specified direction along a reaction coordinate ξ. In principle, the extension of the spring can be used as a measure for the force applied in order to pull the atoms and to induce conformational changes.

Based on SMD, the potential of mean force (PMF) is calculated. The PMF is the free energy difference between two system states as a function of the reaction coordinate with the reaction coordinate being the transformation from one state to the other state. The umbrella sampling (US) method is one important technique, belonging to the free energy methods that allow the calculation of relative free energies of different states. In transition state theory, free energy barriers between the different states correlate to the transition rates. In a canonical ensemble (NVT ensemble), the Helmholtz energy A is calculated from the potential energy U, the absolute temperature T and the entropy S.

$$A = U - TS \tag{2.11}$$

A system's canonical partition function Z is given by the integral over the phase space, i.e. coordinate space and momentum space. With U being independent from the momentum, Z is calculated as follows[132]

$$Z = \int \exp\left[-\beta U(\mathbf{v})\right] d^{\mathrm{DOF}}\mathbf{v} \tag{2.12}$$

Here, $\mathbf{v}$ is the system eigenstate, $\beta = 1/(k_B T)$, k_B is the Boltzmann's constant and DOF the number of degrees of freedom. The Helmholtz energy is related to the partition function via

$$A = -\frac{1}{\beta} \ln Z \tag{2.13}$$

In a system with constant pressure instead of volume, the Gibbs free energy G is calculated analogously.

In the US method, the transition from one state to another is covered by a sequence of windows along the reaction coordinate ξ (Figure 2.11). For each window, an MD simulation with a harmonic bias is performed, that gives a probability distribution of conformational states from which the free energy can be calculated. The bias is an additional energy term linking regions in phase space that are energetically separated.[132]

In an ergodic system, i.e. it is assumed that every state in phase space will be visited at least once, the ensemble average $Z(\xi)$ equals the system's time average probability distribution along ξ, $P(\xi)$.

$$P(\xi) = \lim_{t\to\infty} \frac{1}{t} \int_0^t p\left[\xi(t')\right] dt' \tag{2.14}$$

Here, t is the time and p is a count for the appearance of ξ in a certain interval. With equation (2.14) the free energy along the reaction coordinate can be determined from $P(\xi)$. Since the bias introduced an additional energy term into the system, a biased probability distribution P_i^{bias} along the reaction coordinate is calculated. However, the unbiased free energy $A_i^{\text{unbias}}(\xi)$ of each window is obtained from the biased MD simulation with

$$A_i^{\text{unbias}}(\xi) = -\frac{1}{\beta} \ln P_i^{\text{bias}}(\xi) - U_i^{\text{bias}}(\xi) + n_{i,\text{us}} \tag{2.15}$$

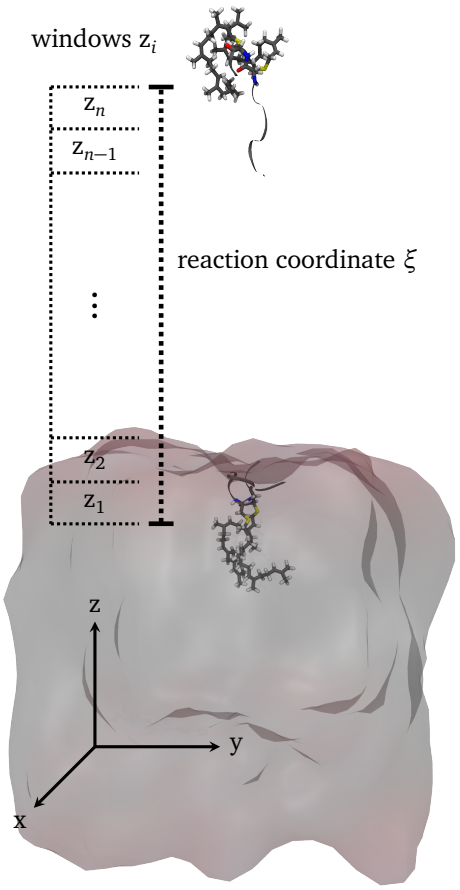


Figure 2.11: Schematic representation of an umbrella sampling calculation. The reaction coordinate ξ is divided into equidistant windows and a harmonically position-restrained MD simulation is performed in each window. Thus, the relative free energy between two states is calculated.

Here, the bias potential U_i^{bias} for each window i is an analytical term and the expression $n_{i,\text{us}} = -(1/\beta)\ln\langle\exp[-\beta n_i^{\text{bias}}(\xi)]\rangle$ is independent of ξ. The expressions $n_{i,\text{us}}$ are constants which link the free energy curves of the individual windows and have to be calculated by self-iteration when determining the global free energy profile.

In the present study, the free energy profile was obtained by combining the probability distributions from the individual windows using the weighted histogram analysis method[133,134] (WHAM). It is important in WHAM that the distributions of the individual windows are overlapping. Here, weights w_i are introduced that minimize the statistical error of the individual windows unbiased probability distributions $P_i^{\text{unbias}}(\xi)$.

$$P^{\text{unbias}}(\xi) = \sum_{i}^{\text{windows}} w_i(\xi) P_i^{\text{unbias}}(\xi) \tag{2.16}$$

Details concerning WHAM analysis can be found in the mentioned articles and reviews.[132–134]

2.2.5 The Martini model for coarse-grained MD simulations

Coarse-grained modeling approaches are based on the idea of reducing the number of particles by mapping single atoms into groups, thereby decreasing the complexity of a model. This allows an extension of the temporal and/or spatial description in the micro range and thus, far beyond full-atomistic simulations. The Martini force field is a popular coarse-grained model for biomolecular MD simulations developed by Marrink and Tieleman in 2007[135] and the successor of a lipid-only coarse-grained force field published in 2004.[136] It was developed based on atomistic models in a top-down approach by comparing and parameterizing the non-bonded interactions between the chemical components according to experimental thermodynamic data like partitioning free energies.[137] Reference full-atomistic simulations provide the template for the bonded interactions.

The following description of the Martini model is based on the 2013 review article of Marrink *et al.*,[137] which may be consulted for further details. Basically, a four-to-one mapping is used in Martini, that is four heavy atoms and their associated hydrogen atoms are combined in one bead that represents a single interaction center (Figure 2.12). Accordingly, four water molecules are represented by one coarse-grained water bead. Ions are mapped in one bead together with their first

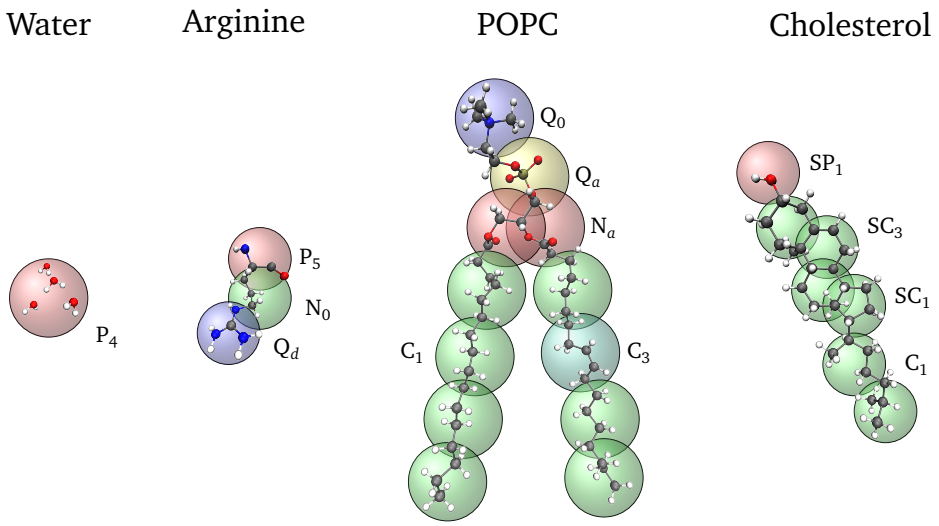


Figure 2.12: Mapping of atoms into beads according to the Martini force field. The topologies for lipids correspond to the 2007 release of version 2.0,[135] whereas protein topologies have been improved in version 2.2 from 2013.[138]

hydration shell. For molecules containing ring structures the mapping scheme is refined to allow a higher resolution with fewer heavy atoms in one bead. There are four main types of interaction sites: polar (P), non-polar (N), apolar (C), and charged (Q) bead types. In addition, with the hydrogen-bonding capability denoted by a letter (d = donor, a = acceptor, da = both, or 0 = none) several subtypes are defined. Further subtypes arise from the specification of the degree of polarity as a number ranging from 1 (low polarity) to 5 (high polarity). Non-bonded interactions are modeled by a LJ potential. Charged group interactions are modeled by a Coulombic potential. The Martini beads radii are defined by the minimum value of the van der Waals potential, namely 0.26 nm for regular and 0.23 nm for small beads. Point charges are located at a bead's center of mass. Similar to all-atom force fields, the bonded interactions are calculated based on mathematical expressions for the potential energy. In general, dihedral angle parameters are optimized to implement the peptide backbone secondary structure and impede artificial torsions of planar groups. This is required because directional hydrogen bonds, which are responsible for correct protein folding, are not modeled in Martini. In addition, the structure of a protein is further constrained to a native state by an elastic network. Thus, changes in secondary structure or the prediction of protein folding cannot be realized using the Martini model.

Besides the reduction of the particle number, the integrator time step may be increased considerably in coarse-grained simulations compared to full-atomistic simulations. Using the Martini model it may be increased to 20 fs when ring-like structures are considered or even to 40 fs in all other cases.[137] However, the advantages of coarse-graining a system, e.g. the reduction of computing time and therefore the feasibility of studies with more particles for longer simulation times, come along with some limitations. Obviously, coarse-graining results in a lower spatial resolution which may impair chemical accuracy. Less degrees of freedom affect the entropy which requires the adaptation of enthalpic terms in the Martini model. As a consequence, although free energy differences will be calculated accurately, the influence of enthalpic and entropic contributions to the free energy may be different compared to full-atomistic calculations. In addition, the energy landscape is smoothened which accelerates kinetics and allows an enhanced sampling compared to all-atom systems within an equal period of time. From lateral diffusion studies of lipids[135] and peptides[139] in membranes, the Martini speed-up factor was estimated to be 4. However, this should be considered with caution since it depends on the molecule type and differs with different degrees of freedom.

3 Methods and technical details

Chapter 3 gives general recurring methods which are the basis of the present thesis. For a more comprehensive description of the system-dependent preparation, implementation and analysis the reader is referred to chapters 4 to 6.

This chapter provides a general workflow showing the preparation, modeling and analysis of MD simulations which applies for all simulated systems. The second part of this chapter gives an overview of the parameter settings for protein MD simulations using the software NAMD. Furthermore, selected characteristics of membrane-protein simulations are described in more detail. The calculation of relative free energies from steered MD and umbrella sampling as well as general statements about MD simulation analysis are provided in the last two sections of this chapter.

3.1 General workflow and software

An overview of the different steps of preparation, performance and subsequent evaluation of an MD simulation is given in Figure 3.1. Prior to simulating a protein-membrane system with MD, models for the individual components, i.e. the bilayer and protein, were prepared. In general, protein experimental structures derived from X-ray crystallization or NMR spectroscopy are used as high-resolution input structures for MD simulations. However, due to internal flexibilities of the macromolecule or inadequate experimental conditions, the derived structures are not always complete or of sufficiently high quality. In this study, the experimental structures were used as templates for creating homology models of the full-length protein. The procedure of protein modeling using the software package Modeller version 9.12[140] and Schrödinger's MacroModel version 10.0 is described for Rab5 in chapter 4 and for GDI in chapter 6. The assembling of the model lipid bilayers was done with the web-based CHARMM-GUI[141] version 1.5 membrane builder.[142] The membrane and protein systems were combined, solvated and ionized using TCL (Tool Command Language) scripts via the TCL scripting interface Tk Console in VMD (Visual Molecular Dynamics) version 1.9.1.[143] The computer simulations, i.e. the energy minimization, heating and MD simulation, were performed with the scalable molecular dynamics code NAMD version 2.9.[144] Details on the preparation and implementation of the full-atomistic MD simulations are given in section 3.3 and the respective chapters. Visual analysis of the MD trajectories was done with VMD, processing of data was performed using the python based library MDAnalysis,[145] python and TCL scripting as well as the software MATLAB R2014b. Data plotting was performed in LaTeX using the 'pgfplots' package.[146]

3.2 MD simulation setup in NAMD

All full-atomistic MD simulations were performed with the software NAMD v2.9. Minimization, heating and short equilibration periods were run on the 'otto' high-performance computing (HPC) cluster at the Max Planck institute for Dynamics of Complex Technical Systems, Magdeburg on 240 parallel CPUs. Long equilibration and production runs were performed on 960 CPUs on the HPC system 'HYDRA' provided by the Max Planck Computing and Data Facility in Garching. After the installation of the new extension cluster 'DRACO' in May 2016, umbrella sampling simulations were run on 'DRACO' using NAMD v.2.11 on 960 CPUs.

Usually, in MD simulations without external control, the equations of motion are solved in a micro-

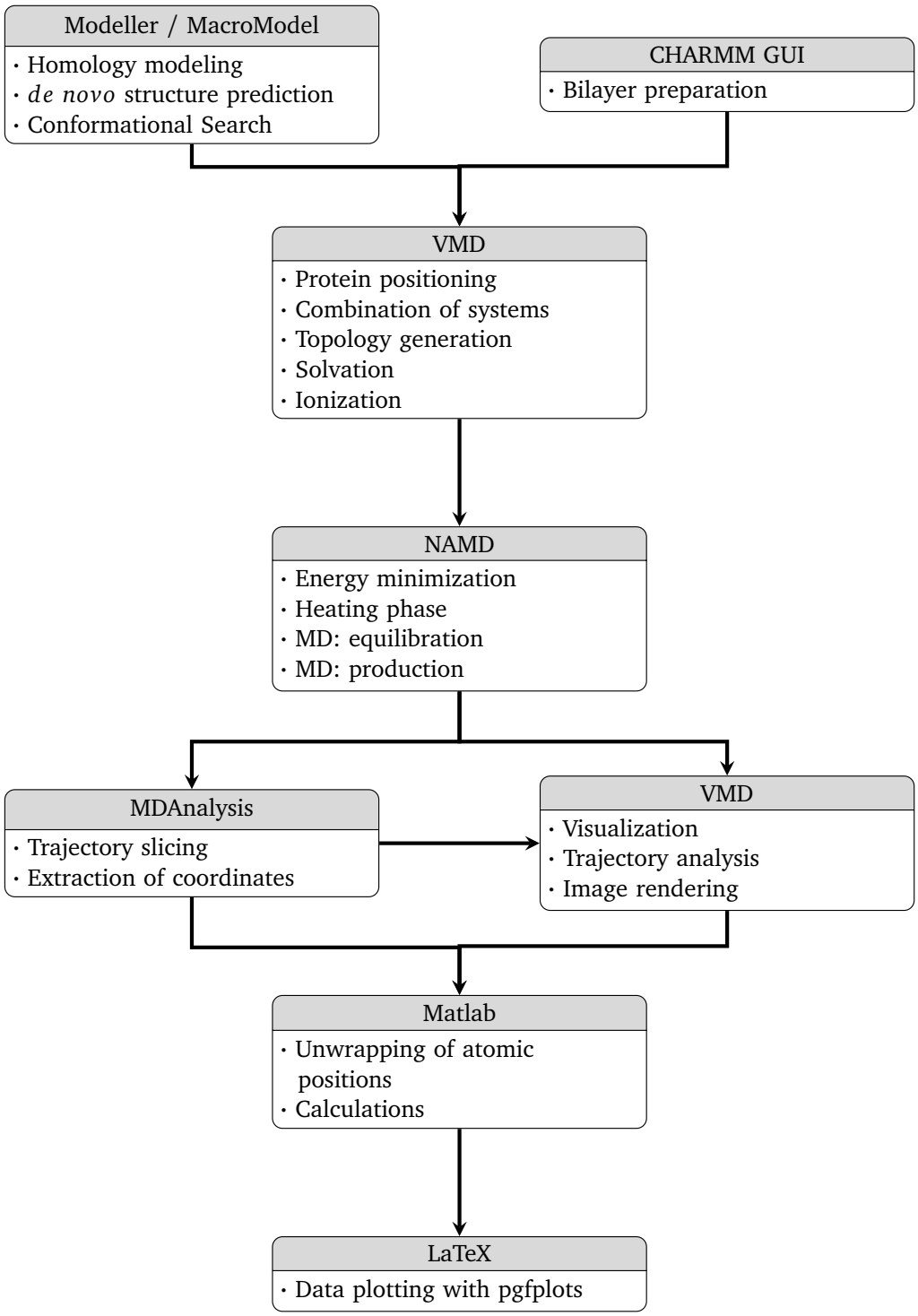


Figure 3.1: Workflow for the preparation, modeling and analysis of protein-membrane MD simulations.

canonical ensemble (NVE ensemble), which means the number of particles, volume and energy are constant for an isolated system. In order to match experimental conditions, it is more appropriate to perform simulations at constant temperature and pressure (isothermal-isobaric ensemble, NPT).[147] Temperature control is achieved by coupling the system to an external heat bath of constant temperature. Additional forces are introduced to the system in order to scale atomic velocities and maintain a constant temperature using Langevin dynamics.[148] In constant pressure simulations, the pressure is controlled by adjustment of the unit cell size and rescaling of atomic coordinates. Thus, similarly to temperature control, the system is coupled to an external piston, i.e. a barostat. In membrane systems, the method of pressure coupling is of prime importance.[149] One distinguishes between i) isotropic, ii) semi-isotropic, and iii) anisotropic pressure coupling. With isotropic pressure coupling, the scaling factor is equal for the contributions in all three directions (x, y, z), that is all directions are coupled and only small box size changes are expected. The isotropic pressure coupling is recommended for simulations in water boxes, but not suitable for membrane simulations since fluctuations in the surface area cannot occur. Semi-isotropic pressure coupling scales the pressure contributions in x, y directions independently from the z direction, i.e. x, y directions are coupled. With this, area fluctuations are allowed which makes semi-isotropic pressure coupling the best choice for membrane simulations. With anisotropic pressure coupling all three direction scaling factors are independent from each other, allowing uncoupled size fluctuations which may lead to extreme distortions of the simulation cell. In the present study simulations were performed in an NPT ensemble with semi-isotropic pressure coupling for membrane systems.

After topology building and setup of the MD simulation box, the potential energy of the system was minimized using the default conjugate gradient algorithm implemented in NAMD. A convergence criterion was specified that defined the minimization to be successful when the energy difference between two steps was $< 10^{-6}$ kcal mol^{-1}. Minimization is done to avoid numerical instabilities of the subsequent simulation due to high interatomic repulsions occurring between atom pairs of unfavourable distances. After convergence, the system was gradually heated from 0 K to a temperature of 310 K, thereby increasing the temperature by 0.001 K in each integration step. Subsequently, the system was equilibrated in a brief MD simulation with fixed protein coordinates. Unconstrained, unbiased MD simulations were then performed for 250 ns at 310 K in an NPT ensemble. The first

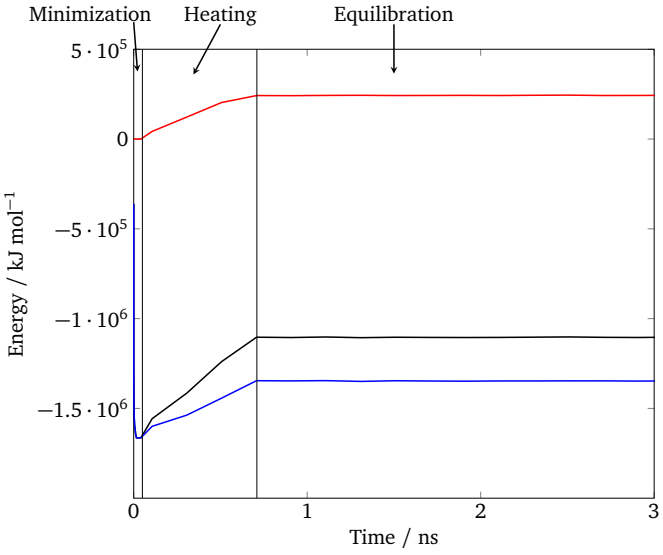


Figure 3.2: Energy changes during the processes of minimization, heating, and molecular dynamics equilibration. The total energy (—) is the sum of potential (—) and kinetic (—) energies. During potential energy minimization the kinetic energy equals zero. While heating the system to a specified temperature, the energy increases. During equilibration the temperature is maintained and the system energy fluctuates around a constant value.

50 ns of the MD simulations were classified as equilibration period and excluded from analysis. The changes in the system's potential, kinetic and total energy over the different simulation phases are shown in Figure 3.2.

General input files for a simulation in NAMD are the CHARMM force field files, a topology file containing information about the molecular structure (PSF file), the initial coordinates of the system (PDB file) as well as a NAMD configuration file. The configuration file includes all settings, options and parameters needed to perform a minimization, heating or MD simulation in NAMD. It is structured in different parts where settings corresponding to numerical integration, calculation of interactions, box shape and periodic boundary conditions as well as temperature and pressure control are specified. An overview of important configuration parameters is provided in the appendix, Table A1.

3.3 Steered MD and umbrella sampling

In preparation for a steered MD (SMD) run and subsequent umbrella sampling (US), the system size was reduced to fit the small part under investigation. Analogous to the preparation of an unbiased MD simulation, this small system was energy minimized, heated to 310 K and equilibrated for 2 ns. A second energy minimization was performed directly before starting the constant pulling velocity SMD simulation. For convenience, the pulling direction was chosen to be the +z direction and the sampled reaction coordinate ξ was the distance between the constrained atom group and the bilayer surface. The phospholipid head groups were harmonically constrained by a force constant of 5 kcal mol^{-1} Å^{-2} in z direction in order to prevent the extraction of bilayer lipids during the pulling procedure. The parameter options used for a SMD run in NAMD are provided in the appendix, Table A2.

The SMD simulations were used to generate input structures for the individual windows in US calculations. The distance between the SMD pulling atom group and the upper bilayer was taken as the reaction coordinate. A predefined number of equidistant windows was defined along this reaction coordinate and in each of the individual windows an energy minimization, heating as well as an US run were performed. In NAMD, the reaction coordinate is called a 'collective variable' and the 'colvar' module may be used to implement a large variety of sampling methods, e.g. metadynamics, adaptive biasing force (ABF) method, SMD, and US. The individual starting conformations were a reaction coordinate distance of 1.3 Å apart from each other and the actual force constant of the harmonic bias on the corresponding atoms was 4 kcal mol^{-1} Å^{-2}. More details regarding the setup of SMD and US calculations are provided in chapters 4 and 6.

3.4 General statements regarding MD simulation analysis

Due to the large size of the investigated systems, the resulting trajectories contained binary data of at least 200 GB which needed to be processed prior to analysis. The python module MDAnalysis[145] was used to save every 100$^{\text{th}}$ frame of the complete trajectory, resulting in a new sliced trajectory with one frame each 0.4 ns. The sliced trajectories were used for visualization in VMD, data extraction with VMD and MDAnalysis and eventually calculation of structural and dynamic properties using MATLAB.

The evaluation of membrane lipid properties derived from MD simulations helps to compare computational results to experimental data from NMR, X-ray or neutron scattering experiments. However, in contrast to experiments that only provide spatial and temporal averages of particular properties, MD simulations are able to also represent local properties. Different membrane properties were primarily investigated in case of the peripheral peptide/Rab5-membrane simulations presented in detail in section 4.4.3. A comprehensive description of the individual MD simulation analysis techniques is provided in the respective chapters where these methods are first applied.

4 Protein-lipid interactions of membrane-bound Rab5

Early endosomes (EEs) are sorting stations for internalized cargo, which is either recycled back to the plasma membrane, fed to the Golgi transport system or sorted for degradation in lysosomes.[113] These sorting processes are controlled by proteins on signaling platforms on the endosome surface which activate intracellular signaling after external or membrane receptor-derived stimuli. The small GTPase Rab5 is the key regulator of EE trafficking and involved, amongst others, in homotypic fusion of EEs, formation of CCVs, endosomal motility and protein sorting within endosomes.[2,3,88,150] In addition, Rab5 indirectly mediates signal transduction via its various interaction partners. For example, Rab5 activity is intertwined with the Ras-activated endocytosis of epidermal growth factor (EGF) receptors via RIN1 (Ras and Rab interactor 1) which in turn interacts with Ras.[151] Moreover, Rab5 is involved in receptor tyrosine kinase induced actin remodeling through the Rab5-specific GAP RN-tre (Related to the N terminus of tre oncogene).[152] Its tremendous spectrum of effector proteins makes correct Rab5 function absolutely crucial for endosomal processes and cell survival. Dysfunction of Rab5 is associated with different diseases of the immune system,[8] cancers[153] as well as endocytosis- and autophagy-related defects that lead to neurodegeneration.[13]

Rab5 activity and function depend on its correct targeting to membranes, its subcellular localization whereby cycling between the bilayer surface and the cytoplasm, and its nucleotide-determined activation state (i.e. inactive GDP-bound vs. active GTP-bound state). Thus, a comprehensive understanding of Rab5 membrane targeting, the association with the EE membrane and the conformational differences between the inactive and active protein states that regulate the interplay with regulatory and effector proteins is essential for treating and, eventually, curing Rab5-related malfunctions.

Generally, protein structural information is provided by experimental techniques like X-ray crystallography or NMR. However, in case of Rab5, the long highly flexible C terminus, the so-called hypervariable region (HVR) was not able to be structurally resolved to date. The HVR connects the catalytically active G domain with two hydrophobic post-translationally attached geranylgeranyl (GG) chains which serve as a membrane anchor. Consequently, crystal structures of Rab5 are missing the membrane-associated protein part which rules out the possibility to predict membrane-induced effects on the protein structure and orientation. The full-length Rab5 structure was generated by comparative modeling in order to allow such predictions. MD simulations have been proven a valuable tool to investigate protein structures, dynamics and protein-membrane interactions on an atomistic level.[154,155] The present simulations improve the understanding of the dynamics of membrane-bound Rab5 HVR and full-length Rab5(GDP) and Rab5(GTP), their association with model bilayers of different composition and may assist in the development of new therapeutic strategies against Rab5-induced disorders.

This chapter is dedicated to the orientation and dynamics of the membrane-associated small GTPase Rab5 (see section 2.1.1, Figure 2.2). Its conformational structure and behavior at different model membranes is investigated thoroughly by MD simulations.

The first part of this chapter focuses on the model generation and structure refinement of human full-length Rab5 by comparative modeling and MD. Section 4.2 introduces model membranes of increasing complexity that were used in this thesis in order to analyze the influence of the lipid composition on the protein. Subsequently, the validation of the force field parameters for the post-translationally attached GG modifications, functioning as the Rab5 lipid anchor, is shown. The main part of this chapter addresses the investigation of membrane-bound Rab5 by MD simulations. First,

the simulation setup and analysis tools are described. Subsequently, the results of the MD simulations are discussed. These comprise the bilayer composition-dependent dynamics of the membrane-bound Rab5 C-terminal HVR as well as of full-length Rab5 in its inactive and active states. Furthermore, the extraction free energy profile of the lipid anchor from the bilayer is determined and discussed with regard to the lipid composition.

Significant parts of this chapter were published in research articles by the author of this thesis.[8,156]

4.1 Full-length Rab5 model generation

Due to highly flexible regions at the protein N- and C-termini the full-length structure of Rab5 could not be determined by X-ray diffraction. The missing residues were modeled using the comparative modeling approach implemented in the Modeller software.[140,157] A further refinement of the terminal structures was performed using a torsional sampling procedure implemented in MacroModel from the Schrödinger program package.

Here, two experimentally derived crystal structures for the Rab5 active and inactive states, hereinafter referred to as Rab5(GTP) and Rab5(GDP), were chosen. They correspond to the protein databank (PDB) entries 1R2Q[72] and 1TU4 (chain A),[73] respectively. For the simulation, the originally bound phosphoaminophosphonic acid-guanylate ester (GNP) in 1R2Q was replaced by GTP. Both crystal structures contain the structural information for the evolutionary conserved G domain comprising residues Gly^{15} to Asn^{184} or Glu^{185} for Rab5(GDP) or Rab5(GTP), respectively. The following steps were performed for both Rab5 states individually. A number of ten individual models was built with Modeller from which the model with the lowest Discrete Optimized Protein Energy (DOPE) score[158] was chosen for further refinement. DOPE is a statistic potential based on a finite sphere reference state with uniform density depending on the native structures sizes which generated the statistical potential. It is implemented in Modeller and used to assess the quality of a structural model. For further refinement a torsional sampling conformational search was done with MacroModel using the OPLS-2005 force field.[159] A conformational search is a tool to explore the potential energy surface of a molecule, to identify lowest energy conformations that may differ in structure. Typically, a conformational search procedure starts with the generation of a new structure which is subsequently energy minimized and then decided to be retained or not. If a structure is retained depends on its energy relative to the previously found conformers and possible structure redundancies. Here, the stochastic searching method Monte Carlo Multiple Minimum (MCMM) was used. In this torsional sampling procedure, different geometries were generated by randomly changing torsion angles. During this procedure approximately 60 distinct conformers were identified which were clustered into five groups according the internal clustering of conformers script. The root mean square deviation (RMSD) matrix was calculated for the flexible N- and C-terminal regions. The conformer that was closest to the RMSD average of the largest cluster was chosen as the final protein model. Subsequently, this full-length Rab5 model was equilibrated in a 50 ns NPT MD simulation in explicit water. The residues comprising the G domain that originated from the X-ray structure were kept fixed during this first MD simulation. Details concerning the setup of MD simulations using the NAMD software are provided in section 3.2. After this 50 ns equilibration the GG chains that serve as Rab5 lipid anchor were covalently attached to the protein cysteine residues Cys^{212} and Cys^{213} at the C-terminus (see section 2.1.1, Figure 2.2). The final model was solvated, energy minimized and modeled for 50 ns in an unconstrained MD simulation at 310 K.

After 50 ns of MD simulation, the conformational structural differences between Rab5(GDP) and Rab5(GTP) were investigated. The structure of the catalytic G domain as well as the superposition of both cytoplasmic Rab5 structures are displayed in Figure 4.1. The structure of the G domain is conserved between different Rab proteins. Five α-helices (α1-α5) form a central core barrel which is permeated by six β-sheets (β1-β6). β1 and α5 demarcate the G domain from the disordered structures of the Rab5 N- and C-termini, respectively. A sequence patch in α1 is also highly conserved

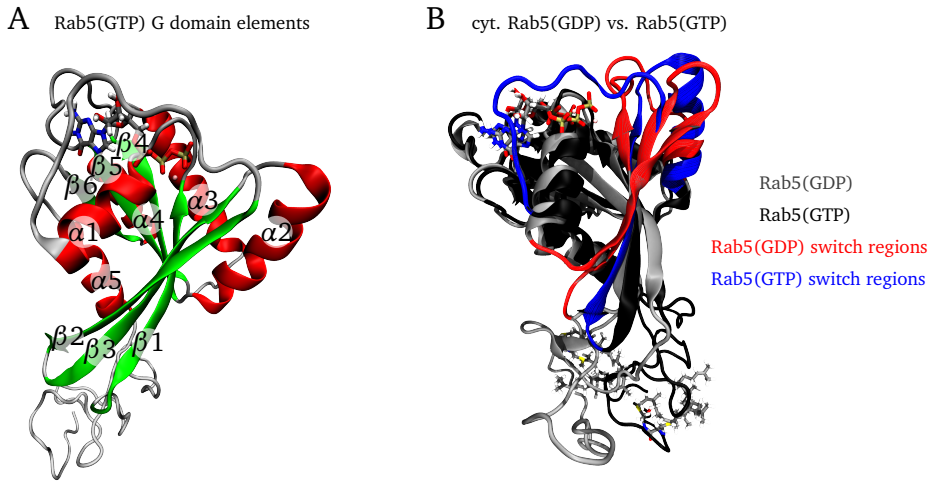


Figure 4.1: A) Secondary structure elements of the Rab5(GTP) G domain are conserved among different Rab proteins. B) Superposition of modeled full-length Rab5(GDP) (gray, switch regions: red) and Rab5(GTP) (black, switch regions: blue) in cytoplasm.

among Rab proteins and serves binding of GTP and Mg^{2+}, the latter being a co-factor required for nucleotide binding and hydrolysis.[160]

Apart from the flexible HVR, most striking differences between Rab5(GDP) and Rab5(GTP) were observed within the switch regions, which comprise the stretches between α1 and β2 (Rab5 residues 44-66, switch I) and between β3, α2, and β4 (Rab5 residues 75-91, switch II). These regions exhibit activation state-dependent structural properties and bind regulatory and effector proteins. In Rab5(GDP), the switch I forms two antiparallel β-strands with the N-terminal part being locally twisted (β2' strand), the second strand being β2. The switch II region constitutes a disordered coil flanking the GDP terminal phosphate group. In Rab5(GTP), the switch I is a long curved coil which coordinated the entire GTP nucleotide and ends in the β2-strand. The switch II region forms a loop and the α2-helix.

4.2 Preparation of membrane models of increasing complexity

In the present study, three different symmetric model membranes of increasing complexity were used. A neutral pure palmitoyl-oleoyl-phosphatidylcholine (POPC) bilayer was built with the VMD Membrane Plugin.[143] A ternary mixture of POPC, cholesterol and palmitoyl-sphingomyelin (PSM) in a ratio 2:2:1 serves as an example for a simple plasma membrane model often used in experiments with the ability to form liquid-ordered (lo) domains.[109] The six-component model membrane features an overall negative charge due to the presence of palmitoyl-oleoyl-phosphatidylserine (POPS) and phosphatidylinositol 3-phosphate (PI(3)P). PI(3)P is an important signaling lipid primarily enriched in the EE membrane where it serves as a recognition signal that recruits cytoplasmic effector proteins to promote membrane deformation, tethering, membrane fusion etc. (for details see section 2.1.4). The six-component membrane was therefore used as a complex model system for the EE membrane. The exact lipid compositions of the model bilayers are shown in Table 4.1. The ternary and six-component bilayers were built using the CHARMM-GUI Membrane Builder.[142,161] The model membranes were solvated, ionized, energy minimized and heated to a temperature of 310 K. First, equilibration was performed for 2 ns in an NVT ensemble with all atomic coordinates kept fixed except for the lipid tails in order to introduce sufficient disorder in the tails region. In a second equilibration step the fully unconstrained membrane system was simulated for another 50 ns in an NPT ensemble. After these steps protein as well as membrane systems were combined.

Table 4.1: Lipid compositions and sizes of model bilayers used in the present MD study. Abbreviations as follows: POPC: palmitoyl-oleoyl-phosphatidylcholine, PSM: palmitoyl-sphingomyelin, POPE: palmitoyl-oleoyl-phosphatidylethanolamine, POPS: palmitoyl-oleoyl-phosphatidylserine, PI(3)P: phosphatidylinositol 3-phosphate.

Membrane system	Lipid type	Lipid number per leaflet	Mole fraction	Lateral (x,y) dimensions / nm
Pure POPC	POPC	273	100.0 %	15.7 x 14.9
Ternary	POPC	166	40.0 %	14.8 x 14.4
	Cholesterol	166	40.0 %	
	PSM	83	20.0 %	
Six-component	POPC	90	17.8 %	16.9 x 16.6
	Cholesterol	150	29.7 %	
	PSM	50	9.9 %	
	POPE	135	26.7 %	
	POPS	55	10.9 %	
	PI(3)P	25	5.0 %	

4.3 Force field parameters for geranylgeranylated cysteine

At the beginning of the present study there were no topologies and force field parameters available for the GG post-translationally modified cysteine amino acid (GG-Cys). The force field parameters required for this residue were adapted from the published CHARMM36 topologies for proteins[162] and lipids.[163] The resulting nomenclature for GG-Cys is given in Figure 4.2.

Quantum mechanical DFT (density functional theory) calculations were performed using TURBOMOLE V6.6 in order to validate the newly derived force field parameters. Geometry optimizations used a split-valence basis set (def2-SVP)[164] and the pure BP86 functional.[165,166] The hydrophobic core region of the surrounding membrane lipids was mimicked using the conductor-like screening model (COSMO)[167] with a dielectric constant of ε=2.[168]

In the MD simulations, four critical dihedral angles along the GG chain were identified (Figure 4.3). The torsion types reflect the repeating units of the GG chains. These involve rotations around:

A) a carbon-carbon single bond with one terminal carbon-carbon double bond

B) a carbon-carbon double bond

C) a chain of single bonded carbon atoms including one central sp^2 hybridized carbon atom

D) a chain of single bonded carbon atoms with two terminal sp^2 hybridized carbon atoms

Potential energy scans were performed using a triple-zeta basis set (def2-TZVP) to investigate the energy barriers of the torsion associated with the respective dihedral angles. The resulting energy profiles were compared to the probability distributions of finding the corresponding dihedral angle in the MD trajectories (Figure 4.4).[156] Dihedrals which corresponded to the energy minima of the quantum chemically calculated torsional profiles were most frequently visited in the MD simulations. This demonstrates the suitability of the GG-Cys force field parameters to accurately reproduce quantum mechanical results. Therefore, the chosen parameter set was used for simulating the Rab5 GG anchor in membranes as part of the full-atomistic protein-membrane MD simulations.

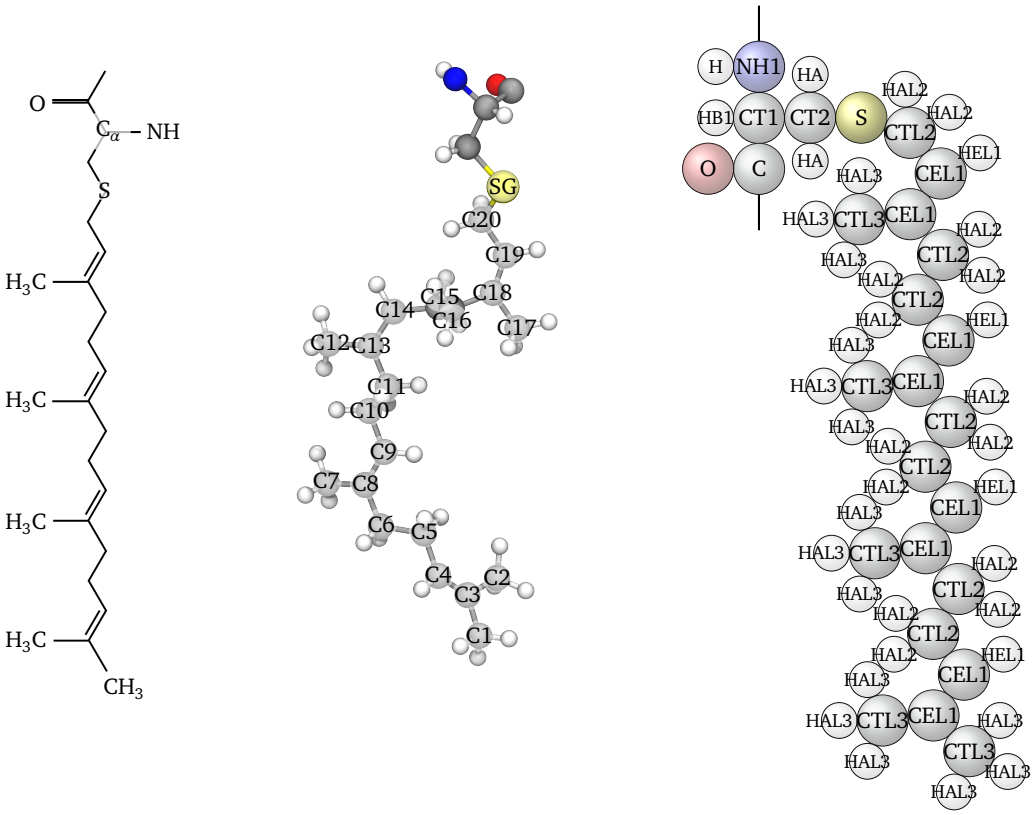


Figure 4.2: Nomenclature of the GG-Cys residue as used in the simulations (atom names are shown left). Published parameters and topologies of the CHARMM36 force fields for proteins[162] and lipids[163] were combined to assemble the GG-modified residue (right).

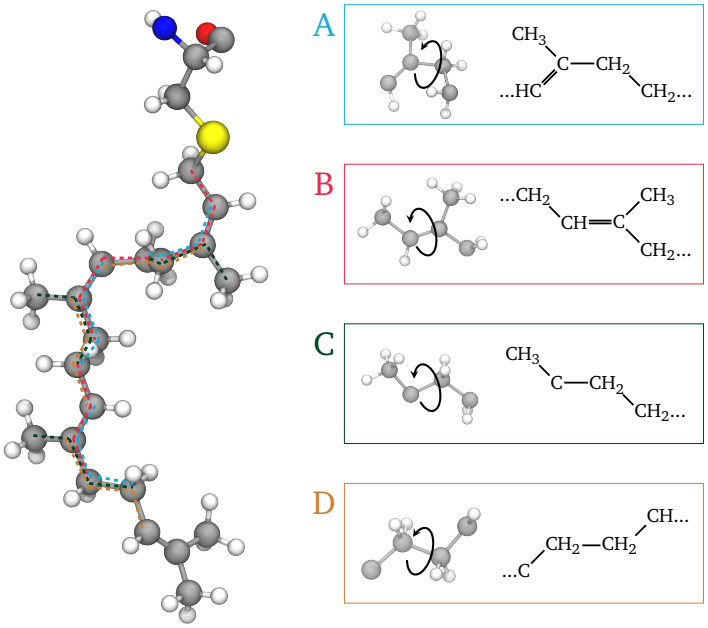


Figure 4.3: The GG chain is composed of four major repeating chemical units. The free energy of torsion around the dihedrals A-D was investigated by means of QM calculations and compared to the MD-sampled conformations.

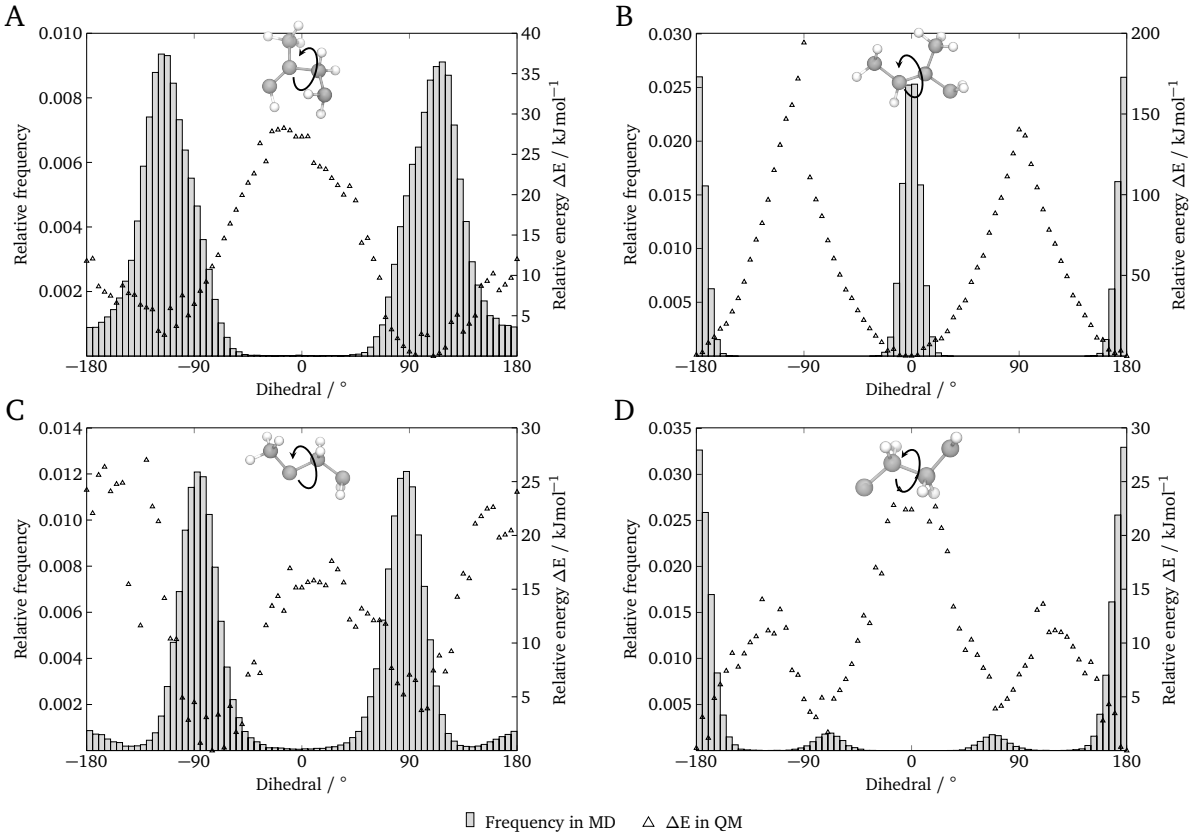


Figure 4.4: The relative frequency of dihedrals of types A-D from MD simulations (histograms) was compared to the torsional energy profile derived from QM calculations (Δ).[156]

4.4 MD simulations of membrane-bound Rab5 HVR and full-length Rab5

The Rab5 lipid anchor is located within the structurally highly flexible HVR which comprises Rab5 residues 186-215. Due to the lack of experimental structural data, little is known about the lipid interactions of the Rab5 HVR apart from membrane insertion via the GG chains. In this section, the interactions of a truncated Rab5 HVR (residues 206-215, hereafter called $HVR^{206-215}$) are investigated by full-atomistic and coarse-grained MD simulations. The results are compared to full-length Rab5(GDP) and Rab5(GTP) MD simulations with special emphasis on the role of the HVR for the protein orientation. The contribution of the HVR to membrane binding is further stressed by the determination of free energy profiles, providing a measure for the strength of peptide-lipid interactions in different model membranes. The present study on the Rab5 HVR lines up with further computational and experimental studies which investigated the membrane-bound C-terminus of different small GTPases using a small peptide sequence.[169–171]

4.4.1 Full-atomistic MD simulation setup

In all protein-membrane simulations the bilayer was placed in the x, y plane and the $HVR^{206-215}$ or full-length protein, respectively, was localized in +z direction above the membrane surface. Membrane anchoring occurred via the two C-terminal GG chains which were inserted by superposing both prenyl groups with the lipid tails of one POPC molecule. Afterwards, the corresponding POPC lipid was deleted.

- In case of the HVR calculations, six $HVR^{206-215}$ replicates were inserted into each model mem-

brane. For each system 200 ns of production MD simulation were performed.[156]

- In the case of full-length protein, one Rab5 entity was membrane anchored by two GG-Cys residues and the G domain of Rab5 was translated to a position perpendicular to the membrane. For each model membrane and each nucleotide state three independent 200 ns production MD simulations were performed. Rab5 simulations in the six-component membrane were extended by another 250 ns each, yielding a total simulation time of 1.35 µs for both Rab5(GDP) and Rab5(GTP).[8]
- Additionally, mono-GG Rab5(GDP) and Rab5(GTP), respectively, were anchored to each membrane model and subjected to 200 ns production MD simulation.

In addition to these unbiased MD simulations (Figure 4.5), some biased SMD simulations as well as US runs were performed in order to calculate the relative free energy of the GG anchor bilayer-to-water transfer in the three different model membranes. The structure of one HVR$^{206-215}$ replicate as well as a small bilayer patch of surrounding lipids and a water layer of at least 10 nm height above the membrane surface was prepared. Details of the US calculations setup are provided in Table 4.2.

Topologies and structures were created using the VMD psfgen plugin, which reads atomic coordinates in PDB format and force field topology files. For topology generation and subsequent MD simulations the CHARMM36 force fields sets were used.[162,172–179] Similar to the GG-Cys residue, the parameters for the nucleotides GDP and GTP associated with Rab5 were not available when the present study was started.[8] Therefore, the required topologies were adapted from the CHARMM36 nucleic acid

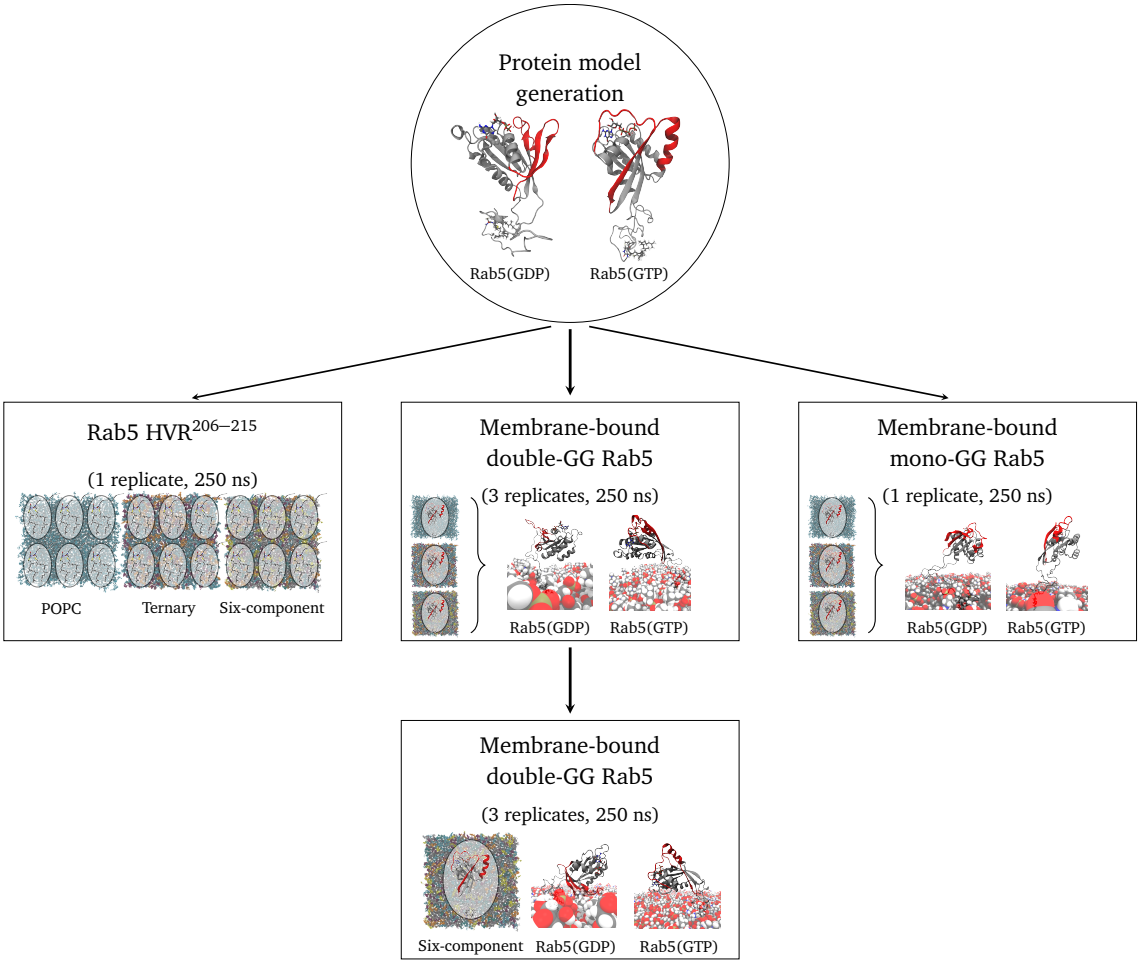


Figure 4.5: Full-atomistic MD simulations performed for a truncated Rab5 HVR$^{206-215}$, full-length mono-GG Rab5(GDP) and Rab5(GTP) as well as full-length double-GG Rab5(GDP) and Rab5(GTP) in three different model membranes.

Table 4.2: Setup of Rab5 HVR systems for which binding free energies were calculated using US calculations.

Simulated system	$HVR^{206-215}$		
Calculated PMF	Binding free energy of GG anchor to bilayer		
Membrane model	Pure POPC	Ternary mixture	Six-component membrane
Number of replicates per membrane	3	3	3
Reaction coordinate ξ	Z-distance between upper membrane layer (P atoms) and Cα atoms of GG-Cys residues		
Sampled distance on ξ / nm	-0.14 3.00	-0.40 3.00	-0.27 3.60
Number of windows	25	27	31

force field[180] and combined with phosphate group parameters from ADP/ATP.[181]

The protein-membrane system was solvated with explicit water using the transferable intermolecular potential 3P (TIP3P) water model[172] and the solvate plugin. The water intruding the bilayer was deleted based on geometric restrictions in the z direction. Finally, the system was neutralized by adding counter ions and subsequently ionized to a physiological salt concentration of 0.15 M sodium chloride with the autoionize plugin.

4.4.2 Rab5 HVR coarse-grained MD simulation setup

Coarse-grained MD simulations of the membrane-bound Rab5 $HVR^{206-215}$ were run with the software GROMACS[182] v. 5.0.7 using the coarse-grained force field Martini version 2.2 topologies and parameters.[138] These simulations were set up and performed by Eric Schulze as part of his master thesis project and further studies. Membrane models comply with the bilayers in the full-atomistic simulations, i.e. a pure POPC bilayer, a ternary mixture composed of POPC, cholesterol and PSM, as well as a six-component bilayer resembling the composition of the EE membrane. In addition, similar to the six-component bilayer a five-component membrane was investigated, in which the threefold negatively charged signaling lipid PI(3)P was replaced by single negatively charged POPS.[156] Using this approach the influence of PI(3)P compared to another negatively charged lipid should be elucidated. The python tool 'insane.py'[183] was extended to include PI(3)P and used to carry out membrane assembly and the insertion of the GG anchor chains. The systems were solvated by Martini water which is composed of nine parts solvent and one part anti-freeze particles. Subsequently, the systems were neutralized and ionized to a salt concentration of 0.15 M sodium chloride. The lipid types POPC, POPE, POPS, and cholesterol were modeled with the Martini lipid topologies version 2.2.[183] The DPSM lipid topology version 1 was used for the PSM lipid. The topologies of PAPI version 1 and POP1 version 1 were combined to model the PI(3)P tails and backbone as well as the phosphorylated inositol head group, respectively. The Martini model for the Rab5 $HVR^{206-215}$ was generated in accordance with the full-atomistic full-length Rab5 protein structure. Coarse-graining was done using the 'martinize.py' tool version 2.4.

Coarse-grained MD simulations were performed using current Martini parameter settings[184] in GROMACS (see appendix, Table A3). The 1000 step minimization was done using the steepest descent method. The MD simulation parameters for non-bonded interactions and neighbor searching were set to a Verlet cut-off scheme with neighbor list updating every 20 steps. Electrostatic interactions were treated with a reaction-field with a 1.1 nm Coulomb cut-off. Van der Waals interactions were

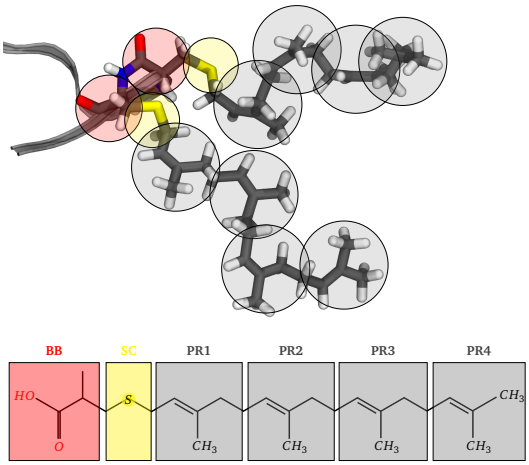


Figure 4.6: Martini bead types used to describe the GG-Cys residue in coarse-grained MD simulations.[156]

subject to a Verlet shifting function and decayed to zero at a 1.1 nm cut-off. The biologically different components of the system, i.e. HVR$^{206-215}$, membrane and solvent, were individually coupled to velocity-rescaling thermostats of 310 K. Pressure coupling and integrator time step were adapted to the respective MD simulation phase. After minimization, the systems were first equilibrated in an NVT ensemble for 0.5 ns with a time step of 10 fs. Subsequently, a 30 ns NPT equilibration with a 20 fs time step was performed using the semi-isotropic Berendsen barostat for pressure control.[148] The 5 µs production runs were performed in an NPT ensemble with a 20 fs time step and pressure coupling to a Parrinello-Rahman barostat.[185]

The topology and parameters of the GG-Cys residue were not part of the Martini force field when this study was set up. Therefore, the cysteine backbone and the thiol group were modeled according to the Martini protein force field,[186] whereas the repeating units of the GG chain were mapped as four C3 beads (Figure 4.6). In accordance with current Martini rules, bond lengths were defined by harmonic potentials and angles by cosine-based functions based on the comparison with the full-atomistic MD simulations. When GG-Cys was introduced into the Rab5 protein model it was treated as a standard cysteine residue to determine the protein secondary structure using the DSSP (define secondary structure of proteins) algorithm.[187] Structural parameters of the coarse-grained GG chains, i.e. bond lengths, bond angles and torsion angles, were found to reproduce the results of the full-atomistic simulations very well (see appendix, Figure B1).[156]

4.4.3 Analysis of peripheral protein-membrane simulations

Bilayer thickness and area per lipid In the present study, the bilayer thickness was calculated as the z-distance between the coordinates of the phosphorous atoms of the phospholipid head groups in the top and bottom layer, respectively. Results from this calculation were found to be very similar compared to the estimation of bilayer thicknesses from electron density profiles.

The area per lipid was calculated as the area of the membrane plane, i.e. the x and y dimensions of the simulation box, divided by the number of lipids in one membrane leaflet. It provided a measure to probe if a membrane system was well equilibrated.

Lipid order parameters The order within the bilayer, or more exactly the acyl chain order, is measured from NMR quadrupolar splitting in experiment.[188] For this purpose one determines the spatial motion of a carbon-hydrogen vector in the lipid acyl chain (Figure 4.7). With the assumption that

both hydrogen atoms are equivalent (i.e. in case of a saturated hydrocarbon chain), the deuterium order parameter S^{CD} is given by

$$S^{\mathrm{CD}} = \frac{1}{3}\langle 2S_{xx} + S_{yy}\rangle \tag{4.1}$$

The order parameter tensors S_{xx} and S_{yy} for the molecular coordinate axes x, y are defined as

$$S_{xx} = \frac{1}{2}(3\cos\theta_x \cdot \cos\theta_x - 1) \tag{4.2}$$

and

$$S_{yy} = \frac{1}{2}\left(3\cos\theta_y \cdot \cos\theta_y - 1\right) \tag{4.3}$$

with θ_x and θ_y being the angle between the bilayer normal and the molecular axis x or y, respectively.

The calculation of lipid order parameters was performed using the approach of Douliez *et al.*[189] for the saturated *sn-1* palmitoyl chains of POPC, PSM, POPE and POPS. A medial acyl chain order parameter was determined by averaging over all order parameters of one palmitoyl chain (carbon atoms C2 to C15).

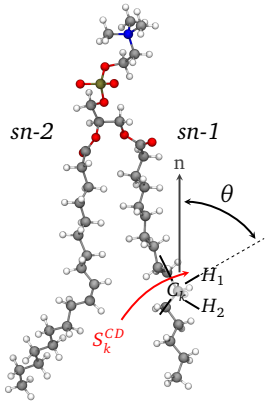


Figure 4.7: POPC lipid and schematic representation of one *sn-1* segment from the acyl chain. The order parameter S_k^{CD} is calculated using the angle θ formed between the bilayer normal $\vec{n}$ and the vector describing the C-H bond.

Lipid orientation Lipid head group orientations for PC, PE, PS, and PSM lipids were calculated as the angle between the vector connecting the phosphate P atom with the head group nitrogen atom and the bilayer normal (Figure 4.8A). Cholesterol ring tilt angles are described by the vector between the C3 and C17 atoms with respect to the membrane normal (Figure 4.8B). PI(3)P orientation is defined by the vector between C1 and C4 atoms in the inositol ring with respect to the membrane normal (Figure 4.8C).

In addition, phospholipid tail orientations were determined by the angle formed between the vector connecting the central glycerol backbone carbon and the last fatty acid carbon atoms of both chains and the membrane normal (Figure 4.8D). The angle $\theta_{A,B,C}$ formed by three atoms A, B, C was calculated as follows:

$$\theta_{A,B,C} = \arccos\left(\frac{\vec{AB}\cdot\vec{BC}}{\|\vec{AB}\|\|\vec{BC}\|}\right) \tag{4.4}$$

Voronoi tessellation The idea behind the Voronoi assignment model is to geometrically subdivide a plane into regions (i.e. cells), and every point of the plane is assigned to its nearest central site of the region. For a detailed introduction into Voronoi diagrams, the reader may be referred to reference,[190]

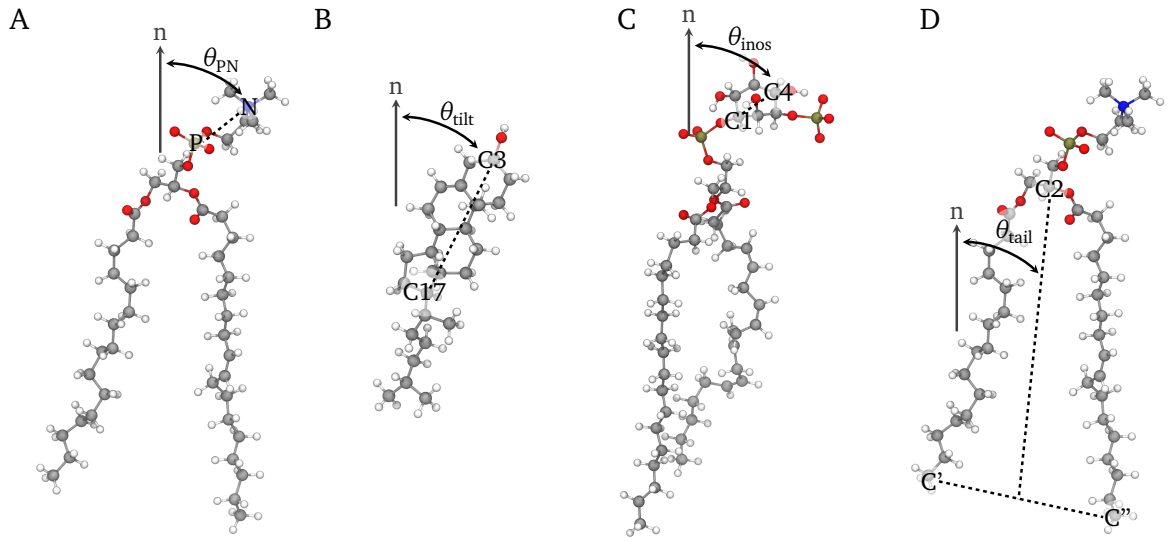


Figure 4.8: θ was defined as the angle between the membrane normal and the corresponding direction vectors describing A) the head group orientation of POPC, POPE, POPS, PSM, B) the tilt angle of cholesterol, C) the inositol ring orientation of PI(3)P, and D) the phospholipid tail orientation of POPC, POPE, POPS, PSM, and PI(3)P.[156]

in a book about computational geometry. Voronoi tessellation was used to assign individual grid cells to each lipid molecule in the membrane top layer. To do so, the phospholipid phosphorous atoms were mapped to their nearest points on a 200x200 grid spanning the average lateral box dimensions, thereby considering the periodical boundaries. The idea is similar to that of the GridMAT-MD (Grid-Based Membrane Analysis Tool) program[191] and used to determine local membrane properties. The local membrane thickness was calculated as the z-distance between each lipid and the closest lipid on the opposite monolayer, which was mapped on the grid. Local acyl chain order parameters were determined for the lipids in the top layer, averaged over the whole acyl chain and mapped onto the grid points. Local properties were averaged over the whole production period and shown as deviations from the global average value. In addition, by evaluating the grid property map, the dependency of the different observables from the cut-off around the GG chains was determined.

Diffusion coefficients The movements of individual particles follow Brownian motion, i.e. the motion's directions are independent from previous directions and random, which means that they can be described by random walk theory. After a sufficiently long time, the position of a particle is normally (Gaussian) distributed with mean zero from the original location.[192] The diffusion coefficient D is derived from the Einstein relation on Brownian motion by considering the number of translational degrees of freedom, DOF, and the current position of the particle at time t, $s(t)$.[188]

$$2\mathrm{DOF} \cdot D = \lim_{t\to\infty} \frac{1}{t} \left\langle (s(t) - s(0))^2 \right\rangle \tag{4.5}$$

In the case of lipids only lateral (in the x, y plane) diffusion is considered which gives DOF=2. Lateral diffusion coefficients are often calculated by analyzing the mean square displacement (MSD) of a molecule. The square displacement (SD) is calculated for the movement of each particle between the current time step and all previous time steps, i.e. the different lag times, while subtracting the motion of the whole layer, which is e.g. the water box or whole membrane in case of water molecules or lipids, respectively. The whole trajectory was divided into blocks of 50 ns for which the MSD and the corresponding diffusion coefficients were calculated. Excluding the first 50 ns the final diffusion coefficient was determined by averaging over the individual block diffusion coefficients. D is determined from the slope of a molecule's MSD averaged over all lag times and the whole set of

particles N in a particular time window:

$$\mathrm{MSD}(t) = \frac{1}{N}\frac{\Delta t}{t_{total}-t}\sum_{i=1}^{N}\sum_{t_0=0}^{t_{total}-t-1}|s(t_0)-s(t_0+t)|^2 \tag{4.6}$$

The expression Δt corresponds to the time between two conformations; the second sum runs over time frames prior to time t, with t_{total} being the total sampling time.[188] The MSDs in a two-dimensional, homogeneous system are in a linear relation to time t and the diffusion coefficient.[193]

$$\langle \mathrm{MSD}(t) \rangle = 4Dt \tag{4.7}$$

However, since MSD leads to averaged diffusion coefficients it gives imprecise results for different particle types or particles with different modes of motion. MSD was also shown to give large deviations of D for small ensembles with <100 particles or short trajectories.[194] In experiment, single-particle tracking is used to track individual molecules and to resolve different motion modes including non-Brownian motion like confined motion, directed motion or anomalous diffusion. The jump distance analysis (JDA) method is based on the same idea, i.e. to track the movements of individual particles throughout the simulation and to identify subpopulations by curve fitting.[195,196] This JDA approach was used in the present study to investigate the diffusion of the protein and the GG anchor. Here, the probability that a particle starting at the origin is found within a shell of radius r and width dr at time t is determined by[197]

$$p\left(r^2,t\right)dr^2 = \frac{1}{4Dt}\exp\left(-\frac{r^2}{4Dt}\right)dr^2 \tag{4.8}$$

In principle, the jump distances (i.e. Euclidean distances) that a particle covered within a given interval $[r,\, r+dr]$ after a specific time are counted, providing a frequency distribution corresponding to the probability distribution function. Fitting equation (4.8) to the measured jump distances provides the diffusion coefficient of a particle ensemble. Fitting the integrated distribution $P(r^2,t)$ is advantageous because it is independent of the bin sizes.[198,199]

$$P\left(r^2,t\right) = \int_0^{r^2} p\left(r^2\right)dr^2 = 1-\exp\left(-\frac{r^2}{4Dt}\right) \tag{4.9}$$

For least-square fitting an initial value of $1\cdot10^{-7}\,\mathrm{cm^2\,s^{-1}}$ for the lateral lipid diffusion was specified. Jump distances were calculated for consecutive frames a certain lag time, at least 1 ns, apart from each other. From all calculated jump distances with lag times ≥5 ns a weighted average depending on the number of considered jumps was determined that gave the final diffusion coefficient.

Structural parameters of the GG chains GG chain structural parameters like bond lengths, bond angles and dihedrals of the backbone carbon atoms were calculated from the full-atomistic MD trajectories and compared with results from QM calculations. The bond angle was calculated according to equation (4.4) (see above). The torsion $\omega_{A,B,C,D}$ between four subsequent carbon atoms A, B, C, D was determined using the normal vectors $\vec{n}_1$ and $\vec{n}_2$ to the planes containing the vectors $\vec{AB}$ and $\vec{BC}$ as well as $\vec{BC}$ and $\vec{CD}$, respectively. An orthonormal frame is formed by the vectors $\vec{n}_1$, $\vec{BC}$ and $\vec{m}$. The torsion was calculated using the inverse tangent of the y- and x-coordinates in the interval $[-\pi,\pi]$ (equation (4.10)).

$$\begin{aligned}
\vec{n}_1 &= \vec{AB}\times\vec{BC}\\
\vec{n}_2 &= \vec{BC}\times\vec{CD}\\
\vec{m} &= \vec{n}_1\times\vec{BC}\\
x &= \vec{n}_1\cdot\vec{n}_2\\
y &= \vec{m}\cdot\vec{n}_2\\
\omega_{A,B,C,D} &= \arctan(y,x)
\end{aligned} \tag{4.10}$$

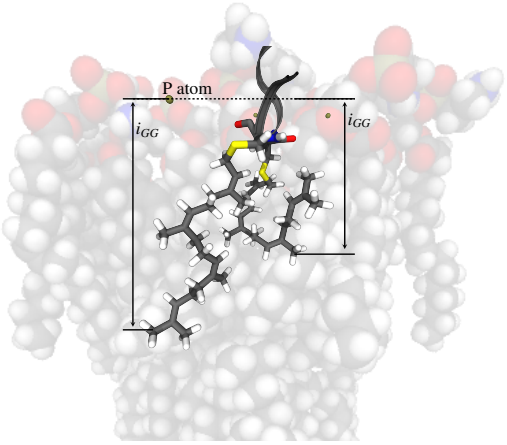


Figure 4.9: The GG chain insertion depth was calculated as the z-distance between the neighboring phospholipid head groups and the lowest carbon atom of the GG chain.[8]

The anchor penetration was defined as the z-distance between the surrounding lipid phosphorous atoms and the GG carbon atom most deeply inserted within the bilayer (Figure 4.9).

Protein RMSD and RMSF The RMSD of atomic coordinates determines the distance between two equal atoms of two structures. Within the scope of protein MD simulations, the RMSD of the backbone alpha carbon (Cα) atoms at each time step t is often calculated relative to the first frame structure at $t_0 = 0$. Thus, the RMSD serves as a measure for the structural stability of a macromolecule over time.

$$\mathrm{RMSD}(t) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(s_i(t) - s_i(t_0))^2} \tag{4.11}$$

In equation (4.11), N is the number of Cα atoms, $s_i(t)$ the position of a particular atom i at a certain time step t, and $s_i(t_0)$ the position of the same atom of the reference structure at time zero. Prior to the calculation of RMSD values the trajectory protein structures were aligned to the first frame reference structure in order to exclude translation-based structural deviations.

The root mean square fluctuation (RMSF) determines the average atomic mobility of the Cα atoms during an MD simulation. Similar to the RMSD calculation an alignment of the protein trajectory to the reference structure is required. The RMSF is calculated for each atom i over the entirety of all frames t_{total}:

$$\mathrm{RMSF}(i) = \sqrt{\frac{1}{t_{total}}\sum_{t=1}^{t_{total}}(s_i(t) - \overline{s_i})^2} \tag{4.12}$$

In the present study, the structure alignment was done using the VMD measure fit command or the MDAnalysis rms_fit_trj built-in function. MDAnalysis was also used to calculate the RMSD, whereas the RMSF was determined with the VMD measure rmsf command.

Radius of gyration and protein orientation The protein radius of gyration r_{gyr} is described as the distribution of atoms around the protein axis and is calculated as the atoms mass-weighted root mean square distances from the protein center of mass COM, with m_i being the mass of the i-th atom (equation (4.13)). It therefore characterizes the compactness of a protein structure.

$$r_{\mathrm{gyr}}(t) = \sqrt{\frac{\sum_{i=1}^{N} m_i (s_i(t) - \mathrm{COM}(t))^2}{\sum_{i=1}^{N} m_i}} \tag{4.13}$$

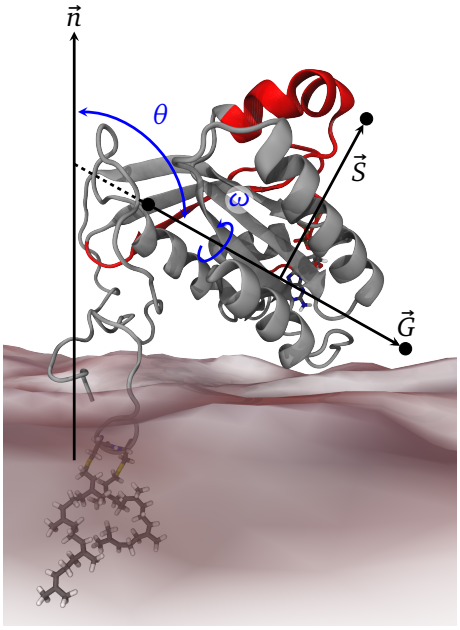


Figure 4.10: The Rab5 orientation at the membrane is described by the pivot angle θ which characterizes the G domain orientation $\vec{G}$, as well as the dihedral ω as a measure for the torsion between membrane normal $\vec{n}$, $\vec{G}$, and the switch regions $\vec{S}$.[8]

The radius of gyration was calculated for all protein atoms in each frame using the VMD command measure rgyr.

For the evaluation of protein position and orientation relative to the membrane, the protein-membrane distance was measured as the z-distance between the amino acid centers of mass and the P atoms of the phospholipids. Furthermore, the protein tilt angle θ and the torsion ω between membrane, G domain and switch regions was determined (Figure 4.10). Tilt angle and torsion were calculated analogously to the bond angle and torsion of the protein anchor GG chain using equations (4.4) and (4.10).

Electrostatic potential maps by using the Adaptive Poisson-Boltzmann Solver (APBS) The protein electrostatic properties were calculated using the PDB2PQR version 2.0.0 code[200,201] and the Adaptive Poisson-Boltzmann Solver (APBS) software package version 1.4.1.[202]

The PDB file was converted into a PQR file with PDB2PQR. PQR files have the same format as PDB files, while the columns for occupancy and B-factors are replaced by per-atom charge and radius. The conversion was done using CHARMM27 force field parameters and pK_a values were predicted via PROPKA[203] at pH 7 in order to assign protonation states. Besides the PQR file, PDB2PQR generates an APBS input file containing parameters for the polar solvation calculation. The APBS was used to calculate the protein charge distribution and to generate the electrostatic potential map.

Solvent-accessible surface area (SASA) and lipid contacts In order to characterize the solvent exposition of the whole protein or particular regions the SASA was calculated using the VMD command measure sasa. A solvent probe radius of 1.4 Å was used. The absolute SASA values for the different amino acid residues were normalized by the experimentally derived values for a Gly-X-Gly tripeptide with X being the corresponding amino acid by Miller *et al.*[204] For the non-standard GG-modified cysteine residue, a separate Gly-GG-Cys-Gly tripeptide with a completely stretched GG chain was solvated, energy minimized, and the resulting SASA was used for normalization.

Dihedral principal component analysis (dPCA) Principal component analysis (PCA) is a well-known tool to reduce the dimensionality of a complex system. It is applied in MD trajectory analysis to extract the free energy surface of a molecule as a function of only a few coordinates.[205] Here, PCA was applied to the Rab5 HVR[206–215] to identify membrane composition-dependent stable conformations.

Instead of using Cartesian coordinates for the PCA it has been suggested by Mu *et al.*[206] to use protein backbone dihedral angles. This approach is called dihedral angle principal component analysis (dPCA) and benefits from the fact that these internal coordinates do not suffer from an insufficient separation of overall and internal motion.[206–208] In dPCA, each protein backbone dihedral angle ψ and ϕ was sine and cosine transformed to give $2N$ variables q_n, with N being the number of dihedral angles (Figure 4.11).

The covariance matrix was formed using the mean-centered variables q_n, with each matrix element reflecting the relationship between two of the internal coordinates in their variation from the mean. Diagonalization of the covariance matrix resulted in a set of eigenvectors $\mathbf{v}^{(k)}$ and corresponding eigenvalues λ_k. The eigenvectors were ranked in descending order of their eigenvalues giving the order of significance. That is, the eigenvector with the highest eigenvalue is the principal component of the data set. Dimensionality reduction is achieved by ignoring the eigenvectors with lesser significance. The principal components V_k of the data set q_n were calculated as follows.[207]

$$V_k = \mathbf{v}^{(k)} \cdot q_n = y_1^{(k)} \cos\varphi_1 + y_2^{(k)} \sin\varphi_1 + \ldots + y_{2N-1}^{(k)} \cos\varphi_N + y_{2N}^{(k)} \sin\varphi_N \tag{4.14}$$

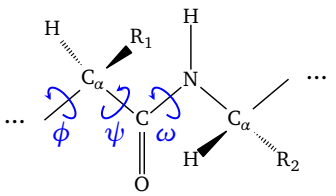
The principal component coefficients $y_1^{(k)}$ and $y_2^{(k)}$ of the eigenvector $\mathbf{v}^{(k)}$ are a measure for the influence $\Delta_1^{(k)}$ of the corresponding dihedral angle φ_1 on that particular k^{th} principal component, which is defined as

$$\Delta_1^{(k)} = \left(y_1^{(k)}\right)^2 + \left(y_2^{(k)}\right)^2 \tag{4.15}$$

When using only the first two eigenvectors the dimensionality is reduced to two and the free energy surface along the first two principal components (V_1, V_2) was determined with equation (4.16).

$$\Delta G\,(V_1, V_2) = -k_B T\,[\ln \rho\,(V_1, V_2) - \ln \rho_{max}] \tag{4.16}$$

The probability density function $\rho\,(V_1, V_2)$ was derived from the data distribution histogram. The density maximum ρ_{max} was subtracted in order to set the free energy minimum to zero.[207] The diagonalization of the covariance matrix was done using the MATLAB built-in pca function, which returns the principal component coefficients, the matrix of principal components (i.e. the eigenvectors) and the eigenvalues.



Protein backbone angles:

$\{\varphi_n\} = \{\psi_1, \phi_2, \psi_2, \phi_3, ..., \psi_{N-1}, \phi_N\}$

sin- and cos-transformation:

$q_{2n-1} = \cos\varphi_n$
$q_{2n} = \sin\varphi_n$
$n = 1, \ldots, N$, with N the number of protein dihedral angles

Covariance matrix:

$\sigma_{ij} = \left\langle (q_i - \langle q_i \rangle)(q_j - \langle q_j \rangle) \right\rangle$

Figure 4.11: The protein conformation depends on the rotatable backbone dihedral angles ψ and ϕ. In dPCA these dihedrals are sine and cosine transformed, providing internal coordinates that allow constructing the free energy surface of the investigated molecule. In order to do so, the covariance matrix formed by the transformed variables is diagonalized. The resulting eigenvectors and eigenvalues represent the collective motion modes and the corresponding principal components give the free energy surface.

Calculation of the potential of mean force (PMF) US runs were performed in order to calculate the free energy profile of the GG chain release from different membranes. The distributions from the individual windows were combined and analysed using the WHAM implementation version 2.0.9 of Alan Grossfield. This code is based on the procedure and equations published in the papers of Shankar Kumar[209] and Benoit Roux.[210] Input parameters are the boundaries of the combined windows histogram, the number of histogram bins (i.e. which specify the number of points in the final PMF), the temperature as well as the metadata file including the path to the NAMD output 'colvar' trajectory file. Each line in the metadata file corresponds to one umbrella sampling window. Furthermore, the biasing potential spring constant which was used in the umbrella sampling simulation is provided; here, the spring constant was $4\,\mathrm{kcal\,mol^{-1}\,\AA^{-2}}$.

The WHAM output file contains the position along the reaction coordinate in the first column and the corresponding free energy in the second column. It is followed by the statistical uncertainty of the free energy. Moreover, the probability distribution and its statistical uncertainty are provided. The statistical uncertainty computed using Monte Carlo bootstrap error analysis is based on the standard deviation of a specific number of randomly generated fake data sets. In the present study, the WHAM output file was used in MATLAB to plot the free energy as a function of the reaction coordinate. Furthermore, from the NAMD colvar trajectory the overlap of the individual umbrella sampling window distributions was checked.

4.4.4 The Rab5 C-terminal hypervariable region

Rab5 is a peripheral membrane protein associated with the membrane via a long and flexible HVR which includes the GG anchor (for the protein sequence see section 2.1.1, Figure 2.2). In the three model membrane systems, i.e. pure POPC, the ternary and six-component membranes, the $\mathrm{HVR}^{206-215}$ conformation and its influence on the membrane lipids were investigated. Data were gathered from the last 200 ns of each MD simulation. The evolution of the global membrane properties thickness and lipid area within this 200 ns demonstrate sufficient equilibration of the systems (see appendix, Figure B2 and Figure B3).

In the present study, local membrane property deviations within a cutoff radius of 0.5 nm around the GG anchor were analyzed using Voronoi tessellation. Local differences of the lipid acyl chain order parameters in proximity to the $\mathrm{HVR}^{206-215}$ (highlighted by black crosses) are shown in Figure 4.12. Over the whole membrane the local lipid order parameters varied by up to ±20 % with the deviations being most prominent in pure POPC. Independent of the lipid composition the acyl chain order was reduced in vicinity to the GG anchor.

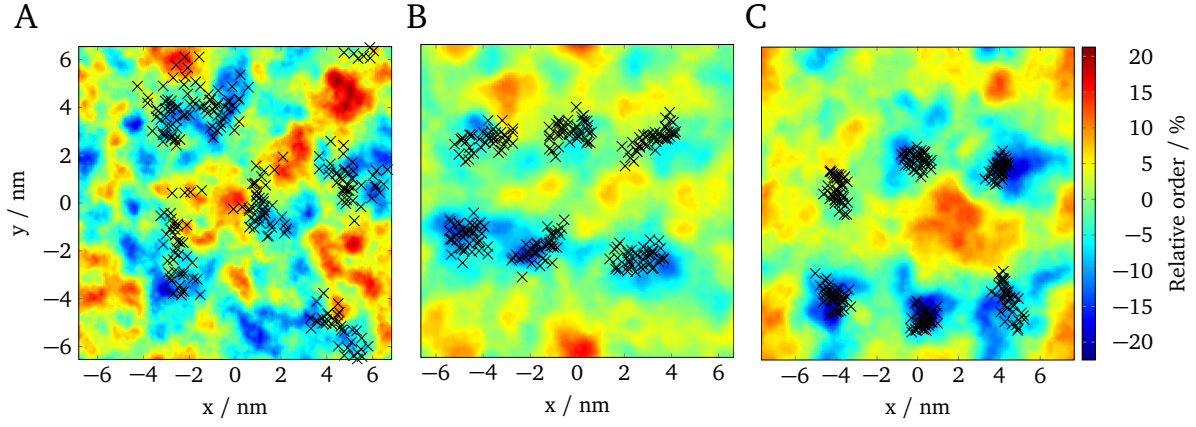


Figure 4.12: Local lipid order deviations in A) pure POPC, B) the ternary mixture, and C) the six-component membrane. Black crosses show the positions of the GG anchors during 200 ns of MD simulation.[156]

Apart from the lipid order parameters, the membrane thickness, lipid head group orientation and acyl chain orientation in a cutoff radius of 0.5 nm around the GG anchor were investigated (Figure 4.13). Data were averaged over the six $HVR^{206-215}$ replicates in each bilayer, respectively.

In close proximity to the lipid anchor the membrane thickness was decreased by up to 2 %. This was a consequence of the decrease in lipid acyl chain order. Obviously, highly ordered lipids exhibit fully extended acyl chains, leading to an increase in membrane thickness in the corresponding region. On the other hand, introducing disorder caused a tilting of the chains and thus a decrease in membrane thickness. Furthermore, close to the GG anchors a decrease of the head group orientation angle by up to 6 % was observed. That is, the head groups were less flexed towards the membrane plane. In contrast, the acyl chain orientation angle was up to 14 % larger around the $HVR^{206-215}$, i.e. the lipid chains were more tilted compared to the global average. As mentioned before, this was directly related to the acyl chain order decrease. To sum up, all local deviations from the global properties were related to the reduction of lipid order parameters around the GG anchor and thus, internally consistent.

The local deviation's magnitude depended on the membrane composition (Figure 4.13). In the first place, this was due to the different cholesterol concentrations of the investigated model membranes. Cholesterol functions as a fluidizer of tightly packed sphingolipids in the solid gel phase.[101,111] Its fluidizing effect is exerted on lipids below their melting temperature, i.e. the temperature T_m at which the gel to liquid-crystalline phase transition takes place, when the lipid fatty acid chains are fully extended. Intercalation of cholesterol leads to a thinning of the membrane due to the introduction of gauche conformers, i.e. the membrane is converted into the liquid-ordered, lo, phase. Heating above T_m introduces gauche conformers as well, leading to the liquid-crystalline phase also called liquid-disordered, ld, phase. When cholesterol is added to lipids in the ld phase, it has a condensing effect, thereby reducing the amount of gauche conformers.[211] In the present study, the cholesterol condensing effect was observed when comparing the global thickness, lipid area and lipid order parameter of the three investigated model membranes (see appendix, Figure B2, Figure B3, and Figure B4). Compared to pure POPC, cholesterol insertion in the other bilayers caused significant increases in global membrane thickness and lipid acyl chain order and a decrease in the area per lipid. Another consequence of the cholesterol condensing effect was a slowdown of the lateral lipid diffusion. In pure POPC, the diffusion determined by JDA was $1.36 \cdot 10^{-7}\,\mathrm{cm^2\,s^{-1}}$ (Table B2) which is in very good agreement with experimental NMR studies.[212] In the cholesterol-enriched ternary and

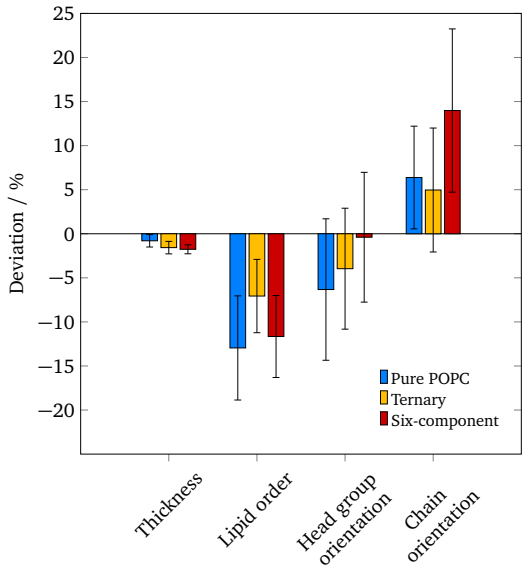


Figure 4.13: Local percentage variance (cutoff radius = 0.5 nm) of different membrane properties from the global average in the three investigated membrane models.

six-component membranes, POPC diffusion was $0.78 \cdot 10^{-7}\,cm^2\,s^{-1}$ and $0.74 \cdot 10^{-7}\,cm^2\,s^{-1}$, respectively; that is a reduction of approximately 43 %. Among the different lipids, POPC diffusion was fastest and cholesterol diffusion was slowest. However, the spectrum of diffusion coefficients was narrow ranging from $0.59 \cdot 10^{-7}\,cm^2\,s^{-1}$ to $0.74 \cdot 10^{-7}\,cm^2\,s^{-1}$ in the six-component membrane.

Compared to adding cholesterol to a bilayer, insertion of the GG anchor had the opposite and quantitatively smaller local effect on the directly surrounding phospholipids. It decreased lipid order accompanied by a modest thinning of the membrane. The disordering effect of the GG anchor was most pronounced in pure POPC which was not 'pre-ordered' by cholesterol, but less prominent for the cholesterol-stabilized phospholipids in the ternary mixture and the six-component membrane. In contrast to integral membrane proteins, which significantly alter the surrounding lipid structure, the influence of peripheral membrane proteins on the lipid environment was rather small. Previous MD simulations as well as experimental Fourier transform infrared (FTIR) spectroscopy and solid-state NMR showed that global properties of a DMPC lipid bilayer are hardly influenced by the insertion of an N-Ras (neuroblastoma rat sarcoma) C-terminal peptide.[169–171] The present results agree well with findings from other groups who found a slight thickness decrease in pure DMPC after insertion of a Ras peptide.[170] However, comparison of the six $HVR^{206-215}$ replicates induced effects showed apparent discrepancies within one membrane. These may arise from different peptide conformational states which have been studied by various techniques, e.g. dihedral principal component analysis (dPCA). Dihedral PCA allows to examine the combination of dihedral angles in the $HVR^{206-215}$ by constructing its free energy landscape, thus characterizing the peptide's overall structure and dynamics.[206–208] First, the $HVR^{206-215}$ free energy landscape in pure POPC was investigated to rule out membrane composition derived structural variations. Figure 4.14 shows the free energy surfaces of the six $HVR^{206-215}$ replicates along the first two eigenvectors ($V1$, $V2$) in pure POPC.

The first two eigenvectors explained approximately 45 % of the peptides' structural variance. In pure POPC, where structural variations caused by different membrane compositions could be excluded, dPCA revealed several distinct peptide conformational states. This indicated a great structural flexibility of the $HVR^{206-215}$, whose secondary structure was hardly stabilized by intramolecular hydrogen bonds or membrane interactions. Besides the distinct conformational states resolved by dPCA, the peptide radius of gyration and RMSF confirmed the enormous structural flexibility of the $HVR^{206-215}$ in POPC (Figure 4.15). Along with that, the GG anchor insertion depth, which was assumed to influence the extent of local membrane distortions, varied between the six peptide replicates. However, a direct correlation between the insertion depth and the reduction in lipid order in close proximity to the GG anchor was not observed. The individual anchor chains were highly dynamic, revealing coordinated anti-correlated motions such as one chain was deeply inserted within the membrane while the other was bend towards the membrane surface.[156] This is in agreement with ^2H solid-state NMR studies of lipidated Ras peptides uncovering that the lipid modifications undergo extensive motions which increase the peptide entropy, thus decreasing the protein free energy at the membrane.[213]

Investigation of the SASA of the HVR replicates in pure POPC as well as the number of neighboring lipids supported the great structural diversity of the peptides (see appendix, Figure B6).

Considering the highly flexible structure of the $HVR^{206-215}$ in pure POPC, the question arose if the membrane composition may influence the peptide dynamics and possibly stabilize one single or a few conformational states. Therefore, peptide properties like the GG anchor insertion depth, the peptide radius of gyration and the SASA were compared for the three model membranes (see appendix, Figure B10). The average GG chain insertion depth in each bilayer was influenced by the absolute thickness of the membrane and was about 38 %-39 % of the respective bilayer thickness.[156] A term often used in this context is that of the 'hydrophobic mismatch' which describes the length difference between the hydrophobic anchor chains and the hydrophobic bilayer core. In case of the Rab5 HVR, hydrophobic mismatch between the GG chains and the lipid tails was mostly reduced by the adaptation of the anchor properties, i.e. insertion depth and order parameter, instead of adapting the surrounding lipids' z dimension. Thus, the only marginal effect of the GG anchor on

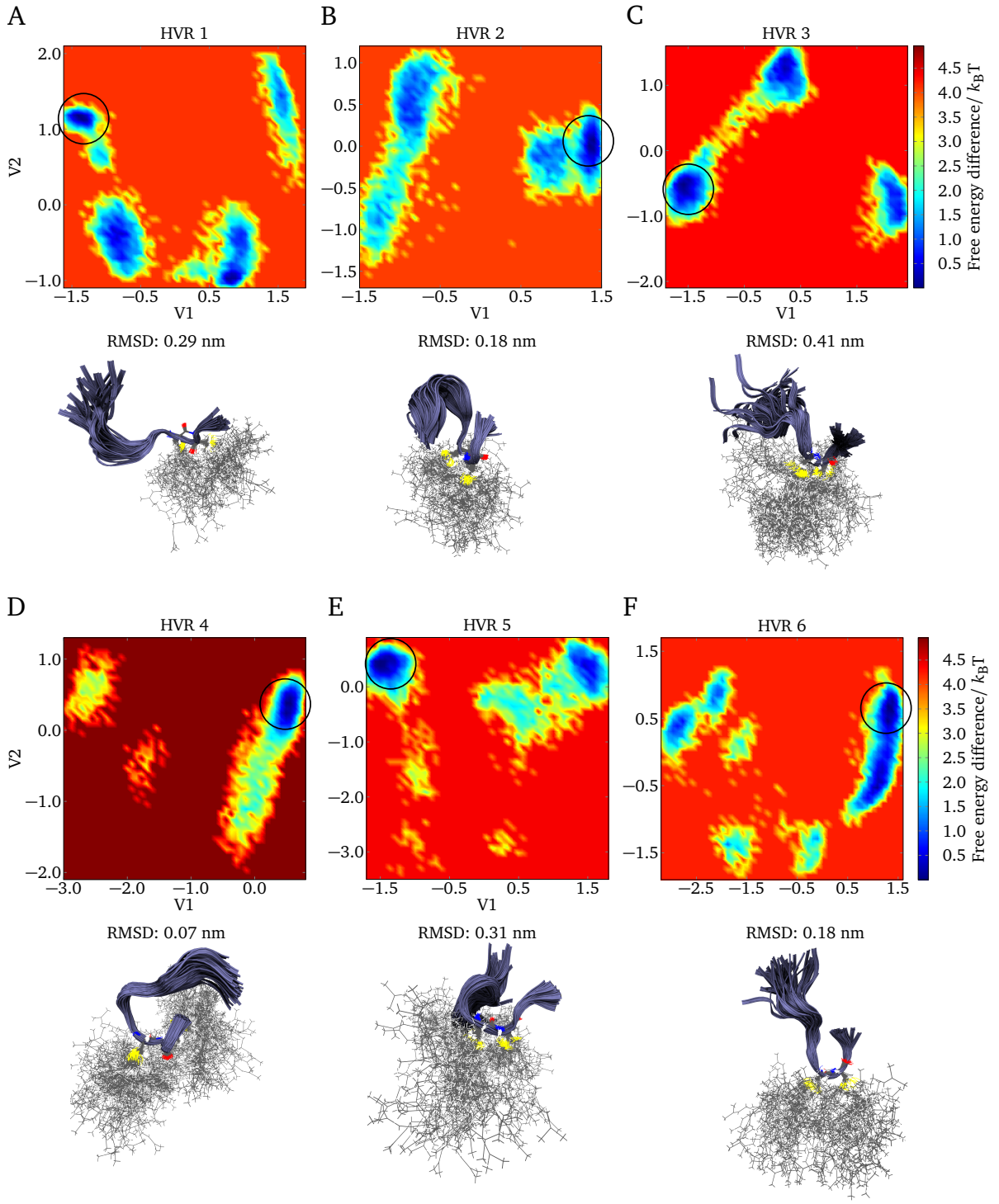


Figure 4.14: Dihedral PCA free energy landscapes of the six Rab5 HVR$^{206-215}$ replicates (A-F) in pure POPC plotted as a function of the first two principal components $V1$ and $V2$. Lower panel: The lowest energy structures and corresponding RMSD of one minimum energy conformation (circle).[156]

the surrounding lipid environment may be explained by the fact that in particular the GG anchor itself adapts to the lipid chains length and not vice versa.[169,170,214] In agreement with that, relatively low order parameters have been found for the lipid modifications of Ras in ^{2}H solid-state NMR studies.[215] Furthermore, these studies together with ^{1}H magic-angle spinning NOESY (nuclear Overhauser effect spectroscopy) experiments demonstrated that the Ras backbone created a free volume within the membrane lipid water interface where hydrophobic amino acid side chains and the lipid

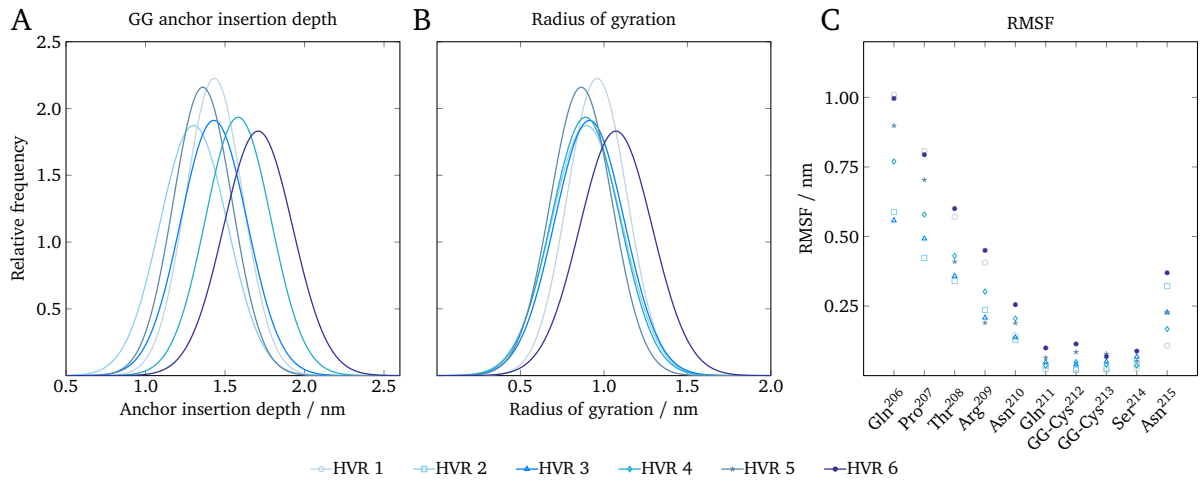


Figure 4.15: Distribution of A) the GG chain insertion depths and B) the peptides radii of gyration of the individual $HVR^{206-215}$ replicates in pure POPC. C) The RMSF of the peptide backbone Cα atoms is another measure for the HVR flexibility.

modifications are allowed to move freely.[169] As a measure for the $HVR^{206-215}$ structural flexibility, the peptide radius of gyration was determined in all membrane models (appendix, Figure B10). It was highest in the physiological six-component membrane indicating a greater HVR conformational freedom compared to the peptides in the uncharged bilayers. This was confirmed by the absolute SASA which suggested that the peptide residues were most water exposed in the negatively-charged six-component membrane and thus, more mobile.

In order to explore a potential membrane composition-dependency of the peptide conformational space, all $HVR^{206-215}$ replicates in pure POPC, the ternary or the six-component membrane, respectively, were subjected to one dPCA yielding a total sampling time 1.2 µs (6 x 200 ns). The corresponding free energy landscapes revealed 4 to 5 energy minima in each membrane system (Figure 4.16A). The energy barriers between different minima were highest in the ternary mixture and lowest in the physiological six-component membrane. The influence of the HVR dihedral angles on the principal components V1 and V2 is shown in Figure 4.16B.

The first two eigenvectors were equally influenced by almost all HVR dihedral angles in pure POPC and in the ternary mixture. In contrast, the peptide structures in the six-component membrane were more strongly affected by backbone torsions from the N-terminal residues Pro^{207}-Asn^{210}. Especially the impact of the Arg^{209} torsion is striking, indicating a particular role in $HVR^{206-215}$ dynamics in the negatively charged six-component membrane. Arg^{209} is the only positively charged amino acid within the $HVR^{206-215}$ peptide. Thus, its interactions with the bilayer lipids were investigated in detail. Representative configurations of the HVR-lipid interactions in different bilayers are displayed in Figure 4.17. Not surprisingly, Arg^{209} formed electrostatic contacts with the phospholipid phosphate groups in pure POPC and the ternary mixture. Interestingly, in the negatively charged six-component membrane electrostatic interactions were predominately formed between Arg^{209} and the protruding phosphate groups of the PI(3)P inositol ring. Compared to the uncharged bilayers Arg^{209} was prevented to insert deeply into the membrane but rather remained in the water-lipid interface. This is the reason for the greater peptide conformational flexibility in the six-component membrane. Further comparing investigations of the HVR structure and dynamics are given in the appendix, e.g. residue-based SASA and lipid neighbors (Figure B7), RMSF (Figure B8), peptide-membrane distances as a function of time (Figure B9).[156] Detailed information about the exact lipid composition around each $HVR^{206-215}$ replicate and local properties are listed in the appendix, Table B4.

In order to investigate the impact of the different Arg^{209}-phospholipid binding modes on the peptide anchoring, the HVR membrane binding Gibbs free energies for the three model membranes were determined. For each membrane the peptide bilayer-to-water transition of three $HVR^{206-215}$ repli-

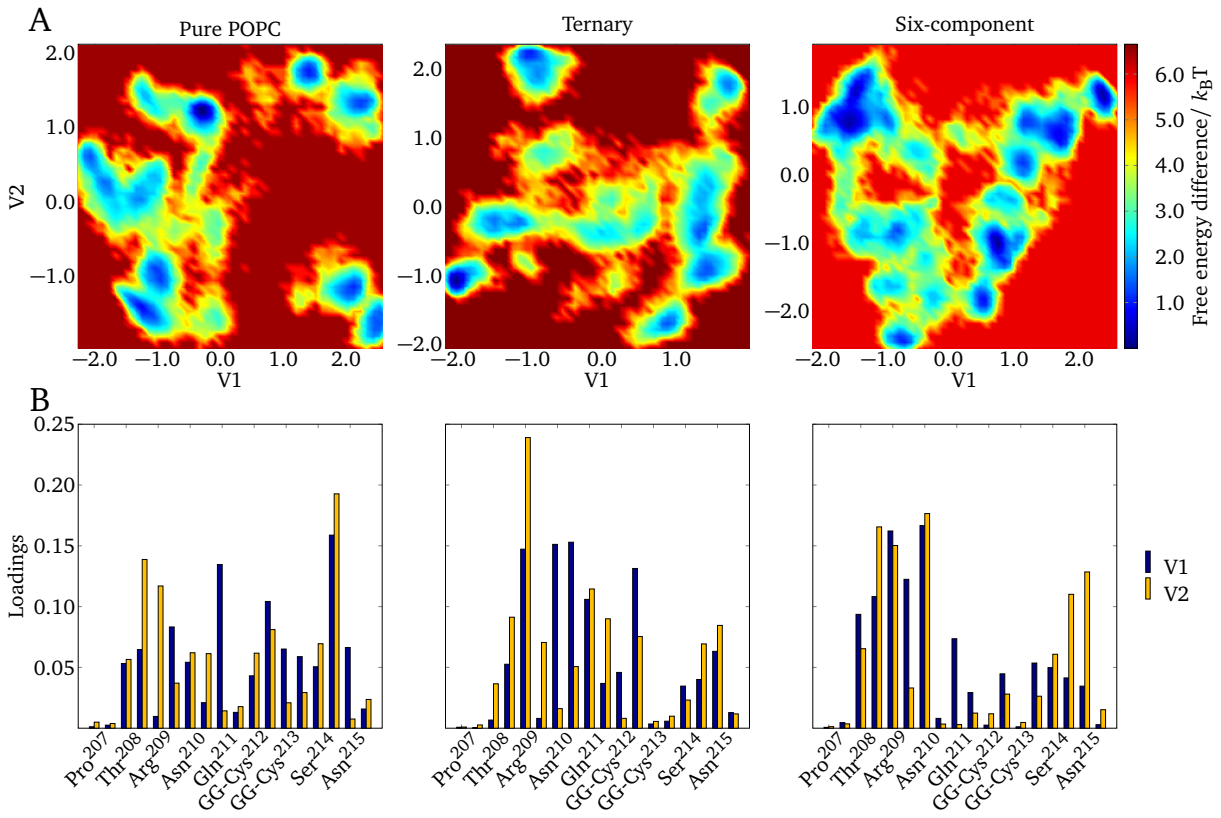


Figure 4.16: A) Dihedral PCA free energy landscapes of all Rab5 HVR$^{206-215}$ replicates in each membrane system plotted as a function of the first two principal components V1 and V2. B) The loadings are the linear combinations coefficients of the sin- and cos-transformed dihedral angles forming V1 and V2 that characterize the influence of the peptide dihedral angles on the first two principal components.

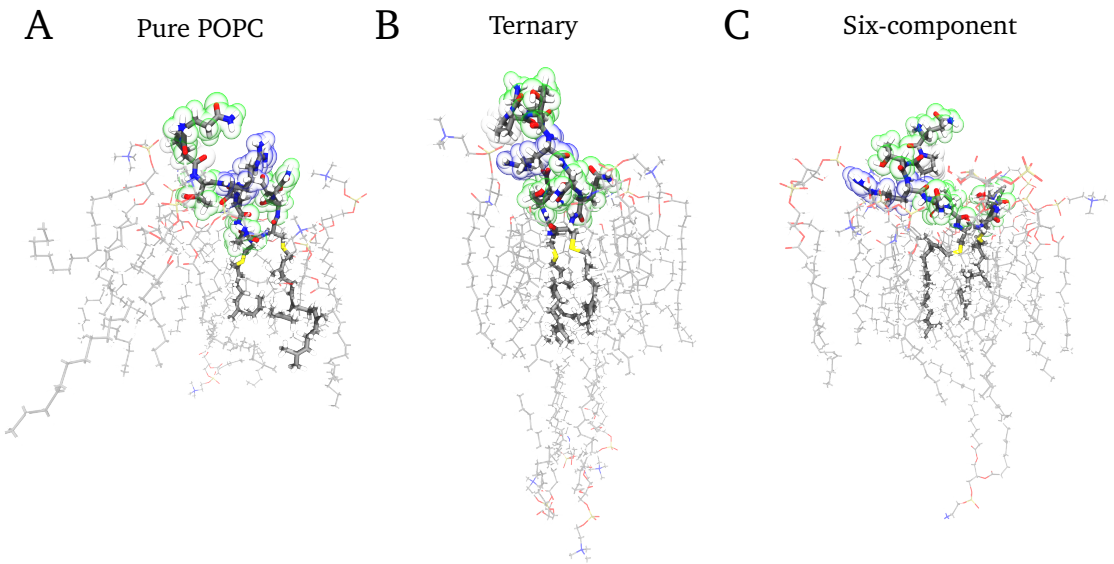


Figure 4.17: Selected configurations of the HVR$^{206-215}$ in A) pure POPC, B) the ternary mixture and C) the charged six-component membrane. Polar residues are represented by green spheres, the basic arginine is colored in blue.

cates was monitored and the resulting average extraction energy profile is given in Figure 4.18. The minimum energy structure corresponded to a peptide configuration with the GG chains completely inserted into the bilayer hydrophobic interior. During extraction the peptide Gibbs free energy rose until all amino acid residues and the peptide's GG chains were fully solvent exposed. For the uncharged pure POPC bilayer and the ternary mixture, the energy profiles reached plateaus at approximately 116 kJ mol^{-1} and 119 kJ mol^{-1}, respectively, for z distances larger than 1.6 nm.[156] Thus, the presence of cholesterol within the membrane did not affect the transition free energy difference. However, for the charged six-component membrane the maximum energy plateau was reached at 124 kJ mol^{-1} for distances larger 2.8 nm.

This was due to a stretching of the peptide backbone to ensure a continuous interaction between the Arg^{209} and the inositol phosphate group of PI(3)P. The interaction between Arg^{209} and the lipid head groups was disrupted when the HVR backbone was in an almost fully extended. Contacts between the peptide and the bilayer lipids were formed by long-range electrostatic interactions and short-range hydrogen bonds between Arg^{209} and PI(3)P. Apparently, electrostatic interactions between basic Arg^{209} and the zwitterionic phospholipid phosphate groups were not established, probably because the phosphates were shielded by the partly bulky lipid head groups. The strong deviation between the individual replicates in the six-component membrane may be explained by the large variance in the composition of the surrounding lipids. The presence of nearby PI(3)P lipids led to the formation of various contacts, which improved membrane binding and caused a larger Gibbs free energy difference of the peptide-to-water transition. HVR$^{206–215}$ replicates which were not located in close proximity to PI(3)P lipids showed a similar energy profile compared to those in the uncharged bilayers. Prior to the HVR calculations, simulations with a GG-modified tripeptide were performed. The results obtained for this peptide could directly be compared to experimental findings and thus, served to verify the Gibbs free energy differences obtained for the GG chain membrane release by umbrella sampling. Those early experiments were performed by Silvius *et al.* who investigated the insertion of an either C-terminally methylated or demethylated tripeptide in a POPC/POPE (9:1) bilayer.[216,217] The free energy change upon extraction from POPC using umbrella sampling was 69 kJ mol^{-1} or 61 kJ mol^{-1} (see appendix, Figure B11) for the methylated or demethylated peptide, respectively, compared to the experimental values for the reverse insertion process of −50 kJ mol^{-1} and −44 kJ mol^{-1}. Thus, the simulation results showed a slight overestimation of absolute binding free energies. However, the overall trend of free energies was correctly predicted. Consequently, the umbrella sampling simulations of the present study that elucidate membrane composition-dependent

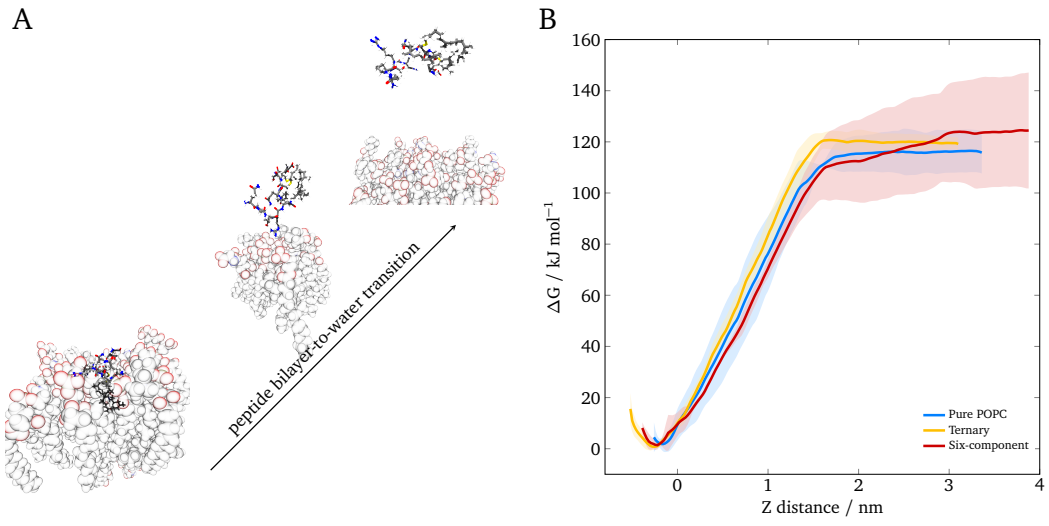


Figure 4.18: A) Procedure of the GG chain extraction and peptide removal from the six-component membrane. B) Bilayer-to-water transition free energy of the GG-modified HVR in the three investigated model membranes. Data are averaged over three single HVR replicate simulations.

relative free energies are expected to perform very well. Inaccuracies in the determination of binding free energies from simulations are often caused by convergence issues and sampling errors. For solute insertion into a bilayer it has been shown that insufficient sampling significantly influences Gibbs free energy differences.[218,219] Special attention should be paid to the initial starting conformation and an adequate equilibration time per umbrella window which may introduce systematic errors. Studies have shown that simulation times >70 ns resulted in converged PMF profiles for long amphiphiles interacting with membranes.[220] This comparably long equilibration time was necessary to allow the membrane to reorganize after the extraction process. In the present study, extraction of the GG chains from the model membranes was expected to cause only small structural perturbations and lipid reorganization was observed during the bilayer-to-water transfer process. Furthermore, since the study aimed to investigate Gibbs free energy differences between differently composed bilayer systems, it was not mandatory to determine accurate absolute binding free energies. Here, systematic errors introduced by shorter simulation times are assumed to similarly influence all membrane models and did not affect relative energy differences.[156]

As seen above, analysis of lateral lipid diffusion coefficients revealed values of approximately $0.6 \cdot 10^{-7}\,\mathrm{cm^2\,s^{-1}}$ in the ternary and six-component membranes (see appendix, Table B2 and Table B4). Conversion into units more suitable for the peptide-membrane system yielded an average diffusion coefficient of $6 \cdot 10^{-3}\,\mathrm{nm^2\,ns^{-1}}$, that is, a lipid covered an area of 1.5 $\mathrm{nm^2}$ during the complete 250 ns MD simulation. Thus, lipid diffusion coefficients were too small to allow free mixing of the lipids within the bilayer. This led to the observation that the local membrane composition in close proximity to the $\mathrm{HVR^{206-215}}$ replicates was unchanged during the 250 ns full-atomistic MD simulation and predefined by the initial setting. Since the local lipid composition was crucial for peptide-membrane interactions as demonstrated for the relative binding free energies, 5 µs coarse-grained MD simulations of one membrane-anchored $\mathrm{HVR^{206-215}}$ in different model membranes were performed to obtain insight into the long-term evolution of the systems. The membrane composition within a distance cutoff around the peptide averaged over the 5 µs is shown in Figure 4.19A-C.

In all membrane models, the amount of cholesterol was highest in close proximity to the peptide and converged to the average value for distances >7 nm. In the six-component membrane, the signaling lipid PI(3)P was also significantly enriched around the peptide. Compared to average the PI(3)P ratio increased by a factor of 4 within 1 nm distance around the peptide in the six-component membrane. The other lipids were less frequently found in close neighborhood to the $\mathrm{HVR^{206-215}}$. Initially, PI(3)P lipids started from a random distribution but enriched significantly in close vicinity to the peptide within 5 µs (Figure 4.19D). The enrichment of PI(3)P close to the $\mathrm{HVR^{206-215}}$ was reproduced in two further independent simulations which started with different initial lipid distributions of the same composition. Since the simulations yielded also transient phases where no PI(3)P lipid was found close to the peptide, an artificial clustering due to an overestimation of Martini interactions could be excluded. Interestingly, an accumulation of the likewise negatively charged POPS lipid was not observed. This implies that an electrostatically-driven recruitment of PI(3)P included the formation of interactions exclusively between the PI(3)P inositol ring phosphates and the positively charged $\mathrm{Arg^{209}}$. Apparently, the phosphate groups of the zwitterionic lipids were too deeply buried within the membrane to be part of strong electrostatic interactions that could induce an active recruitment towards the HVR. Protein clustering with certain lipid types in microdomains is a common motif in membrane-associated signaling. Examples are the recruitment of negatively charged lipids by the HIV-1 Gag polyprotein precursor nucleocapsid to improve lipid binding[221] or the attraction of cholesterol and anionic $\mathrm{PIP_2}$ by the cytokine receptor glycoprotein 130.[222] Ionic protein lipid interactions have also been observed for the SNAP receptor protein syntaxin-1A which forms microdomains with $\mathrm{PIP_2}$ to improve membrane fusion.[223] The other lipid recruited to the $\mathrm{HVR^{206-215}}$ in the present study was cholesterol. Binding of cholesterol to transmembrane proteins is described in several cases, e.g. to the transmembrane domain of the $\mathrm{Na^+K^+}$ATPase[224,225] and to G-protein coupled receptors like human β_2 adrenergic receptor,[226,227] human $\mathrm{A_{2A}}$ adenosine receptor[228] and bovine rhodopsin.[229,230] The protein-cholesterol interaction is found to influence the conformation of those transmembrane proteins by increasing their stability and thus, preserve protein function.[231] Apart from that, high

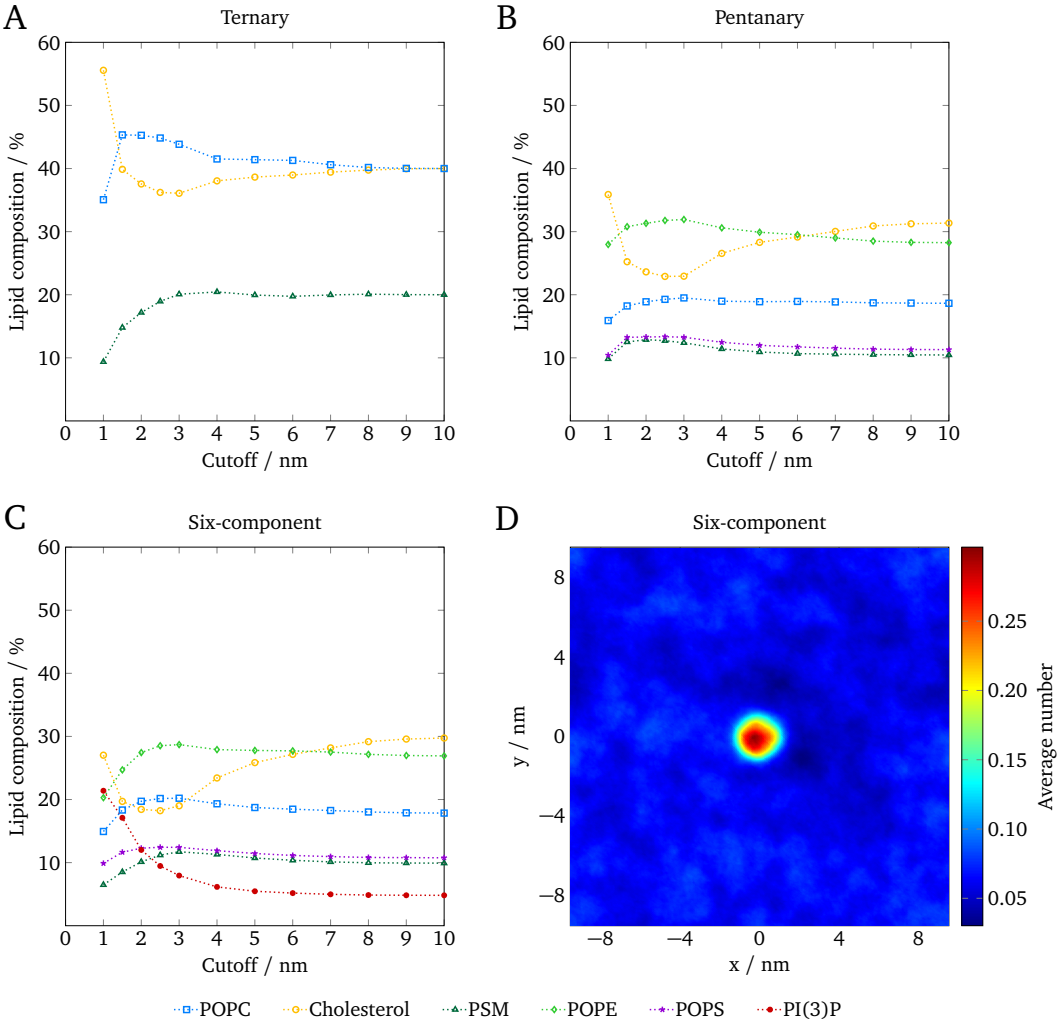


Figure 4.19: Lipid composition within a certain distance cutoff around the HVR$^{206-215}$ in A) the ternary mixture, B) the pentanary membrane, and C) the six-component membrane (dotted lines to guide the eye). D) Top view on the six-component membrane with the HVR centered showing the distribution of PI(3)P averaged over 5 µs of coarse-grained MD simulation.

levels of cholesterol increase the lipid order, thereby reducing the conformational flexibility of the protein. This effect was also observed for the HVR$^{206-215}$ replicates, since diffusional analysis of the full-atomistic simulations yielded reduced peptide diffusion coefficients in the six-component membrane compared to pure POPC and the ternary mixture (see appendix, Table B3). Surprisingly, peptide diffusion was slowest in the six-component membrane, although the cholesterol ratio was higher in the ternary bilayer. It may be speculated that the specific electrostatic interactions between PI(3)P and the basic Arg^{209} further reduced the diffusion coefficient. These interactions are supposed to stabilize membrane anchoring and together with the recruitment of additional PI(3)P lipids towards the peptide the foundations for the formation of protein and PI(3)P-enriched docking platforms are laid.

4.4.5 Membrane composition-dependent full-length Rab5 orientations

Based on the observations for the truncated Rab5 HVR$^{206-215}$ which revealed specific electrostatic interactions between basic arginine and PI(3)P, the influence of the lipid composition on the conformational orientation and dynamics of full-length Rab5(GDP) and Rab5(GTP) was investigated. All membrane-bound Rab5(GDP) and Rab5(GTP) MD simulations started from a perpendicular protein orientation relative to the bilayer surface (Figure 4.20).

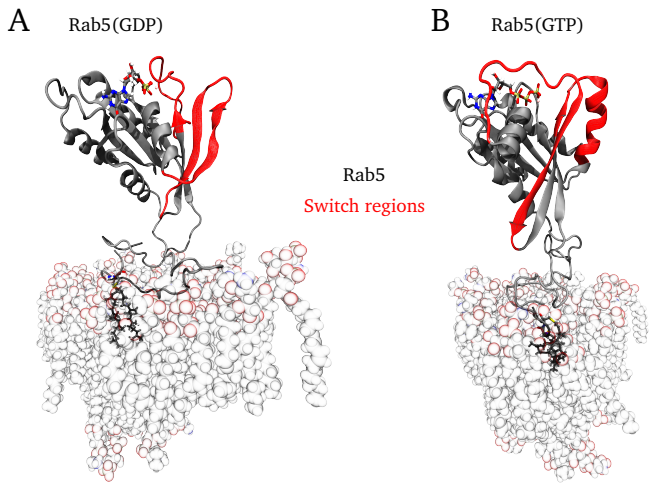


Figure 4.20: Initial conformation of full-length A) Rab5(GDP) and B) Rab5(GTP) in the six-component membrane.

The RMSD of the G domain Cα atoms revealed only slight fluctuations around an average value after 50 ns of MD simulation, indicating a sufficient equilibration time (see appendix, Figure B12). Internal structural deviations of the protein during the simulation were monitored using the RMSF (Figure 4.21).

As expected, structural deviations were most pronounced at the extremely flexible N- and C-termini. The membrane-dependency was only minor, revealing the highest flexibility of the terminal protein regions in pure POPC. However, in the charged six-component membrane a flexibility reduction at the extreme C-terminus (residues 208-215) was observed, pointing to a stabilization of this region by protein-membrane interactions. Within the G domain, the switch regions exhibited strongest structural fluctuations, which were more significant for Rab5(GDP) compared to Rab5(GTP). This is in agreement with the fact that Rab5(GTP) switch regions were stabilized by binding to the γ-phosphate of the GTP nucleotide which is obviously not existent in GDP. Thus, compared to the GTP-loaded state, Rab5(GDP) switch regions exhibited a rather disordered structure.

To obtain a first hint of the Rab5 structural evolution during the MD simulation, visual inspections of the trajectory were carried out and exemplary final configurations of Rab5(GDP) and Rab5(GTP) are shown in Figure 4.22 for the six-component membrane and in the appendix, Figure B13 for the uncharged model bilayers. Originating from a protein orientation almost perpendicular to the mem-

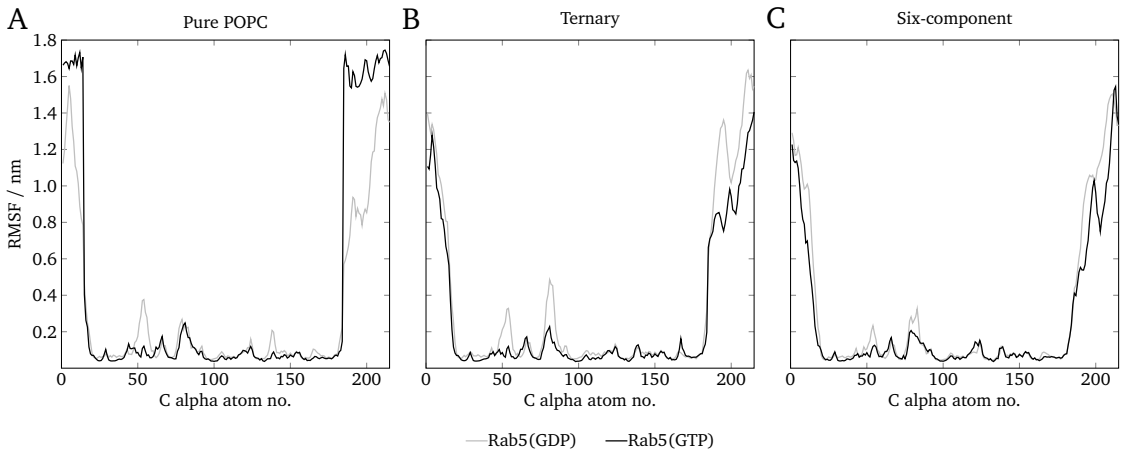


Figure 4.21: The RMSF of Rab5(GDP) and Rab5(GTP) averaged over three independent simulations in A) pure POPC, B) the ternary mixture, and C) the six-component membrane.

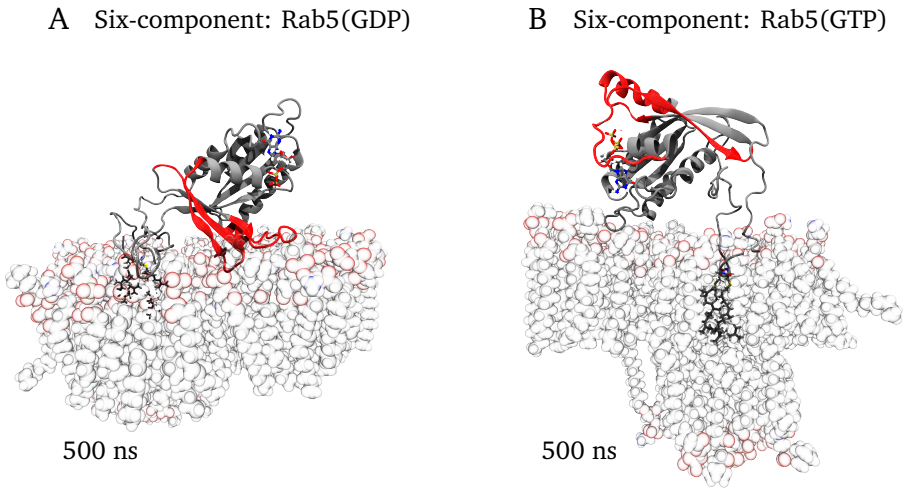


Figure 4.22: Exemplary final configurations of A) Rab5(GDP) and B) Rab5(GTP) in the six-component membrane. The switch regions (red) are hidden in Rab5(GDP) but solvent-exposed in Rab5(GTP).[8]

brane, the Rab5 G domain approached the bilayer surface during the MD simulation. This trend was rather weak in the uncharged bilayers but significant in the charged six-component membrane. In order to confirm these visual observations, the temporal evolution of the protein-membrane distances was monitored over 200 ns in the uncharged model membranes and over 450 ns in the six-component membrane (Figure 4.23).

In the uncharged membranes, i.e. pure POPC and the ternary mixture, Rab5(GDP) and Rab5(GTP) performed quite similar. Protein parts close to residues 50-60, 75-90 and 100-120 and partly around residue 140 had largest distances to the membrane surface. This involved also residues within or close to the switch regions. In contrast, in the six-component membrane, where protein tilting to the bilayer surface was most pronounced, clear activation state-dependent differences could be distinguished.[8] Apart from being closer to the bilayer in general compared to the other membrane models, Rab5(GTP) residues 50-60 and 75-92, roughly representing the switch regions, were most distant to the bilayer surface, too. Surprisingly, for Rab5(GDP) the distance profile was inverse, that is, residues which were closer to the bilayer surface in Rab5(GTP) were more distant in Rab5(GDP) and vice versa. Besides the Rab5(GDP) termini the protein domains close to residues 50-60 and 80-90, approximately representing the switch I and switch II regions, were closest to the charged bilayer surface. Apparently, the switch from the perpendicular protein orientation to the tilted orientation occurred after ≈120 ns for Rab5(GDP) and ≈200 ns for Rab5(GTP) in the six-component membrane.

In order to evaluate the protein orientation more precisely, the G domain orientation pivot angle θ and the internal dihedral ω were defined (see section 4.4.3, Figure 4.10). The dihedral ω as a function of the pivot angle θ is given in Figure 4.24. Independent of the membrane system, the G domain orientation took values between $80°<\theta<160°$ for Rab5(GDP) and $50°<\theta<180°$ for Rab5(GTP). For θ being 90° the G domain adopts a perfectly parallel orientation with respect to the bilayer surface. The internal dihedral ω enabled the distinction between a 'syn' and an 'anti' conformation of the switch regions, i.e. the switch regions either facing the bilayer surface or pointing towards the solvent.

In the uncharged model membranes for both Rab5(GDP) and Rab5(GTP) a broad spectrum of internal dihedrals was observed with the most populated clusters between $0°<\omega<120°$ and $280°<\omega<360°$. The same held true for Rab5(GTP) in the six-component membrane. These 'anti' conformations were characterized by a strong solvent exposure of the switch regions. However, interestingly, for Rab5(GDP) another large population was found at $\omega\approx190°$. These configurations belong to the switch regions 'syn' conformations facing the membrane surface. The combination

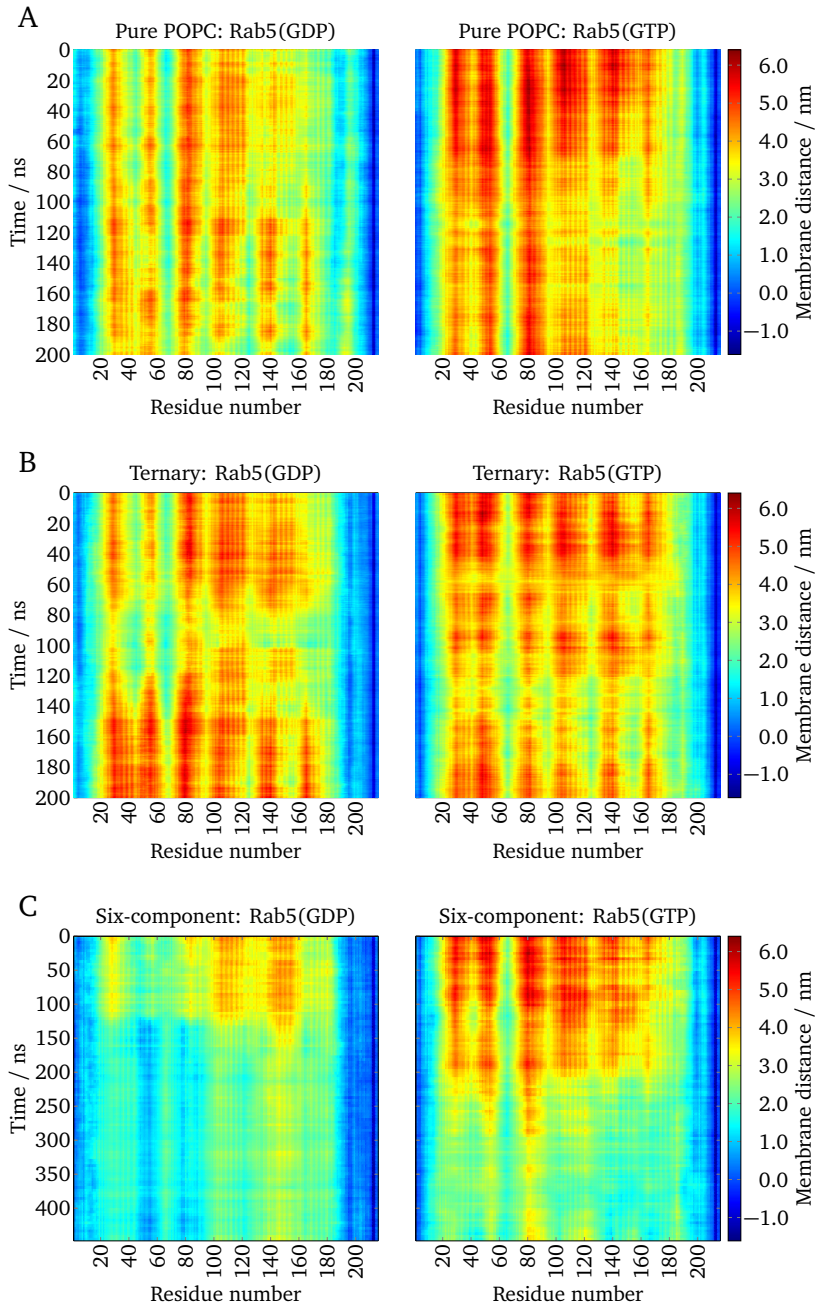


Figure 4.23: Residue-wise protein-membrane distances in A) pure POPC, B) the ternary mixture, and C) the six-component membrane. Data were averaged over three independent simulations for both Rab5(GDP) (left) and Rab5(GTP) (right).

of a G domain orientation almost in parallel to the bilayer surface and the 'syn' conformation of the switch regions caused efficient screening of the Rab5(GDP) switch regions from solvent in the six-component membrane. During the phase of permanent interactions between the Rab5 G domain and the membrane lipids, i.e. the last 400 ns of simulation, a detailed contact analysis was performed (Figure 4.25). Rab5(GDP) formed a greater number of longer lasting contacts predominately via residues within the switch regions compared to Rab5(GTP). Rab5(GDP) residues that remained within 0.5 nm of the membrane for >30 % of the time were Phe^{48}, Ser^{51}, Ile^{53}, Gly^{54}, and Ala^{56} of the switch I region and Ala^{86}, Phe^{87}, Met^{88}, and Arg^{91} of the switch II region.

The interaction forming pattern was mainly made up of polar amino acid residues which formed contacts with the lipid head groups. In addition, patches encompassing the positively charged residues Arg^{39}, Lys^{42}, His^{46}, Arg^{81}, His^{83}, and Arg^{91} were frequently found in close proximity to the membrane

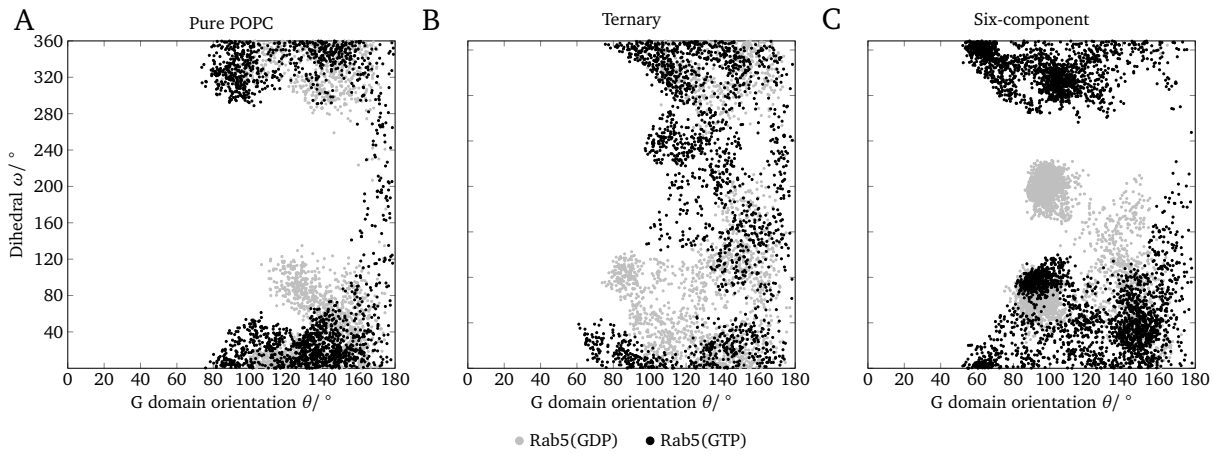


Figure 4.24: The Rab5 protein orientation in A) pure POPC, B) the ternary mixture, and C) the six-component membrane is defined by its internal torsion ω as a function of the G domain orientation θ relative to the membrane surface.

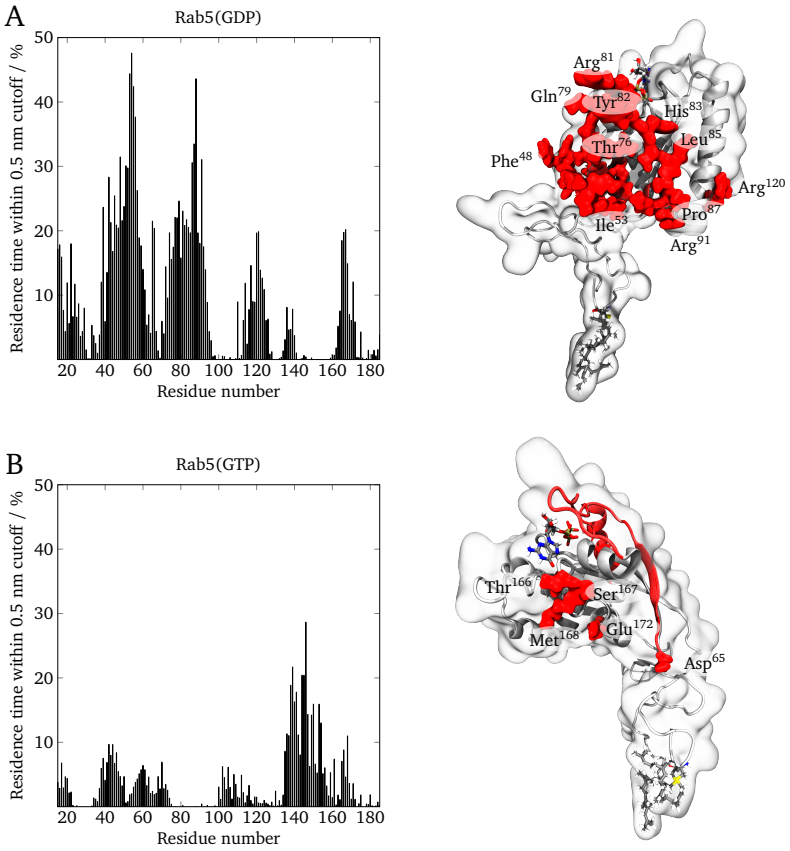


Figure 4.25: Rab5 G domain residues which were found within a 0.5 nm cutoff from the bilayer surface were observed to form protein lipid interactions. Left: Residue percentage residence time during the last 400 ns of A) Rab5(GDP) and B) Rab5(GTP) bound to the six-component membrane averaged over three simulations, respectively. Right: Residues within 0.5 nm from the bilayer in the final protein conformation were colored in red.

surface. This is consistent with previous findings indicating that polar residues play an important role in the reversible adsorption of peripheral membrane proteins.[232] The absolute distance between the Rab5(GTP) G domain and the bilayer surface was larger than in Rab5(GDP) resulting in less extensive contacts with the switch regions being hardly involved. The Rab5(GTP) residues Asn^{139},

Asp^{144}, Phe^{145}, and Gln^{146} were within 0.5 nm of the bilayer lipids in >20 % of the time.

Investigation of the SASA of the full-length protein as well as the switch regions and the C-terminus (residues 185-215) revealed only minor membrane composition-related differences (Figure 4.26). The overall SASA was slightly larger for Rab5(GDP) compared to Rab5(GTP) in all bilayers. The solvent accessibility of the switch regions was similar for both active and inactive Rab5 in the uncharged membrane models but considerably smaller for Rab5(GDP) in the six-component membrane. This is in agreement with the Rab5(GDP) internal 'syn' conformation impeding the attainability of the switch regions. Independent of the membrane composition, the SASA of the C-terminus was marginally larger for Rab5(GDP) over Rab5(GTP). This may be a consequence of the somewhat more extended structure of the HVR in Rab5(GDP).

Nucleotide-dependent binding modes of small GTPases have been described earlier. MD simulations of full-length H-Ras (Harvey rat sarcoma) revealed distinct protein orientations associated with particular interactions formed between the DMPC bilayer and the Ras catalytic domain.[233] Due to the rather short simulations times of 40 ns this study was insufficient to reliably associate the different orientations with either the GDP- or GTP-bound state. In agreement with this, the activation state-dependent lipid protein interactions in the present work occurred after 120 ns in Rab5(GDP) and after 200 ns in Rab5(GTP). Another MD study with full-length GTP-bound K-Ras (Kirsten rat sarcoma) in an anionic membrane detected two dominant protein orientations featuring significant differences in their switch regions accessibility.[234] The conformation with hardly accessible switch regions exhibited a more extended structure of the HVR. This agrees well with the Rab5(GDP) state in the six-component membrane observed in the present work. The nucleotide state-dependent arrangement of lipidated small GTPases in different membrane-associated orientations seems to be a general concept in regulation and signal transduction related to this class of enzymes. Distinct G domain orientation dynamics of small GTPases have been reviewed recently.[235] Their effects on the binding of regulatory and effector proteins may represent an important mechanism in the functionality of small GTPases. Here, in the negatively charged membrane, only the Rab5(GTP) switch regions were accessible to potential binding partners whereas Rab5(GDP) was rather effector inaccessible. This raises the general question, how binding partners recognize either the active or inactive Rab protein state. The switch regions are specific markers for the enzyme's activation state and the site where Rab-effector interaction takes place. Among the Rab protein family the long-range electrostatic potential is highly conserved between residues of the switch I and switch II regions.[93] The electrostatic potential maps of membrane-bound Rab5(GDP) and Rab5(GTP) are provided in the appendix, Figure B14 and illustrate the Rab5 charge distributions. The Rab5 protein has an overall positive charge of plus two at pH 7. The blue surface areas indicate regions of positive charge

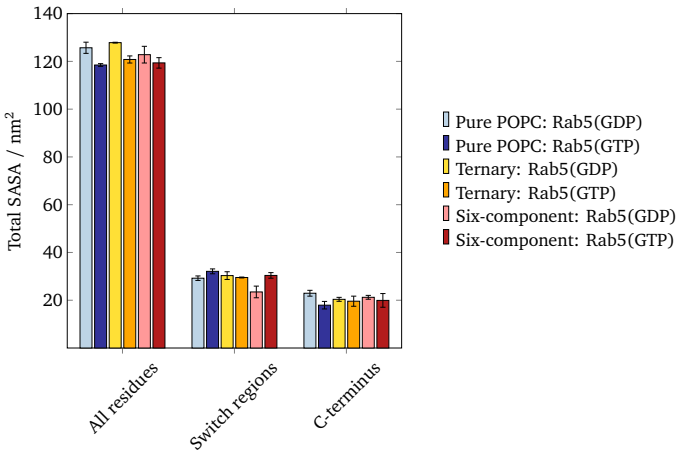


Figure 4.26: Absolute SASA of the whole protein, the switch regions as well as the C-terminus (residues 185-215) of double-GG Rab5(GDP) and Rab5(GTP) in the three investigated membrane systems.

whereas red regions indicate a negative potential. Thus, the Rab5 regions colored in blue were expected to preferentially interact with the negatively charged phosphate groups within the bilayer. The nucleotide state-dependent recognition of binding partners is expected to occur predominately via electrostatic recognition. In addition, nonspecific electrostatic interactions between basic amino acid residues and negatively charged phospholipids are required for peripheral membrane protein regulation and activity.[236] An appropriate orientation of the protein relative to the membrane surface thus increasing the probability of protein-lipid interaction is realized by long-range electrostatic attraction. The rather short-range nonpolar hydrophobic interactions are more variable among Rab family members and come into play when the protein approaches the bilayer surface. In Rab5 they assisted in forming stable interactions between amino acid residues of the catalytically active G domain and bilayer lipids to adapt a tilted membrane orientation with fully solvent-exposed switch regions in Rab5(GTP) and rather shielded switch regions in Rab5(GDP). Hydrogen bonds were preferentially formed between basic protein residues of the G domain and the N and C-termini and the inositol ring phosphate group of PI(3)P in the charged six-component membrane (see appendix, Figure B15).

The question remains why protein-lipid interactions were more pronounced in Rab5(GDP) compared to Rab5(GTP). This may be due to the secondary structure elements formed by these regions. The Rab5(GDP) switch I region formed two β-strands which were internally twisted and, in contrast to Rab5(GTP), not stabilized by contact formation with the nucleotide. This resulted in a disordered and more flexible structure which was to some extent pulled away from the protein core reaching towards the bilayer surface. The same was true for the Rab5(GDP) switch II region which formed a disordered coil and was in a rather extended shape. Consequently, the switch region amino acid residues were more solvent-exposed in the initial protein conformation and appropriate to establish electrostatic interactions with lipids in an early stage. In contrast, switch region residues in initial Rab5(GTP) were predominately involved in intramolecular interactions and contacts with the nucleotide and magnesium ion and less protracted to the membrane.

However, regulatory protein binding to Rab5 is not limited to the GTP-loaded state but does also occur in the GDP-bound state.[8] One example is the GEF Rabex-5 which helps activating Rab5(GDP) by inducing the exchange of GDP with GTP. The corresponding Rab5 interaction site is located close to or within the switch regions, namely residues Glu^{72}, Asp^{75}, Arg^{81}, and Tyr^{89}.[115] Additionally, RabGDI binds Rab5(GDP) in order to promote its membrane extraction and to keep it soluble in the cytoplasm. Chapter 6 of the present thesis is dedicated to Rab5(GDP):RabGDI interactions and provides a detailed picture on protein-protein interaction sites and the membrane extraction mechanism. It may be speculated that Rab5(GDP) binding partners induce small Rab5 re-orientations which are sufficient to expose the switch regions and allow protein complex formation.

Apart from small GTPases, several examples for peripheral protein binding to membranes are found in literature. Molecular dynamics studies of the homologue oxysterol binding protein of yeast (Osh4) revealed a single stable binding conformation to anionic membranes containing PS and PI(4,5)P_2 but only weak and transient binding to bilayers composed of solely zwitterionic lipids.[237] Several distinct binding sites mostly dominated by positively charged arginine and lysine residues were detected. These basic residues were suggested to attract anionic lipids and to form first contacts to the bilayer surface depending on their conformation to be either solvent-accessible or oriented towards the protein interior. In agreement with the rather extended switch regions in Rab5(GDP), membrane contact occurred exclusively when the basic residues including loops were extended.

In order to examine if the approaching G domain influenced the membrane structure a Voronoi tessellation of the local lipid order was performed (see appendix, Figure B16). In addition to the significant order reduction around the GG chains, the membrane regions close to the tilted G domain were characterized by a decreased lipid order as well. This effect was mainly observed for the six-component membrane. However, absolute local order deviations were most prominent in pure POPC resulting in a less homogeneous spatial order distribution compared to the other model membranes. This may be a result of the faster diffusion of the GG chains in pure POPC, which was 1.8 times higher

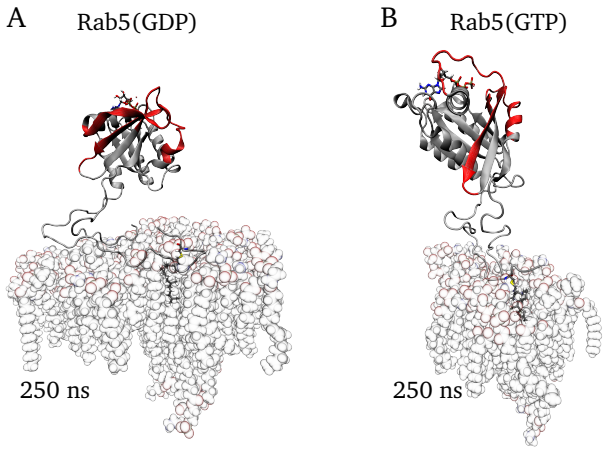


Figure 4.27: Final conformations of A) mono-GG Rab5(GDP) and B) mono-GG Rab5(GTP) in the physiological six-component membrane. In both cases, the switch regions (red) are fully solvent-accessible.

than in the ternary membrane and 2.2 times higher compared to the six-component membrane. A table of GG anchor lateral diffusion coefficients for the different membrane and protein systems is given in the appendix (Table B3). Since MD simulations with the truncated $HVR^{206-215}$ revealed very similar diffusion coefficients in the ternary and six-component membrane, the diffusion reduction in the charged bilayer cannot be explained by the lipid composition alone. It is supposed that electrostatic and hydrophobic interactions which occurred between the peptide and PI(3)P in the $HVR^{206-215}$ calculations slowed down diffusion to a moderate extent. In case of full-length Rab5 a variety of those interactions was formed not only between PI(3)P and the HVR but also between charged lipids and amino acid residues of the protein G domain, thus notably reducing diffusion in the six-component membrane.

Surprisingly, this trend was not observed for mono-GG full-length Rab5. In order to investigate the effect of a GG chain number reduction on membrane anchorage and protein dynamics, MD simulations of mono-GG Rab5(GDP) and Rab5(GTP) in different model membranes were performed. The lateral diffusion of the GG anchor was 1.7 times (pure POPC) to 3.5 times (six-component membrane) faster compared to double-GG Rab5 (see appendix, Table B3). Unexpectedly, for mono-GG anchor diffusion was faster in the six-component membrane compared to the ternary mixture. This indicated that protein lipid interactions which slowed down diffusion in case of double-GG Rab5 did not occur to this extent between mono-GG Rab5 and the six-component bilayer. Monitoring the protein-membrane distances in case of mono-GG Rab5, no membrane composition-dependent differences were observed (appendix, Figure B17). For the six-component membrane the final Rab5 conformations after 250 ns MD simulation are shown in Figure 4.27.

The mono-GG Rab5 G domain hardly approached the bilayer surface in all membrane models. In particular, the significant G domain tilt revealed for double-GG Rab5(GDP) in the physiological six-component membrane was absent for mono-GG Rab5(GDP). This indicated that Rab5 membrane anchorage with only one GG chain was indeed adequate to associate the protein with the bilayer during 250 ns MD simulation but not sufficient to induce a G domain bending toward the membrane surface. It may be speculated that the number of anchor chains significantly affected anchor and protein dynamics. The distributions of the anchor insertion depths of mono-GG Rab5 compared to double-GG Rab5 are given in the appendix, Figure B18. In contrast to double-GG, whose insertion depths clearly depended on the lipid composition and the bilayer thickness, in case of mono-GG the insertion depths showed no correlation with lipid composition and rather broad bimodal distributions. This extreme flexibility may be the reason for the higher structural fluctuations revealed by the RMSF at the C-terminus of mono-GG Rab5(GDP) in the six-component membrane (Figure 4.28A). However, the general Rab5 structure compactness measured by the radius of gyration (Figure 4.28B) was sim-

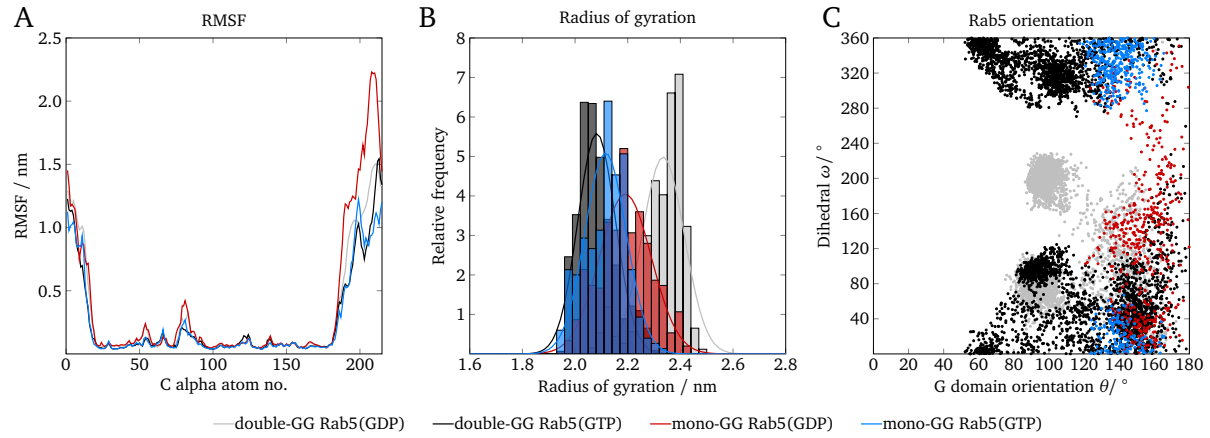


Figure 4.28: Comparison between double-GG Rab5 (450 ns) and mono-GG Rab5 (200 ns) in the six-component membrane with regard to A) the RMSF, B) the protein radius of gyration, and C) the internal torsion ω as a function of the G domain orientation θ.

ilar for mono-GG Rab5 in both nucleotide states and in agreement with double-GG Rab5(GTP). The larger gyration radius in double-GG Rab5(GDP) was a consequence of the less ordered switch regions structure compared to Rab5(GTP) and the rather extended C-terminus. Comparison of the Rab5 orientations in the six-component membrane revealed that mono-GG Rab5 was unlikely to adopt a parallel G domain orientation with respect to the bilayer surface (Figure 4.28C). Furthermore, the switch regions were more frequently found in the 'anti' conformation being fully solvent-accessible. Consequently, in contrast to double-GG Rab5(GDP), mono-GG Rab5(GDP) was at no time found in the tilted conformation with solvent-screened switch regions.

Most Rab proteins undergo doubly geranylgeranylation to be correctly targeted and inserted within the membrane. Only a few Rab proteins possess a single GG motif responsible for membrane anchoring, e.g. Rab8 and Rab13.[238] It has been shown that different membrane binding motifs are correlated with the protein cellular localization.[107] There are examples for additional peripheral membrane proteins for which only one covalently attached lipidation is described. In addition to this isoprenyl group which partitions into the bilayer hydrophobic region to anchor the protein, small GTPases like K-Ras4B, Rac-1, Rho-A, or Rap1-A often contain an adjacent cluster of basic amino acid residues to bind negatively charge membranes.[239] On the other hand, N-Ras and H-Ras are farnesylated and exhibit an additional palmitoylation at a C-terminal cysteine residue. When the palmitoylation is missing both Ras proteins are incorrectly targeted to the cytoplasm and non-functional.[240,241] MD simulations of a lipidated Ras peptide have shown that a single lipid modification is not sufficient to permanently anchor N-Ras to the membrane due to an inadequate hydrophobic binding energy which causes an equilibrium of adsorption and desorption processes.[170] Correctly double-GG Rab5 is targeted to the EE where it regulates vesicle fusion and transport.[3] An experimental study with wild type Rab5 and different Rab5 mutants where the C-terminal double-GG motif was replaced by double-cysteine and mono-cysteine motifs of other Rab proteins showed that those proteins were efficiently prenylated.[238] However, functionality and correct targeting to the EE was only ensured for the Rab5 proteins with di-cysteine motif. Mono-cysteine motif proteins were non-functional and targeted to the endoplasmic reticulum membrane. Pursuing studies with hybrid Rab proteins, i.e. proteins in which the N-terminal region of one Rab protein was fused with the C-terminal HVR of another Rab family member, investigated whether the HVR was responsible for targeting the GTPase to the corresponding subcellular compartment.[242] In case of Rab5 fused with the HVR of Rab1a or Rab2a, targeting to the early endosomal membrane was not disturbed. Conversely, hybrid proteins with a Rab5 HVR were not targeted to the EEs. Consequently, a general membrane targeting signal is not exclusively part of the HVR. Nevertheless, Rab family-specific (RabF) and Rab subfamily-specific (RabSF) sequences were identified whose substitution among different Rab proteins caused a redistribution of the hybrid proteins to other subcellular compartments. It was shown that RabF regions,

which cluster around the switch regions, were involved in effector binding and allowed a discrimination between the nucleotide states. On the other hand, effector binding to the RabSF regions ensured effector specificity.[243] More recent studies demonstrated that the role of the HVR regarding membrane targeting depends on the specific Rab protein.[244] For Rab1 and Rab5 the HVR is primarily a linker and not necessary for targeting to the correct subcellular compartment. In case of Rab7, the HVR interacts with the effector Rab-interacting lysosomal protein (RILP) which in turn is essential for correct Rab7 targeting. In Rab35 the HVR is involved in electrostatic interactions with the bilayer via a polybasic cluster.

In the present study, the six-component membrane composition was chosen to mimic the EE membrane, i.e. the bilayer Rab5 is naturally associated with. Assuming that the tilting of the Rab5 G domain toward the membrane surface and the presence of the two distinct orientation modes of Rab5(GDP) and Rab5(GTP) are essential for correct protein function, only double-GG Rab5 would be able to recruit effector proteins to the membrane surface. Mono-GG Rab5 revealed no nucleotide-dependent orientation modes in none of the investigated model membranes. This is in agreement with the experimental findings mentioned above, demonstrating that mono-cysteine motif containing Rab5 was non-functional.[238] The present MD simulations may provide part of the structural explanation for this observation, that is, the impeded formation of G domain-lipid interactions due to faster protein diffusion and less electrostatic interactions between basic amino acid residues and EE-specific PI(3)P signaling lipid.

4.5 Summary

In this chapter, the structure and dynamics of membrane-bound Rab5 HVR$^{206-215}$ as well as full-length Rab5 were thoroughly investigated by full-atomistic and coarse-grained MD simulations. The focus was on addressing the following issues:

- Determination and validation of force field parameters for the Rab5 GG-Cys residues
- Investigation of peptide-lipid interactions between Rab5 HVR$^{206-215}$ and differently composed model membranes in full-atomistic detail as well as determination of the binding Gibbs free energy profile of the double-GG peptide
- Elucidation of lipid aggregation behavior in close proximity to the HVR$^{206-215}$ in coarse-grained MD simulations
- Comparison of mono- and double-GG full-length Rab5(GDP) and Rab5(GTP) dynamics in different model membranes

Independent of the membrane composition, the GG-modified Rab5 C-terminal HVR$^{206-215}$ exhibited a highly flexible, intrinsically disordered structure with several distinct conformational states. The structure of the surrounding bilayer lipids was only marginally influenced by the peptide, indicating that the GG anchor adapts to the membrane environment by adjusting the insertion depth and reducing the chain order parameter. Specific peptide-lipid interactions were observed between the basic Arg^{209} residue and the PI(3)P signaling lipid inositol head group of the negatively charged six-component membrane. These electrostatic interactions persisted during the whole GG anchor extraction process as revealed by SMD and US. Consequently, the HVR$^{206-215}$ binding free energy was increased in the physiological membrane compared to the uncharged bilayers. Furthermore, compared to the simply zwitterionic bilayers, Arg^{209}-PI(3)P interactions in the six-component membrane prevented a deeper insertion of the HVR residues into the lipid head group region, thus further increasing the flexibility of the peptide. In addition, long-term coarse-grained MD simulations showed an electrostatically driven recruitment of anionic PI(3)P as well as cholesterol to the HVR$^{206-215}$ which may serve initiating PI(3)P and Rab5-enriched signaling platforms.

For full-length double-GG Rab5, clear membrane composition-dependent orientation differences be-

tween Rab5(GDP) and Rab5(GTP) were revealed. An evolution from an initially perpendicular G domain orientation toward a tilted conformation was detected primarily in the negatively charged six-component bilayer. Exclusively in the physiological membrane, a Rab5(GDP) G domain rotation was observed, resulting in a 'syn' configuration of the switch regions. Due to a simultaneous parallel alignment of the G domain to the membrane surface, the switch regions were partially buried between protein and bilayer and, thus, hardly solvent- and effector accessible. Mainly positive surface potentials around the Rab5 switch regions were expected to be responsible for the membrane association and approaching of the catalytically active G domain toward the negatively charged bilayer via long-range electrostatic interactions. In immediate proximity to the membrane surface additional short-range hydrophobic interactions were formed between basic or nonpolar amino acid residues and bilayer phospholipids. The reason behind the fact that exclusively Rab5(GDP) was found in the 'syn' conformation with inaccessible switch regions is assumed to be the secondary structure of the switch regions. Compared to Rab5(GTP), the Rab5(GDP) switch regions were rather disordered with less contacts formed with the protein interior or the nucleotide and thus initially more solvent-exposed. Therefore, electrostatic interactions between the G domain residues and the lipids were more likely to occur in the GDP-loaded state.

The long C-terminal HVR served as a flexible linker between the membrane GG anchor and the G domain. It allowed a great degree of freedom, high amplitude motions and orientation changes of the G domain like a balloon on the cord. In Rab5(GDP), the HVR was rather extended providing a potential association site for binding partners exclusively interacting with the GDP-bound Rab5 state.

In contrast to double-GG Rab5, mono-GG Rab5 revealed only one membrane-binding mode irrespective of the nucleotide state or bilayer composition. A re-orientation from the initial perpendicular toward a tilted conformation was not observed. It is supposed that due to the single GG chain anchorage mono-GG Rab5 diffusion was significantly faster, thus impeding the formation of long-range electrostatic interactions between the G domain residues and the phospholipids.

5 The dimerization of membrane-associated Rab5 proteins

Since 1988[245] and 2000[246] , it was speculated that Ras small GTPases form dimeric complexes at membrane surfaces. A recent study revealed that GTP-loaded K-Ras dimers are required for activation of the mitogen-activated protein kinase (MAPK). Thus, inhibition of Ras dimerization may represent a strategy in fighting mutant Ras-driven cancers.[247] These findings created renewed interest in the field and several subsequent experimental studies reported also evidence for the existence of N-Ras[248] and H-Ras[249] dimers. However, it still remains elusive whether Ras dimerization is driven by direct protein-protein contacts in the catalytic domain or rather indirectly by aggregation of membrane-associated complexes.[250] Several reports focused on the dimerization of Ras-related proteins like Rab9,[251] Arf1,[252] and Sar1[253] for which dimerization via direct protein-protein interactions between the G domains was detected. In contrast, dimerization via polybasic stretches at the membrane-associated C-terminus of one monomer was observed for Rho[254] and yeast Rsr1.[255] Due to its close relationship with Rab9, Rab5 is speculated to form membrane-bound dimers as well. A chemical cross-linking study indicated the existence of Rab5 dimers in 2001,[256] however, the interactions and dynamics underlying dimerization remain elusive. Its role in EE trafficking makes Rab5 an important regulator of cellular transport, recycling and degradation processes.[8] Thus, Rab5 malfunction is often related to endocrinological diseases and certain types of cancer. Similar to K-Ras, inhibition of a potential Rab5 dimerization by therapeutic drugs may represent a valuable starting point for the development of specialized Rab5-focused therapies. A profound understanding of the underlying dimerization mechanism and knowledge of protein-protein contact sites is crucial for such approaches.

This chapter focuses on the question whether Rab5 is able to form membrane-associated dimers or higher-order oligomers. Homotypic Rab5(GDP) and Rab5(GTP) dimers were manually arranged with their switch regions oriented to each other and subjected to 250 ns of MD simulation. The first part of this chapter describes the orientation of the Rab5 monomers relative to each other to enable a potential dimerization process. Section 5.2 is dedicated to the evaluation of the MD simulations and discusses the outcome with regard to recent findings of the closely related small GTPase Ras.

5.1 Modeling the potential Rab5 dimer interface

Today, experimental studies provided evidence that membrane-bound Ras proteins exist as monomers, dimers and are able to assemble into oligomers which form isoform- as well as activation state-specific nanoclusters at the plasma membrane.[257] Ras dimers were detected using super-resolution microscopy and electron microscopy which together with computational approaches provided insight that the K-Ras dimerization interface is made up by the G domain.[250,258] For K-Ras two potential dimerization regions in were identified, namely the $\alpha 3/\alpha 4$ helical and the β-sheet interfaces (Figure 5.1).[258]

Structural superposition revealed a high structure conservation between K-Ras and Rab5, allowing the transfer of K-Ras dimerization regions to the corresponding regions in Rab5. The helical dimerization interface comprised Rab5 residues 109-125 and 148-154, and the β-sheet dimerization interface involved Rab5 residues 38-61 and 69, accounting for the switch I region. Based on these findings, two monomers of membrane-bound Rab5(GDP) or Rab5(GTP), respectively, were placed

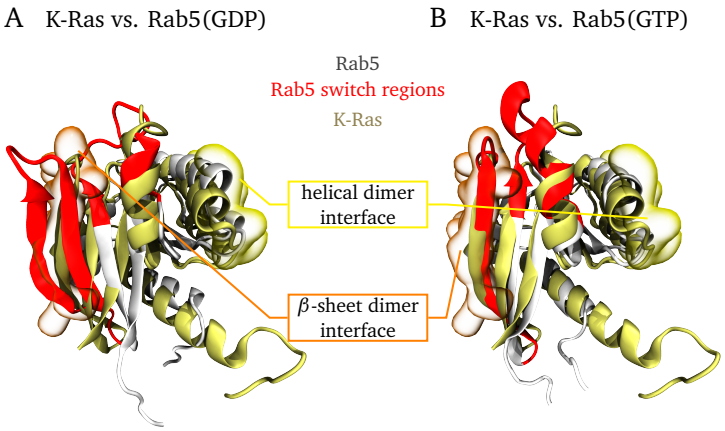


Figure 5.1: Superposition of N-terminal K-Ras (from PDB entry 4DSO[259] colored in dark yellow) with the catalytic G domain of A) Rab5(GDP) and B) Rab5(GTP) (light gray). In K-Ras two dimerization interfaces were identified, namely the $\alpha 3/\alpha 4$ helical interface (highlighted in yellow) and the β-sheet interface (orange).[258]

in close proximity relative to each other, facing each other with the potential dimer interaction sites (Figure 5.2).

Following the MD simulation protocol for un-complexed Rab5(GDP) and Rab5(GTP), the proteins were inserted in model membranes of increasing complexity, i.e. pure POPC, a ternary zwitterionic mixture including cholesterol and a negatively-charged six-component bilayer mimicking the EE membrane (see section 4.2, Table 4.1). Preparation (including energy minimization, heating to 310 K and equilibration) of the protein-membrane systems was done accordingly.

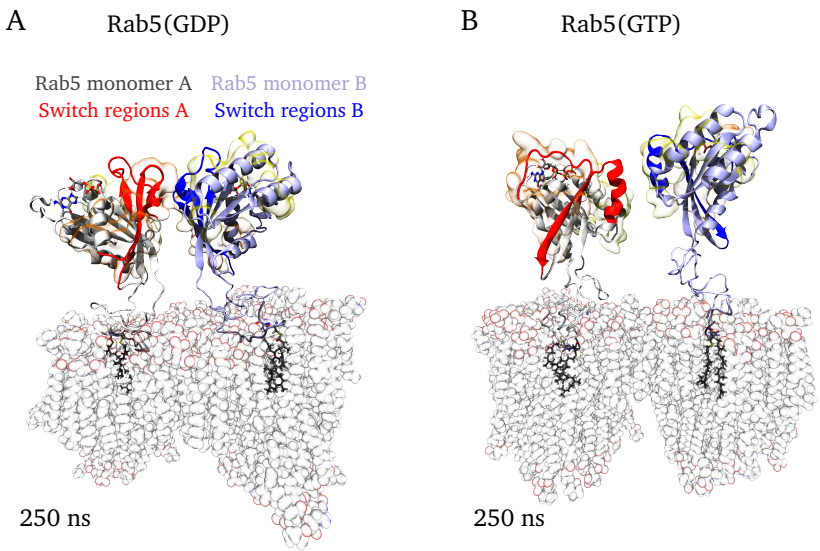


Figure 5.2: Initial arrangement of two monomers of A) Rab5(GDP) and B) Rab5(GTP) in the six-component membrane. The helical and β-sheet dimerization interfaces identified in K-Ras were assigned to the corresponding regions in Rab5 and colored in yellow and orange, respectively.

5.2 MD simulation of membrane-bound dimeric Rab5

In contrast to various Ras proteins, for which dimerization has been observed experimentally,[249,260,261] a similar behavior for Rab5 GTPases still remains uncertain to date. In this section, two membrane-bound monomers of either Rab5(GDP) or Rab5(GTP) were arranged in close proximity and subjected to full-atomistic MD simulations. Interactions between residues of the dimerization sites newly formed or were stabilized during the simulations and provided insights into a possible nucleotide-state dependent Rab5 dimerization on different membrane surfaces.

5.2.1 MD simulation setup

In accordance with the un-complexed full-length Rab5 MD simulations, the bilayer was placed in the x, y plane and both full-length Rab5 monomers were positioned in +z direction above the different membrane surfaces. The following simulation setup was used:

- The pre-oriented Rab5(GDP) and Rab5(GTP) dimers were solvated in explicit water and ions and inserted into either pure POPC, the ternary mixture or the six-component membrane via their GG chains. For each model membrane and each nucleotide state 50 ns of equilibration and 200 ns production MD simulation was performed (Figure 5.3).

The MD simulations were performed analogously to full-length Rab5 (see section 4.4.1) with identical parameter settings.

5.2.2 Analysis of dimeric Rab5 simulations

The protein RMSD, RMSF and SASA were calculated using the equations provided in section 4.4.3. Additionally, the SASA buried at the protein-protein interface (i.e. the buried SASA) was calculated

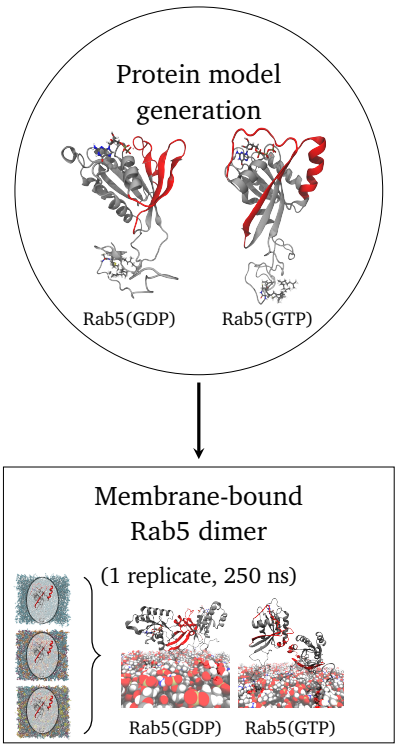


Figure 5.3: Full-atomistic MD simulations performed for Rab5(GDP) and Rab5(GTP) dimers in three different model membranes.

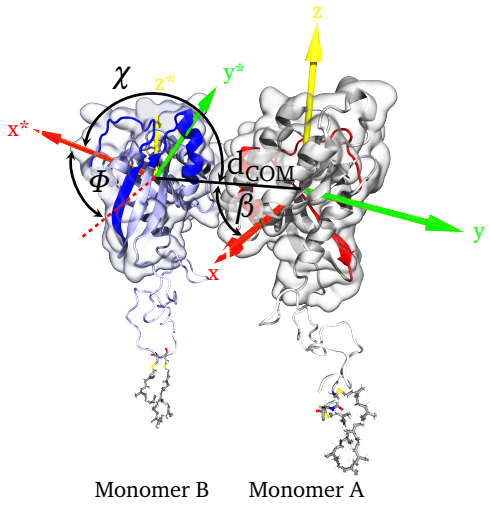


Figure 5.4: The intermolecular distance d_{COM} between the centers of both monomers as well as three relative angles were defined to characterize the dimer conformations. The position of monomer B relative to monomer A was described by the angle β. The phase Φ corresponded to the rotation of monomer B around its z-axis (z*). χ was a measure for the angle under which monomer A bound to monomer B. The definition of relative orientation angles was based on references[262–264].

as the sum of the two monomers' SASA minus the SASA of the dimer.

Two associated Rab5 monomers were denoted as a dimer for intermolecular distances d_{COM} <4.5 nm which applied to all simulations (Figure 5.4). This distance was chosen due to the average spatial extension of the Rab5 G domain which was 4.4 nm, 4.6 nm and 4.0 nm in x, y and z direction, respectively.

Three angles were defined based on the G domains' principal axes (x, y, z in monomer A, x*, y*, z* in monomer B) in order to characterize the relative orientation of both monomers in a dimer.[262–264] 1) The angle β described the position of monomer B relative to monomer A, that is the angle between the axis connecting both protein centers and monomer B principal axis x*. 2) The phase angle Φ represented the rotation of monomer B around its z* axis, i.e. the angular shift between both monomer' x axes. 3) The angle χ was calculated as $\chi = (180° + \beta - \Phi) mod 360$, using the modulo operator and referred to the angle under which monomer A bound to monomer B, relative to the x-y plane of monomer A. All angles were calculated according to equation (4.4) provided in section 4.4.3. Hydrogen bonds and salt bridges between both Rab5 monomer amino acid residues were calculated using the VMD HBonds 1.2 plugin and Salt Brigdes 1.1 plugin, respectively. For hydrogen bonds, a cutoff distance of 0.30 nm and a cutoff angle of 20° between donor and acceptor atoms was used. Salt bridges were formed between oxygen atoms of acidic amino acids (i.e. aspartate and glutamate) and nitrogen atoms of basic residues (i.e. arginine, lysine, histidine) within a cutoff distance of 0.32 nm. APBS (adaptive Poisson-Boltzmann solver) was used to determine the electrostatic contribution of the Rab5 binding energy.

5.2.3 Structural characterization of Rab5 dimers

Independent of the membrane composition, all dimer MD simulations started from a protein-protein orientation perpendicular to the bilayer surface (see Figure 5.2). The final protein complex conformations after 200 ns of MD in three different membrane models are displayed in Figure 5.5.

Pictures of the final Rab5 dimer conformations from both side and top views are provided in the ap-

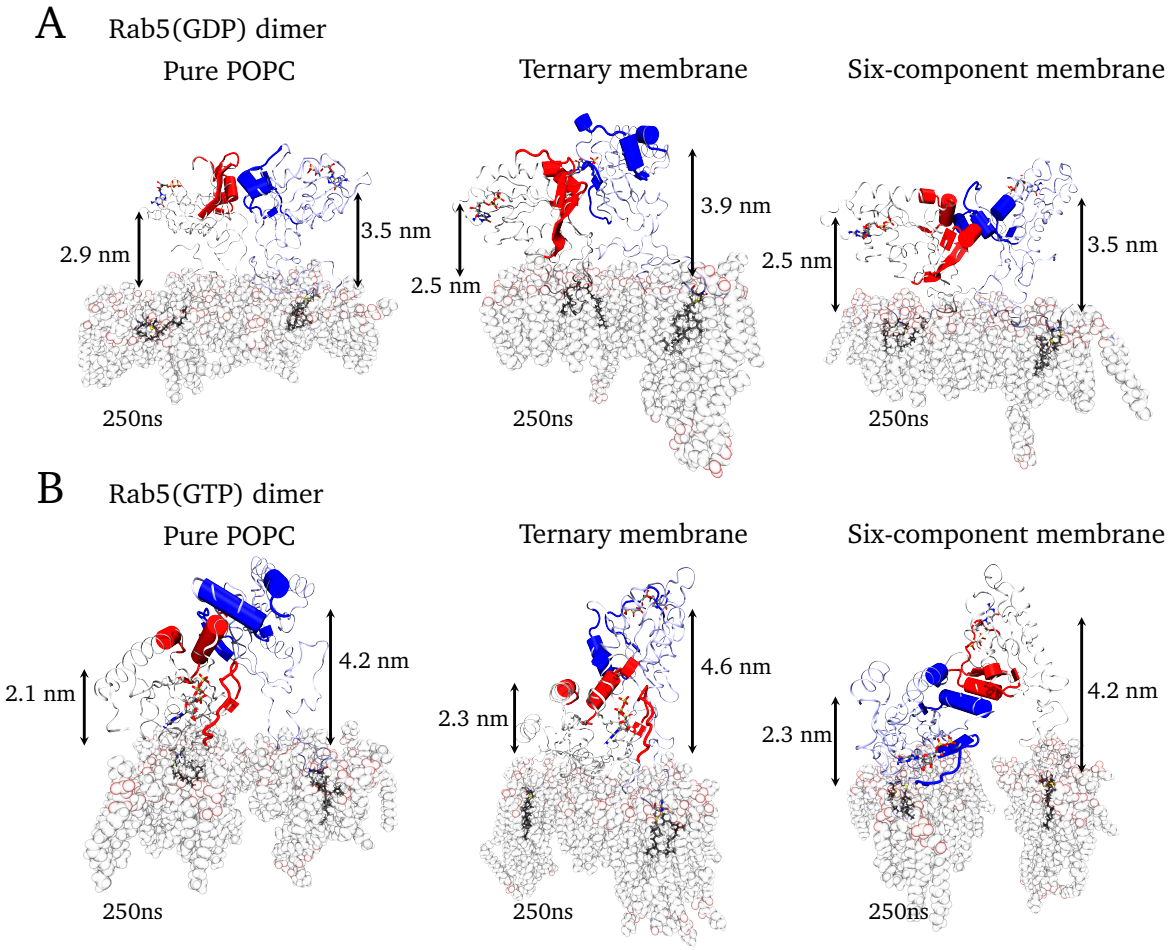


Figure 5.5: Final conformations of the A) Rab5(GDP) and B) Rab5(GTP) dimers in pure POPC (left), the ternary mixture (middle) and the six-component membrane (right). The highlighted regions correspond to the three interactions sites identified in both nucleotide states by protein-protein distance analysis, namely residues 45-58, 82-95 and 118-123 (monomer A: red, monomer B: blue). The distances between the protein centers and the membrane surface are indicated.

pendix, Figure B19. Compared to the initial conformations the relative orientations of both monomers did not change significantly. Parts of the Rab5 switch regions remained the main dimer interaction face during the whole simulation. In the Rab5(GDP) dimeric complexes, both monomers interacted with each other via their G domains at similar heights above the membrane surface (Figure B19A-C). In contrast, in case of Rab5(GTP), the simulations revealed a significant tilting of one monomer towards the membrane surface, while the second monomer stayed in a rather upright position on top of the first like a lid (Figure B19D-E). The protein RMSD (see appendix, Figure B20) represented a measure for the deviation of the monomer structures from their initial conformations. Compared to the GTP-loaded states, Rab5(GDP) structures exhibited a higher RMSD. This was due to the K-Ras structure (PDB entry 4DSO[259]) used as a template for the dimeric interaction face which represented the protein-protein interactions in the GTP-loaded state. Thus, GDP-bound Rab5 structures first needed to re-orient into their preferred relative orientations. Surprisingly, Rab5(GTP) monomers in the six-component membrane revealed higher RMSD values compared to the other membrane models. In order to elucidate nucleotide state- or membrane composition-dependent protein regions of higher flexibility the RMSF was calculated and compared to un-complexed Rab5 from chapter 4 (Figure 5.6). Besides the Rab5 N- and C-termini, the switch I (residues 44-66) and switch II (residues 75-91) regions exhibited greatest structural flexibility, whereby switch II was more disordered than switch I. This may be a consequence of the involvement of switch I residues

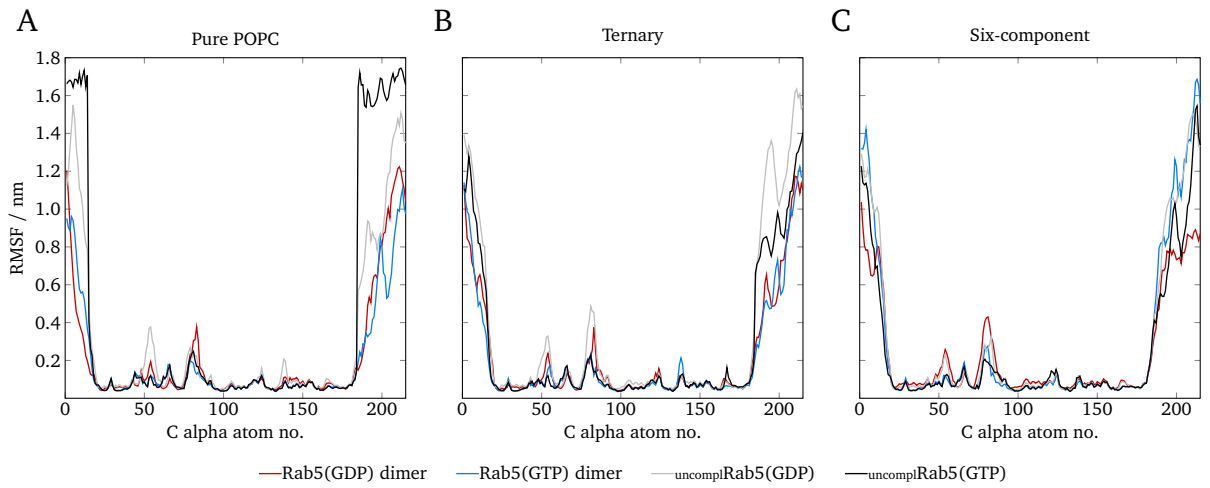


Figure 5.6: RMSF of Rab5(GDP) and Rab5(GTP) monomers in a dimer (red and blue, respectively) compared to a Rab5 monomer (gray and black) bound to A) pure POPC, B) the ternary mixture and C) the six-component membrane. Data from the individual Rab5 proteins of one dimer were averaged.

in the β-sheet dimerization interface and their stabilization by protein-protein interactions. Compared to un-complexed Rab5 the terminal regions were stabilized in the two interacting monomers in pure POPC (Figure 5.6A). In contrast, in the ternary mixture and in the six-component bilayer, the Rab5(GTP) N- and C-termini were as flexible as in the un-complexed protein (Figure 5.6B, C).

The intermolecular distance d_{COM} between the Rab5 monomer G domain geometric centers was investigated in order to estimate the tendency of Rab5 to form dimers (Figure 5.7).

The average intermolecular distance was similar in the GDP- and GTP-bound states in the six-component membrane and with approximately 3.8 nm by 0.2 nm-0.5 nm larger compared to the uncharged bilayers. Shortest intermolecular distances of about 3.3 nm were observed for the Rab5(GDP) dimer in the ternary mixture and the Rab5(GTP) dimer in pure POPC. The differences in d_{COM} indicated different types of protein-protein binding, however, the intermolecular distance alone was not sufficient to distinguish such different interaction modes. Similar to K-Ras GTPases, a participation of the switch regions was expected to be involved in Rab5 dimerization. The distances of each amino

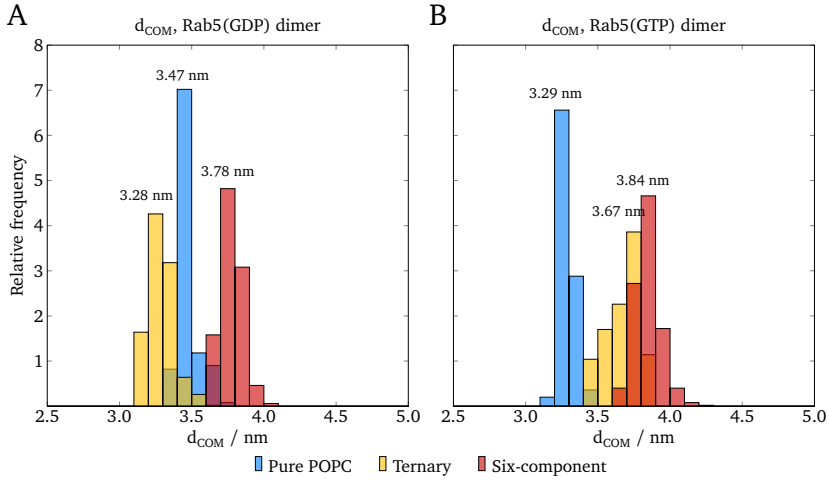


Figure 5.7: The intermolecular distance d_{COM} between the centers of both monomers in case of A) Rab5(GDP) and B) Rab5(GTP) served as a geometric criterion for dimerization. Values above the bars correspond to the average intermolecular distance in the respective model membrane.

acid of one monomer to the center of the second monomer were calculated to get a thorough understanding of which residues promoted Rab5 dimerization (see appendix, Figure B21, Figure B22 and Figure B23). Slight differences between both monomer residue-residue interactions indicated that the dimeric complexes were not perfectly symmetric. The distances between the HVR residues of both monomers were larger in the GTP-bound dimers. The influence of the membrane composition as well as the nucleotide state on the protein-protein distances, and thus on the dimer structure, is shown Figure 5.8. Distances were averaged over the three investigated membrane systems (Figure 5.8A) and revealed only minor standard deviations. Nucleotide state-dependent differences were mainly observed in the N-terminal region, a short patch preceding the C-terminus (residues 185-200) as well as between Rab5 residues 40-70, which comprise the switch I region. Here, protein-protein distances were smaller in the GDP-bound dimers compared to the GTP-loaded states. The influence of the individual Rab5 monomers of one dimer and the membrane composition on the protein-protein distances is broken down in Figure 5.8B and C. In the GDP-loaded state, the distances were only slightly affected by the membrane composition. However, in the GTP-bound dimer, a membrane composition dependency was observed. Protein-protein distances in the N-terminal region including the switch I (residues 1-75) and the extended HVR (residues 160-215) were significantly shorter in the Rab5(GTP) dimer bound to the pure POPC bilayer compared to the other model membranes.

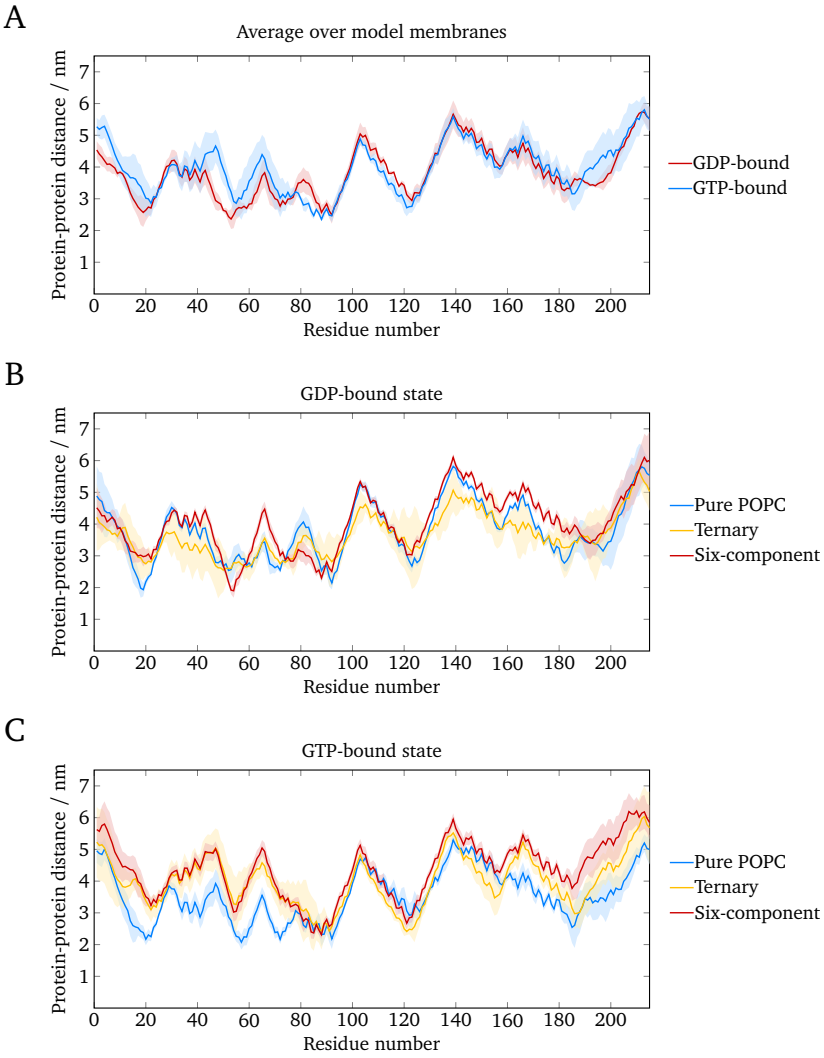


Figure 5.8: Protein-protein distances between associated Rab5 monomers averaged over A) the different model membrane systems and over both monomers in the B) GDP- and C) GTP-bound nucleotide states. The shadowed lines represent the standard deviation between the different monomers and for A) the different membrane systems.

Protein-protein distances between the switch II regions were neither influenced by the membrane composition nor by the nucleotide state. Amino acid residues that exhibited shortest protein-protein distances were located within the following regions:

1. a disordered coil and N-terminal part of the first β-strand (residues Lys^{17}-Lys^{22}),
2. the β-strand of switch I (residues Phe^{45}-Leu^{58}),
3. the coil and short α-helix (α2) of switch II (residues Tyr^{82}-Ala^{95}),
4. the C-terminal part of α-helix 3 (α3) (residues Leu^{118}-Ser^{123}), and
5. an extended disordered region of the HVR (residues Lys^{180}-Arg^{197}).

Interactions in the N-terminal β1 strand and the HVR were predominately found in the GDP-bound states. Three regions were identified as major Rab5 dimerization interfaces in both, Rab5(GDP) and Rab5(GTP) dimers, namely the switch I β interface (β2), the switch II helix α2, and α3. Largest differences originating from the membrane composition were observed in the inter-switch region close to residues Cys^{63}-Gly^{78} and the C-terminal HVR. The inter-switch region comprises β3 of the Rab5 G domain which was rather far from the dimer interaction face in all membranes but farthest in the six-component membrane. The interaction regions identified by protein-protein distance analysis were highlighted in the final conformations yielded by MD simulation (Figure 5.5) to allow a visual inspection of the different complex interaction modes. As mentioned above, the most striking difference between the GDP- and GTP-loaded states was the relative orientation of the monomers with regard to the distance from the membrane surface (see appendix, Figure B24). In the GTP-bound states, one monomer G domain approached the bilayer surface with the dimer interaction sites positioned on a distal side from the bilayer surface, while the other monomer remained in a perpendicular orientation with distances $\geq$4.2 nm above the membrane. In contrast, both monomers remained in a similar position above the bilayer surface in the GDP-bound states. This was also reflected by a smaller standard deviation between the individual GDP-bound Rab5 monomers compared to the GTP-bound dimers. Moreover, in Rab5(GDP), a larger number of C-terminal residues was found within a radius of $\leq$2 nm above the membrane surface (residues 181-215). In Rab5(GTP), residues 197-215 were found in close proximity to the bilayer surface. The protein-membrane distances for both, Rab5(GDP) and Rab5(GTP) were very similar in the ternary and six-component membranes, but different in pure POPC. Especially, in the GTP-bound dimers, protein-membrane distances were significantly smaller in pure POPC compared to the other bilayers. The switch II α2 contributed various residues to complex formation predominately, but not exclusively, in the GTP-bound dimer, and thus, represented one of two major interaction sites. The other important interaction site was formed by two β-strands and the connecting β-turn of the Rab5 switch I region. Dimerization via the switch I β-interface was primarily observed in the Rab5(GDP) dimers. This may be explained by the significant differences between the switch I structures of Rab5(GDP) and Rab5(GTP). The N-terminal region of switch I forms an additional β-strand (β2') in the GDP-bound state which is absent in Rab5(GTP) (see section 4.1). This β2'-strand was largely involved in dimerization in Rab5(GDP). Regarding the orientation of the α2 and β-interface, four predominant interaction modes were observed (Figure 5.9):

1. The switch II α2 of both monomers were arranged in one plane perpendicular to each other (e.g. final conformations of Rab5(GDP) in the six-component membrane and Rab5(GTP) in pure POPC).
2. The switch II α2 of both monomers were aligned in an anti-parallel orientation (e.g. final conformation of Rab5(GTP) in the six-component membrane).
3. The C-terminal part of the N-terminal switch I β-strand (β2') plus β-turn of one monomer interacted with residues of the switch I β-turn of the second monomer (e.g. final conformations of Rab5(GDP) in pure POPC and in the ternary mixture).

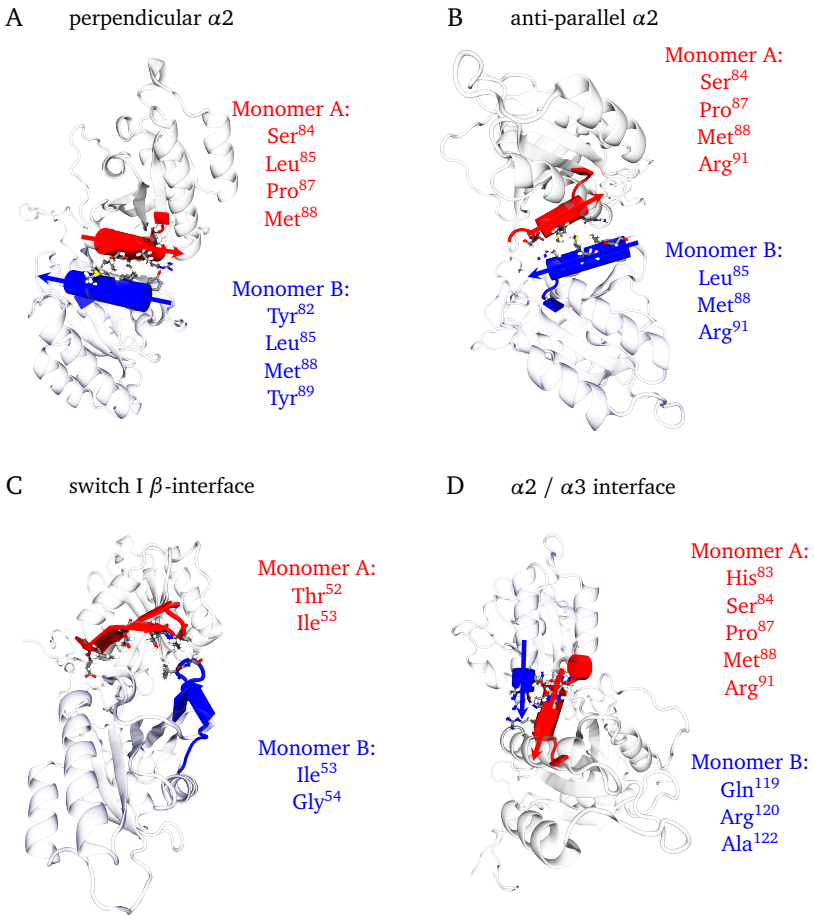


Figure 5.9: Major types of dimeric interactions in Rab5(GDP) and Rab5(GTP) comprise contacts between A) the switch II α2 of both monomers arranged perpendicular to each other, B) α2 of both monomers arranged in an anti-parallel manner, C) the N-terminal switch I β-strand (β2') of one monomer and the adjacent β-turn of the second monomer, and D) the α2 of one monomer and the α3 C-terminal region of the second monomer. Residues of monomer A and monomer B are colored in red and blue, respectively.

4. The α2 of one monomer interacted with the C-terminal part of α3 of the second monomer (e.g. final conformation of Rab5(GTP) in the ternary mixture).

Interactions of type 3 between the β-strands of switch I were in some cases accompanied by contacts between the switch I β-strand of one monomer and α2 of the second monomer.

A large number of residues provided for dimerization were of polar nature, namely Thr^{52} in switch I, Tyr^{82}, Ser^{84}, Tyr^{89} in switch II and Gln^{119} in α3, which participated in the formation of hydrogen bonds. Additionally, hydrogen bonds were formed between terminal residues, e.g. Asn^{192} with Thr^{202} or Glu^{185} with Arg^{8}. The positively charged amino acids His^{83}, Arg^{91} of switch II and Arg^{120} in α3 were found to be involved in electrostatic interactions between both monomers. In the GDP-loaded states, salt bridges were predominately formed by Glu^{47} and Glu^{50} from switch I of one monomer with Lys^{17}/Lys^{22} in a disordered region or with Arg^{91} in the α2-helix of the second monomer. Independent of the nucleotide state, salt bridges were most frequently formed between Glu^{80} and Arg^{81} as well as between Glu^{80} and Arg^{120} in α2 and α3. Binding was furthermore assisted by hydrophobic residues which sought for minimizing their solvent contacts, e.g. Ile^{53} of switch I, Leu^{85}, Met^{88} of switch II and Ala^{122} in α3.

The distribution of the relative monomer orientation defining angles was investigated to possibly reveal nucleotide state- and membrane composition-dependent structural properties over the whole 200 ns MD production simulation (Figure 5.10). The relative binding position of monomer B to

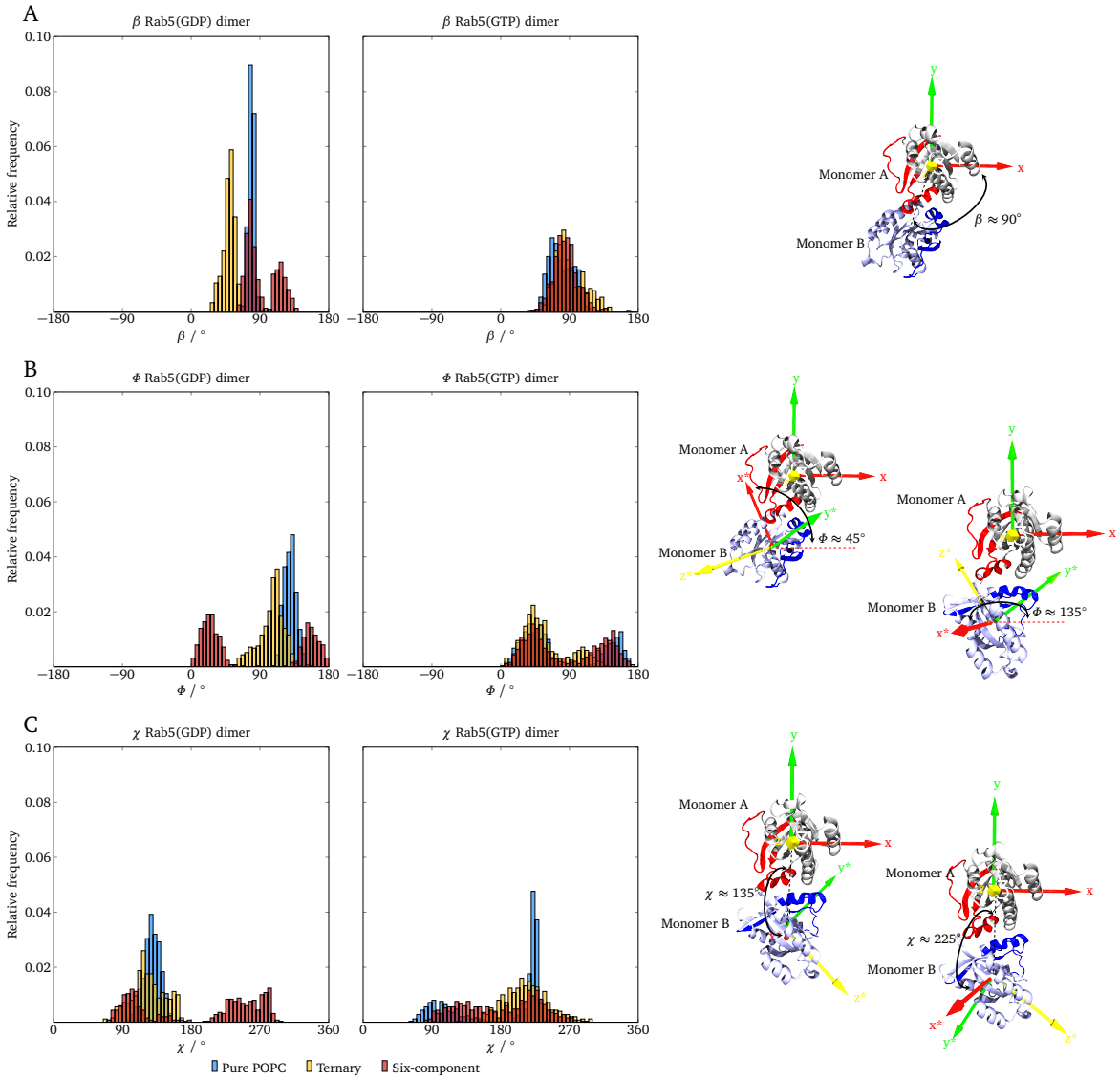


Figure 5.10: Characterization of the dimer interaction modes by A) the angle β, B) the phase Φ and C) the angle χ. Histogram data are shown for Rab5(GDP) (left) and Rab5(GTP) dimers (right) in pure POPC (blue), the ternary mixture (yellow) and the six-component membrane (red) during 200 ns of MD simulation. Right panel: Visualization of the orientation angles for the Rab5(GTP) dimer in the six-component membrane.

monomer A was described by the angle β. In case of Rab5(GDP) in pure POPC and Rab5(GTP) in all three investigated model membranes, there was only one population with β fluctuating around an average of 90°. This corresponded to a monomer B position directly facing switch II and α3 of monomer A. Some populations of Rab5(GDP) in the ternary mixture and in the six-component membrane revealed β values off 90°. For β<90°, monomer B was facing the monomer A α3-helix and for β>90°, monomer B was rather turned towards the switch I and II regions of monomer A. This was consistent with the observation that switch II α2 was involved in dimerization independent of the nucleotide state, either in one monomer or even in both monomers. The phase angle Φ described the monomer B rotation around its own z* axis and determined the position of the monomer B dimerization sites relative to monomer A. For both Rab5(GDP) and Rab5(GTP), Φ was broadly distributed between 0 and 180°. In the uncharged membranes, the GDP-loaded states exhibited a unimodal distribution of Φ with average values of approximately 135° in pure POPC and 110° in the ternary mixture. Rab5(GDP) in the six-component membrane as well as Rab5(GTP) in all model membranes exhibited a bimodal distribution of phase angles with peaks at 45° and 135°. A phase angle of 45°

represented a monomer B orientation with both, α2 and the C-terminal part of α3 turned towards monomer A. The α3-helix moved to the monomer B side averted from monomer A with increasing Φ. Simultaneously, α2 and the β-strands of switch I rotated to a position directly opposite to the monomer A dimerization interfaces for Φ=135°. The angle χ represented the angle under which monomer B bound to monomer A with respect to the plane formed by the monomer A principal x and y axes. Thus, it determined the orientation of the monomer B dimerization sites relative to the monomer A α2. That is, for χ=225° both monomer α2 were almost in parallel to each other, whereas for smaller χ the helices adopted a perpendicular orientation. χ was broadly distributed for Rab5(GDP) in the six-component membrane and Rab5(GTP) in all bilayers and adopted values between 80° and 300°. For Rab5(GDP) in the uncharged membranes, the distribution was narrower and χ fluctuated around approximately 135°. Rab5(GDP) in the six-component membrane was predominately found in two conformations with $\chi \approx 100°$ and $\chi \approx 250°$. Compared to Rab5(GDP), larger χ values were more frequently found for Rab5(GTP) dimers in all membranes, i.e. the orientation of both α2 in parallel to each other was more likely in the GTP-loaded states.

Although Rab5 dimerization as a prerequisite for further effector protein binding is not described in literature so far, it is discussed in case of Ras. Ras GTPases activate kinases of the Raf (rapidly accelerated fibrosarcoma) family in the MAPK signaling pathway for which a side-to-side dimerization was reported.[265] It is assumed that Ras(GTP) dimerization supports the self-association of Raf and thus, activates signaling cascades that promote cell proliferation.[246] If dimerization is a proper mechanism in Rab5 signal transduction the accessibility of the monomers to binding partners is of great importance. Thus, the SASA of the GDP- and GTP-bound Rab5 dimers in the different model membranes was investigated (Figure 5.11). Only minor differences were observed between the individual monomers concerning the solvent-accessibility of the switch regions, the C-terminal HVR and the entire protein. Independent of the membrane composition, the switch I region SASA was larger in the GTP-loaded states compared to the GDP-bound dimers. This was consistent with the finding that the switch I β-interface was primarily involved in dimerization in GDP-bound, but not in GTP-bound Rab5. On the other hand, the solvent accessibility of the switch II region was slightly larger in the GDP-bound dimers. This may be explained by the fact, that protein-protein interactions via the switch II α2-helix were observed in both, Rab5(GDP) and Rab5(GTP), but dominated in the active Rab5 dimers. The SASA of the C-terminal HVR and the entire protein was slightly larger in

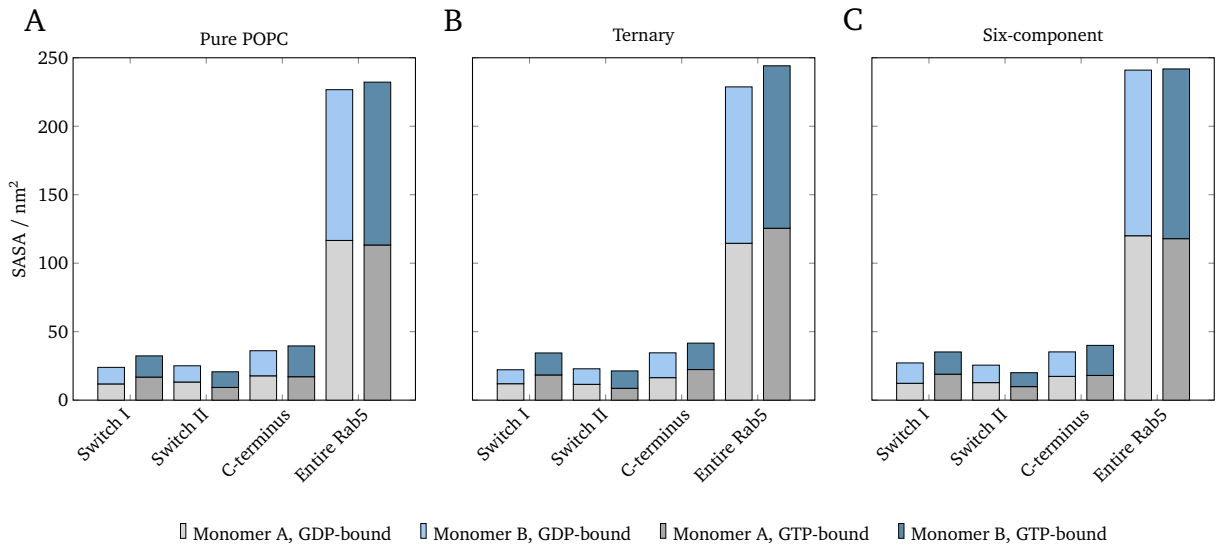


Figure 5.11: SASA of the Rab5 switch regions, the C-terminal region (residues 185-215) and the entire protein of the individual associated monomers (monomer A: gray, monomer B: blue). Data are shown for the protein GDP- (light colors) and GTP-loaded (dark colors) states in A) pure POPC, B) the ternary mixture, and C) the six-component membrane.

the GTP-bound compared to the GDP-bound dimers. In general, membrane composition-dependent differences in the protein SASA were negligible. The buried SASA was calculated in order to further characterize the protein-protein interface (Figure 5.12). The buried SASA varied between 7.7 nm^2 for Rab5(GTP) in the six-component membrane to 17.3 nm^2 for Rab5(GDP) in the ternary mixture. The buried area was significantly larger in the GDP-loaded states compared to the GTP-loaded dimers in all investigated model membranes. Furthermore, the buried SASA of the respective nucleotide state was smaller in the charged six-component membrane compared to the zwitterionic bilayers. In 1999, 75 different protein-protein complexes were investigated regarding the size of their recognition sites yielding a buried area of $(16 \pm 4)\,nm^2$ as the 'standard-size' of the interfacial region.[266] The recognition site size is determined by conformation changes accompanying protein-protein association. For values of approximately 16 nm^2, dimerization was expected to introduce rather small conformational changes. Conformational changes due to the arrangement of Rab5 monomers in a dimer were larger in the GDP-bound states than in the GTP-bound states and also in the uncharged bilayers compared to the negatively charged six-component membrane. A closer look at the specific protein-protein interactions revealed that in case of Rab5(GDP) in pure POPC and in the ternary mixture, which both revealed the largest buried SASA, dimerization occurred predominately via the switch I β-sheet and β-turn. This rather flexible region underwent comparatively large conformational changes upon dimerization. For Rab5(GDP) in the six-component membrane and Rab5(GTP) in pure POPC, the buried SASA was almost identical, which agreed well with the observation that both structures predominately interacted via the switch II α2-helix oriented perpendicular to each other. Thus, the conformational changes associated with a dimerization via α2 were significantly smaller compared to dimerization via the switch I β-interface. Consequently, least conformational changes occurred for Rab5(GTP) in the six-component membrane where dimerization was achieved via an anti-parallel orientation of α2 of both monomers. These findings indicated that dimerization may be more stable in the GDP-bound states than in the GTP-bound states. Calculation of the Rab5 dimer electrostatic binding energies revealed negative values for the Rab5(GDP) and Rab5(GTP) dimers in pure POPC, namely $-26.78\,kJ\,mol^{-1}$ and $-68.50\,kJ\,mol^{-1}$ (Table 5.1). In all other cases, the interaction showed a significant repulsive character, ranging from $48.82\,kJ\,mol^{-1}$ for the GTP-bound dimer in the ternary mixture to $80.41\,kJ\,mol^{-1}$ for the Rab5(GTP) dimer in the six-component membrane.

The binding energy calculation using APBS considered only the electrostatic component of the dimer binding energy, that is, the contribution of polar solvation or more specifically, the change in desolvation energy upon binding. This corresponds to the energy change related to the displacement of water molecules between both monomers to allow the protein-protein interaction. However, the

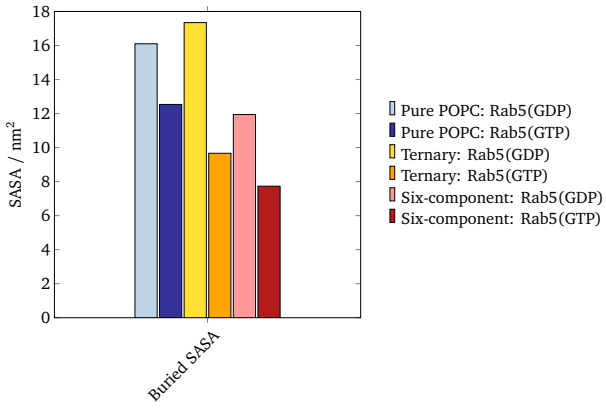


Figure 5.12: The SASA buried at the protein-protein interface provided a measure for the conformational changes associated with dimerization. GDP- (light colors) and GTP-loaded (dark colors) states in pure POPC, the ternary mixture and the six-component membrane are colored in blue, yellow and red, respectively.

Table 5.1: Binding energies of the Rab5 dimers in different model membranes.

Binding energy / kJ mol^{-1}	GDP-bound	GTP-bound
Pure POPC	-26.78	-68.50
Ternary	70.46	48.82
Six-component	68.95	80.41

internal energy changes upon binding and the entropic term were not considered. Obviously, there is a significant polar desolvation penalty associated with Rab5 dimerization. The apolar surface tension energy contributions, which favor binding, were neglected in the calculation. Since these contributions scale with the surface area, they are expected to add an energetically favorable term to the actual binding energy, which depends on the respective buried SASA. In both Rab5(GDP) and Rab5(GTP) dimers, the buried surface areas were large enough to yield negative total binding energies after including apolar surface tension contributions. The GDP-bound dimers exhibited larger buried SASA values compared to the GTP-bound dimers, resulting in more energetically favorable states. These findings indicate that Rab5(GDP) may form more stable dimers compared to Rab5(GTP).

Dimerization of membrane-bound Rab5 may represent a mechanism to impede the tilting of the monomeric Rab5 G domain toward the bilayer surface. Especially for Rab5(GDP), this was associated with a partial burial of the switch regions as shown in chapter 4. Consequently, the question arose whether dimerization could improve the accessibility of the Rab5 switch regions which would then affect effector protein binding via the switch regions. The SASA of the monomers arranged into dimeric structures was thus compared to the SASA of monomeric Rab5 in the six-component membrane from chapter 4 (see appendix, Figure B25). Slight differences were observed regarding the solvent-accessibility of the switch regions in monomeric and dimeric Rab5. The switch region SASA was only marginally larger for Rab5(GDP) in the dimer but lower for Rab5(GTP) compared to monomeric Rab5. As seen above, in the Rab5(GDP) monomer, the switch regions were screened from solvent due to their position partly buried between the membrane and the protein G domain. The same effect, i.e. a partly shielding of the switch regions from surrounding solvent, was achieved by the protein-protein interactions in the Rab5 dimer, independent of the nucleotide state. The SASA of the C-terminus as well as of the entire protein was almost identical for monomeric and dimeric Rab5. As mentioned above, Rab5 dimerization is realized via the switch I β2'-strand, the switch II α2 as well as α3. Consequently, the switch regions are partly screened from solvent. It needs to be clarified whether this is compatible with the recognition by regulatory and effector proteins. Rab5(GTP) effector recognition is primarily achieved via the invariant hydrophobic triad which comprises residues Phe^{57} in switch I as well as Trp^{74} and Tyr^{89} in switch II.[73,267] The asymmetric structure of the GTP-bound Rab5 dimers enabled the hydrophobic triad of one monomer to be effector-accessible in the ternary and six-component membranes (see appendix, Figure B26). In the GDP-bound dimers, those residues were buried at the dimerization interface. This indicated, that dimerization via the α2/α3 interface or an anti-parallel α2 orientation favored the accessibility of the hydrophobic triad and thus effector recognition. On the other hand, dimerization via the β-interface or perpendicular α2 yielded dimeric complexes in which Phe^{57}, Trp^{74} and Tyr^{89} were rather shielded from the solvent environment. Dimerization via the β-interface, like primarily in the GDP-bound Rab5 dimers, enabled a greater solvent exposure of the switch II residues. Physiologically, this makes sense since Rab5 interactions with RabGDI are essential in the GDP-bound state and usually occur via the switch II region (see chapter 6).

The present MD study provided evidence that Rab5 monomers interact in a way that allows formation of homotypic dimers. Comparison of the Rab5 protein-protein interactions to other related small GTPases on a molecular level may help answering the question whether Rab5 dimerization is feasible in nature. Recently, a combined experimental and computational study investigated weak protein-

protein interactions in K-Ras dimers.[250] The major interaction face was assigned to K-Ras helices 3 and 4 (h3/h4) from which various interfacial residues formed stable intermolecular electrostatic interactions (Figure 5.13).

However, dimeric K-Ras crystal structures are lacking the C-terminal HVR and thus, do not consider membrane binding. In the K-Ras dimer displayed in Figure 5.13A, both K-Ras monomer HVRs are pointing in opposite directions which makes binding of the dimer to one and the same membrane rather unlikely. In K-Ras the h3/h4 interaction site involved residues His^{94}-Tyr^{137} which correspond to residues Lys^{112}-Asn^{155} in Rab5 constituting the C-terminal part of α3, a buried β-strand and α4. Sequence alignment of this section revealed a low similarity between K-Ras and Rab5 which explains why, in contrast to K-Ras, only part of α3 but not α4 was involved in Rab5 dimerization. An earlier study observed K-Ras dimerization with nucleotide state-dependent stability and confirmed two major dimerization sites using various experimental techniques.[258] Accordingly, K-Ras dimerization occurred predominately in the GTP-loaded state via a β-sheet extension in switch I and the h3/h4 helices in switch II. Involved residues in K-Ras dimerization correspond to Rab5 residues Leu^{38}-Thr^{61}, Ala^{109}-Asn^{125} and Glu^{147}-Asp^{154}, i.e. the first including Rab5 switch I β-strands and the second coinciding with Rab5 α3. Therefore, Rab5 dimerization is partly distinct from K-Ras dimerization regarding the involvement of specific regions. Thus, the question arose whether the closely related Ras isoforms N-Ras and H-Ras show more significant similarities to the Rab5 dimerization interfaces. Membrane-anchored N-Ras was analyzed by Förster resonance energy transfer (FRET) measurements and MD simulations which disclosed dimer interactions of helices h4 and h5 as well as the β2-β3 loop.[248] Sequence alignment revealed that N-Ras residues Asp^{47}, Glu^{49} (β2-β3 loop), His^{131}, Lys^{135}, Gly^{138} (h4) and Asp^{154}, Arg^{161}, Arg^{164}, Met^{168} (h5) correspond to Rab5 residues Asp^{65}, Thr^{67} (switch I), Gln^{149}, Asp^{153}, Ser^{156} and Glu^{172}, Lys^{179}, Pro^{182}, Pro^{189}. The latter patch representing N-Ras h5 complies with a rather disordered region in Rab5 which formed protein-protein interactions exclusively in Rab5(GDP) dimers. For H-Ras, mutation studies showed that Tyr^{64} located in switch II is essential for dimerization, either by being part of the dimer interface or by being allosterically coupled to the interaction area.[249] H-Ras Tyr^{64} corresponds to Rab5 Tyr^{82} in switch II α2 which indeed was found to be involved in dimerization of all systems investigated in the present study. Furthermore, analysis of the H-Ras dimerization equilibrium revealed that the HVR lipidation as well as the membrane lipid composition did not influence the degree of dimerization and that the nucleotide state dependency was only minor. The comprehensive studies on Ras small GTPases provided evidence that dimer formation is isoform-specific as apart from the β-sheet dimer interface which is similar in all Ras isoforms, K-Ras dimerization occurs preferentially via h3/h4

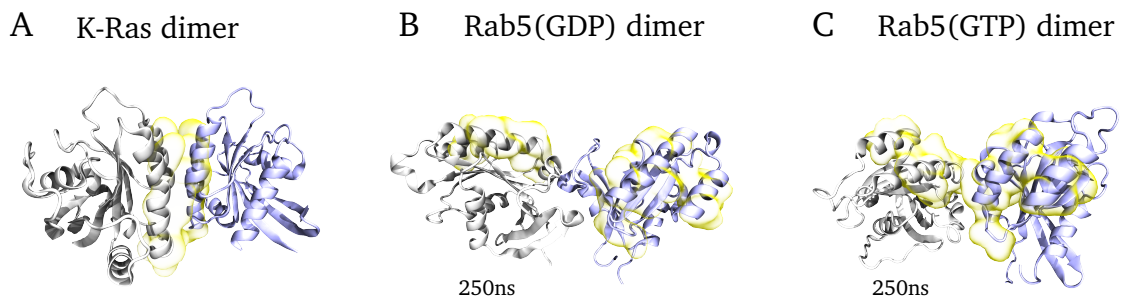


Figure 5.13: A) Structure of dimeric K-Ras co-crystallized with GDP (from PDB entry 4LV6[268]) compared to Rab5 dimers in the B) GDP- and C) GTP-loaded states, bound to the six-component membrane (only G domain residues are displayed here).The yellow regions indicate the h3/h4 dimerization interface of K-Ras which is not located at the dimerization site in Rab5 dimers.

whereas H-Ras and N-Ras associate predominately via h4/h5.[269] Rab5 dimerization is similar to Ras dimerization as protein-protein interactions occur via helical and β-sheet interfaces, however, the specifically involved regions differ between both GTPase subfamilies. Rab5 dimerization interfaces mainly comprise the switch I β2'-strand, the switch II α2-helix and the α3-helix.

In 2001, Rab5 dimers have been observed using chemical cross-linking assays.[256] Dimerization occurred in wild-type Rab5 and constitutively active Q79L mutants, however, mutation of Arg^{81} to alanine abrogated dimerization in a GTP-dependent manner. This is in agreement with the present study, which revealed interactions of Tyr^{81} in the switch II α2-helix of one monomer with α3 residues of the second monomer. However, Tyr^{81} interactions were exclusively found in Rab5(GTP) dimers.

The GDP-loaded states of Rab9[251] and Rab11[270] are examples for Rab small GTPases which have been crystallized in dimeric complexes. However, crystallization of proteins as oligomers is not unusual but does not imply a concrete biological meaning in general.[252] The usually rather disordered switch regions were stabilized by the dimer interface with hydrogen bond formation between a switch I threonine residue and a β-strand associated aspartate or glutamate. The β-strand and β-turn involved in Rab5 dimerization contain the negatively charged Glu^{47} and Glu^{50} residues which were involved in hydrogen bond formation with Lys^{22} and formation of salt bridges with Lys^{17} and Arg^{91}. The Rab11(GDP) dimer interaction face was found to occupy approximately 20 nm^2, thereby involving major parts of the switch I and switch II regions, which, thus, were inaccessible to binding partners. It was speculated that dimerization may represent an alternative mechanism to retain an inactive pool of Rab11(GDP) directly at the membrane without cytoplasmic recycling by GDI.[270] In case of Rab5, the interaction face was smaller ranging from approximately 7 nm^2-17 nm^2 (see above). The GDP-loaded states exhibited larger interfaces compared to the GTP-loaded states. This may be a consequence of a stronger involvement of disordered regions to the dimerization interfaces in Rab5(GDP), namely the additional twisted β2'-strand in switch I. Rab5(GDP) primarily dimerized via the switch I β-interface, leaving parts of the switch II accessible to binding partners. Rab5(GTP) dimers interacted predominately via the switch II α2-helix, allowing the invariant hydrophobic triad, which is essential for effector recognition, to be accessible.

The question remains whether Rab5 dimerization is of physiological relevance in cells. In case of Arf1(GTP) the physiological relevance for dimerization originates from the fact that Arf1 dimers affect membrane curvature which enables the formation and release of coated vesicles.[252] They have shown that monomeric Arf1 was not able to induce the required membrane deformation. For Ras GTPases, single-fluorophore video tracking (SFVT) experiments revealed that approximately one half of the Ras population is monomeric while the other moiety is found as membrane-bound dimers and immobile nanoclusters.[257] Due to the low affinity interaction sites on Ras monomers, dimerization is not stable in solution but possible at the membrane surface which increases the local Ras concentration. Ras dimers were reported to be preferentially formed and more stable in the GTP-loaded states, although K-Ras(GDP) dimer formation via identical interfaces compared to K-Ras(GTP)was predicted using a template-based protein-protein complex structure prediction algorithm.[258,271]The preference for GTP-loaded Ras to dimerize originates from their role in the MAPK cascade, where Ras(GTP) dimers assist Raf dimerization and thus, activation.[272] Raf is the only Ras effector which acts as a dimer and requires prior Ras dimerization. Further Ras-promoted signaling cascades, i.e. the PI3K/Akt/mTOR and the RalGDS/RasGEF (Ras-like guanine-nucleotide-exchange factor)/Ral pathway, are activated by monomeric Ras.[269] Such an effector-dependent dimerization is also conceivable in case of Rab5. However, Rab5 dimerization after binding a certain effector protein being part of a signaling cascade has not been described so far. The Rab5(GTP) effector EEA1 is involved in EE membrane tethering and forms a dimeric long parallel coiled-coil structure.[273–275] EEA1 was shown to bind Rab5(GTP) on vesicles via its N-terminal C_2H_2 Zn^{2+} finger and PI(3)P via the C-terminal FYVE domain on the target EE membrane. According to recent findings,[275] the EEA1 core forms an extended coiled-coil structure with an end-to-end distance of approximately 141 nm reaching through the cytoplasm. Binding of Rab5(GTP) induces conformational changes making EEA1 more flexible which subsequently results in a collapse of the extended structure. It is speculated

that EEA1 works like a spring and that the collapse may pull vesicle and target membrane close to each other to enable fusion. The EEA1 homodimer may be a plausible candidate requiring prior Rab5(GTP) dimerization for inducing the described allosteric conformational changes in a similar fashion described for Ras and Raf dimerization. Comprehensive MD simulations including EEA1 N-terminal dimers associated with monomeric compared to dimeric Rab5(GTP) may shed light on this possible scenario.

Rabaptin5 is another Rab5 effector known to form parallel coiled-coil homodimers at the membrane surface.[73] However, Rabaptin5 dimers complexed with two Rab5(GTP) molecules were shown to organize in a dyad-symmetric arrangement, that is the Rabaptin5 dimer was located in the middle while two Rab5 proteins bound independently to both sides of the effector. In this case, Rab5(GTP) dimers are not required.

A similar physiological role as for Rab11(GDP) may also be plausible for Rab5(GDP) dimers, as they may form an inactive pool at the membrane surface which is accessible and activatable at any time.

However, based on the present MD simulations, a nucleotide state-dependent preference for dimerization in different membranes could only be estimated. The dimerization areas were slightly larger in the GDP-loaded states, which may indicate a higher stability of interactions. In case of Rab5(GDP) dimers, both monomer G domain centers were located at the same distance above the membrane surface, whilst for Rab5(GTP) one monomer was tilted towards the bilayer surface with the second being in a perpendicular orientation, forming a lid-like structure. All observed dimers remained stable during the 250 ns MD simulations. However, it has to be emphasized that 250 ns were not sufficient to allow a complete re-orientation of the Rab5 monomers relative to each other. Thus, the protein-protein interaction sites were biased by the initial conformations. An extension of the simulation time may reveal whether the three identified interaction regions are stable or may converge into different interfaces. Computational protein-protein docking tools like ClusPro[276] or HADDOCK[277] (high ambiguity driven protein-protein docking) aim to predict the structure of protein complexes. These methods generate initial models which are subjected to subsequent MD simulations to be refined and thoroughly characterized. A comprehensive docking study was beyond the scope of the present thesis since Rab5 dimer interfaces were expected to be similar to the related Ras dimers, but nevertheless represents a promising approach for investigating protein-protein complexes. Steered MD and umbrella sampling are tools to determine the PMF of the Rab5 monomer dissociation and thereby allow an estimation of the interaction strengths in the dimers depending on the nucleotide state and the membrane composition.

5.3 Summary

In this chapter, the formation of homotypic membrane-bound Rab5(GDP) and Rab5(GTP) dimers was investigated. The main results addressed the following issues:

- Identification of dimerization interfaces in membrane-bound Rab5 monomers
- Investigation of the nucleotide state- and membrane composition-dependent dimer stability
- Evaluation of Rab5-specific protein-protein contacts compared to Ras small GTPases

Two Rab5 monomers in either the GDP- or GTP-loaded states formed a dimeric complex over a 250 ns MD trajectory. Three distinct dimerization interfaces were identified; 1) the β-strand and β-turn of the switch I region (β-interface), 2) the α-helix of switch II (α2), and 3) the C-terminal part of α3. Protein-protein interactions of the β-interface occurred solely in the GDP-bound dimer via the β2'-strand formed by the N-terminal part of switch I, which is absent in Rab5(GTP). The switch II α2-helix was primarily, but not exclusively, involved in dimerization in the GTP-bound states. The GDP-loaded dimer was further stabilized by interactions of residues located in the Rab5

N-terminal β1-strand and a disordered region of the HVR. Protein-protein contacts between the dimerization interfaces of both monomers could be assigned to four predominant binding modes: A) The α2-helices of both monomers were oriented perpendicular to each other, B) the α2-helices of both monomers were aligned in an anti-parallel orientation, C) the β-interfaces of both monomers interacted with each other, and D) the α2 of one monomer interacted with the C-terminal part of α3 of the second monomer. Different binding modes correlated with different interaction area sizes. That is, involvement of rather disordered regions within the dimerization interface (e.g. the switch I β-turn) resulted in larger buried areas compared to the monomeric protein. Consequently, due to the less ordered structure of the Rab5(GDP) switch regions compared to Rab5(GTP), GDP-loaded dimers exhibited larger buried surface areas forming slightly more stable protein-protein interactions.

Independent of the membrane composition, Rab5 dimerization was stabilized by a complex network of hydrogen bonds, electrostatic and hydrophobic interactions, and salt bridges between both monomer interfaces. Salt bridges between residues Glu^{50}-Arg^{91}, Glu^{80}-Arg^{81}, and Glu^{80}-Arg^{120} were most frequently found.

An apparent difference between the GDP- and GTP-loaded dimers was the orientation of the monomers with respect to the membrane surface. In Rab5(GDP) dimers, both monomer G domains were equally distant from the bilayer surface. In case of Rab5(GTP), one monomer G domain was tilted toward the bilayer while the other monomer covered the bent one in a lid-like manner with an almost fully extended HVR. However, in both nucleotide states, the switch regions responsible for effector and regulatory protein binding were sufficiently solvent-exposed to allow interaction with potential binding partners.

6 Rab5(GDP) membrane extraction by RabGDI

As molecular switches, Rab5 small GTPases cycle between an inactive GDP-bound and an active GTP-bound state, from which only the latter is able to recruit and activate effector proteins at the membrane. This activation cycle is coupled to a transport process comprising the reversible association of Rab5 with the early endosome (EE) membrane. While Rab5(GTP) is exclusively found at the bilayer surface, membrane-bound Rab5(GDP) is recognized by the Rab GDP dissociation inhibitor (GDI), released from the bilayer and recycled back to donor membranes in a soluble cytoplasmic complex.[278,279] The hydrophobic Rab5 GG anchor, which allows membrane anchorage in the membrane-bound states, is accommodated in a GDI prenyl binding pocket during cytoplasm transfer and thus, screened from unfavorable solvent contacts.

X-ray diffraction studies on bovine[280,281] and yeast[282,283] Rab:GDI complexes revealed insight into protein-protein interactions and the localization of the prenyl binding pocket. However, these experimental complex structures are missing the long C-terminal hypervariable region (HVR) of the Rab GTPase due to its high flexibility. Thus, to date, very little is known about the HVR conformation when complexed to GDI and the respective protein-protein contacts. In the present study, full-length human Rab5(GDP) (hereafter called Rab5) interactions with GDI in the soluble state as well as associated with the EE membrane are characterized by full-atomistic MD simulations.[279] Profound knowledge of the protein-protein interactions in the Rab5:GDI complex are of relevance due to the important role of Rab5 in EE transport and signaling (see section 2.1.4). GDI activity is of utmost importance for the delivery of Rab5 to the correct target bilayers and Rab5 recycling between donor and acceptor membranes within the EE. Distortion of the Rab5 membrane extraction process due to GDI or Rab5 dysfunction may cause severe diseases. Thus, a deep understanding of the structure and dynamics of the Rab5:GDI complex may pave the way for the development of new Rab5-specific therapeutics.

This chapter addresses complex formation between Rab5 and GDI by full-atomistic MD simulations. Specific protein-protein interactions are investigated when bound to a six-component bilayer mimicking the EE membrane as well as in the cytoplasmic complex.

The first part of this chapter introduces the model membrane composition and describes the approach of modeling human GDI by comparative modeling as well as the Rab5:GDI interaction interface. In the second section, the human Rab5:GDI complex is associated with the membrane via the Rab5 geranylgeranyl (GG) anchor and different binding modes are elucidated. Subsequently, protein-protein interactions in the soluble Rab5:GDI complex are investigated and compared to those described for the yeast homologue. Based on these findings, a concept of the Rab5(GDP) membrane extraction process guided by GDI is derived in the last part of this chapter. Moreover, the binding free energy profiles of Rab5 to GDI association as well as complex to membrane binding are determined.

Significant parts of this chapter were published in a research article by the author of this thesis.[279]

6.1 Models for human RabGDI and the early endosome membrane

For the membrane-associated Rab5:GDI complex simulations, a large bilayer patch mimicking the EE membrane was prepared consistent with the six-component bilayer model for the stand-alone Rab5 simulations. The larger bilayer size was chosen to prevent self-interaction of the complex' mirror images in PBC due to too small lateral dimensions. The composition, number of lipids and bilayer

Table 6.1: Lipid composition and size of the large six-component membrane used in the Rab5:GDI complex simulations.

Membrane system	Lipid type	Lipid number per leaflet	Mole fraction	Lateral (x,y) dimensions / nm
Six-component	POPC	204	17.7 %	25.2 x 25.3
	Cholesterol	348	30.2 %	
	PSM	72	9.4 %	
	POPE	312	27.1 %	
	POPS	120	10.4 %	
	PI(3)P	60	5.2 %	

patch size are given in Table 6.1. Bilayer preparation and modeling was done in accordance with the other model membranes (for details see section 4.2).

Currently, there is no experimental structure of human GDI available. However, PDB entry 1LV0 corresponds to a high-resolution X-ray structure of bovine GDI in complex with a GG peptide.[281] A pairwise alignment between the amino acid sequences of human and bovine GDI using the software Jalview[284] revealed a sequence identity of >98 % and a discrepancy of only seven residues between both sequences (Figure 6.1). Therefore, 1LV0 was taken as the template structure for GDI model generation with Modeller. Twenty models for GDI were derived based on the alignment of the target sequence with the template. The model with the lowest DOPE score was subjected to a loop refinement procedure in Modeller to give 100 models. These 100 structures were RMSD-based clustered into five groups. A structure belonging to the largest cluster was chosen as the final model for human GDI. Until today, it is debated where the GDI prenyl binding pocket, that accommodates the GG chains, is exactly located within the Rab:GDI complex. Evidence has been found that the two prenyl groups are located in a hydrophobic cavity formed by the GDI domain II. Such an interaction was experimentally resolved in the yeast analogue of Rab, Ypt1, and GDI.[282] The corresponding PDB entry 2BCG was therefore used as a model for the Rab5:GDI interaction and localization of GG chains. First, simulations were performed for Rab5:GDI interactions in the cytoplasm, i.e. without a membrane. With Modeller generated full-length Rab5(GDP) was superposed with Ypt1 from PDB entry 2BCG, whereas the modeled human GDI was superposed with the yeast GDI. The Rab5 GG chains were placed according to the prenyl moieties in the yeast PDB structure. Analogous to membrane-bound Rab5, topologies were built, and subsequently the Rab5:GDI system was solvated and ionized to a salt concentration of 0.15 M sodium chloride.

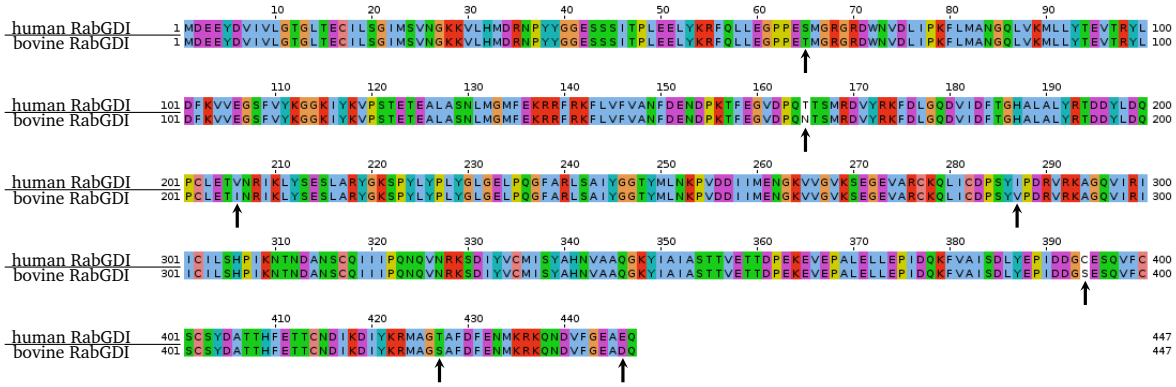


Figure 6.1: Alignment of the amino acid sequences of human and bovine GDI using Jalview. The seven dissimilar residues are highlighted with arrows.

6.2 MD simulations of the Rab5:GDI complex

In contrast to Rab5(GTP), which is exclusively found in a membrane-bound state in order to recruit effector proteins to the bilayer, Rab5(GDP) cycles between cytoplasm and membrane. However, Rab5(GDP) membrane extraction is not spontaneous but depends on binding to GDI, which releases Rab5(GDP) from the bilayer, maintains it soluble in the cytoplasm and recycles the GTPase back to donor membranes.[100]

In this section, the interactions of Rab5(GDP) and GDI when associated with and bound to the six-component membrane as well as in cytoplasm are investigated by full-atomistic MD simulations. Based on these simulations a theoretical model for the Rab5:GDI membrane extraction process is proposed. Furthermore, the free energy profiles of Rab5 GDI separation and GG anchor release from the bilayer are determined.

6.2.1 MD simulation setup

In accordance with the full-length Rab5 MD simulations, the bilayer was placed in the x, y plane and Rab5 was localized in +z direction above the membrane surface. The following simulation setups were used (Figure 6.2):

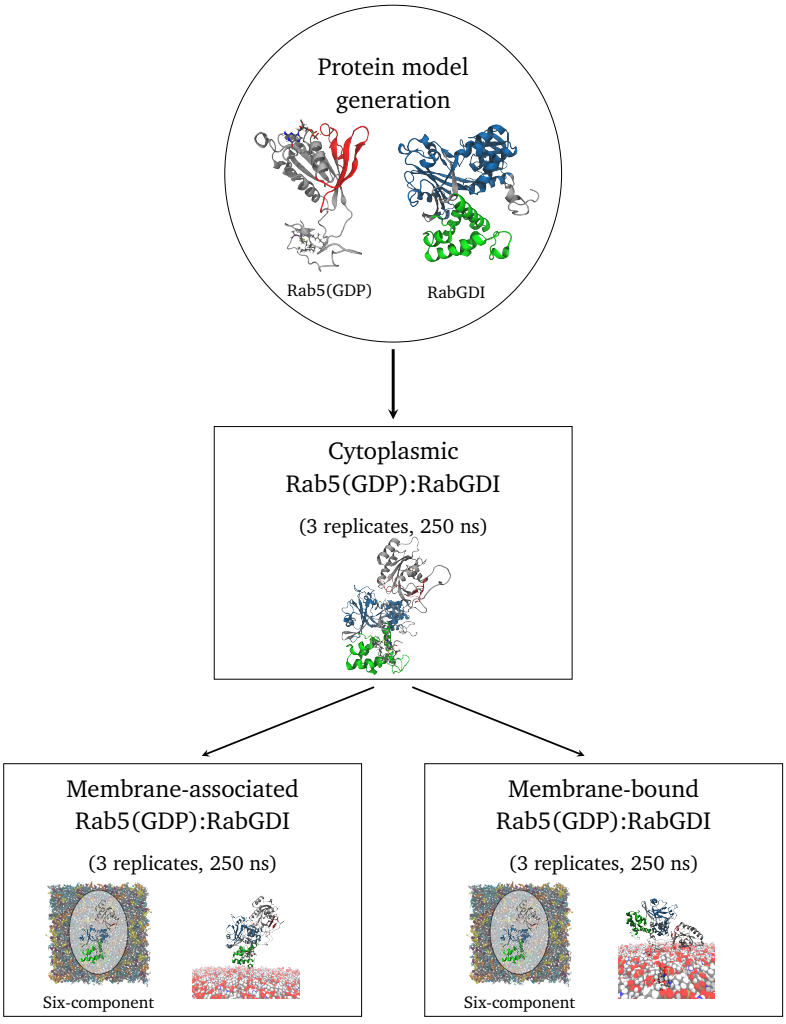


Figure 6.2: Full-atomistic MD simulations of a Rab5:GDI complex in cytoplasm as well as of membrane-associated and membrane-bound Rab5:GDI in the charged six-component bilayer.

- The Rab5:GDI complex was solvated in explicit water and ions to mimic the cytoplasmic environment (hereafter called $_{cyt}$Runs). Three independent simulations with 200 ns production time each were performed.[279]
- The Rab5:GDI complex obtained from each of the three cytoplasm simulation was arranged in +z direction 4 nm above the anionic membrane ($_{assoc}$Runs). These protein complex-membrane systems were simulated for 200 ns production time.
- Based on the GDI-free Rab5(GDP) simulations (see section 4.4.5) and the Rab5:GDI simulations in cytoplasm, Rab5 complexed with GDI was anchored to the six-component membrane via its GG anchor ($_{insert}$Runs). Three MD simulations with 200 ns production time each were performed.[279]

Furthermore, based on the final conformation of the membrane-bound complex ($_{insert}$Run1) two steered MD simulations were performed in order to probe protein-protein binding and to model the membrane extraction process. A force was applied a) to all GDI atoms or b) to Rab5 residues 5-210 (Table 6.2). Moreover, the final Rab5:GDI complex structure after membrane extraction was used as the initial conformation for two further 250 ns MD simulation in cytoplasm. These simulations were meant for revealing the solvent-exposed GG chains behavior when being neither membrane-bound nor inserted within the GDI binding pocket.

Table 6.2: US calculations setup for determining Rab5:GDI complex binding free energies.

Simulated system	Rab5:GDI complex	
Calculated PMF	Binding free energy between Rab5 and GDI	Binding free energy of GG anchor to bilayer
Membrane model	Six-component membrane	
Number of replicates per membrane	1	1
Reaction coordinate ξ	Z-distance between upper membrane layer (P atoms) and GDI COM (Cα atoms)	Z-distance between upper membrane layer (P atoms) and Rab5 COM (Cα atoms, residues 5 to 210)
Sampled distance on ξ / nm	4.26 8.03	2.26ă 14.09
Number of windows	30	92

6.2.2 Analysis of Rab5:GDI complex simulations

The protein RMSD, RMSF and SASA were calculated in agreement with the equations provided in section 4.4.3.

Three geometrical parameters were defined to allow an accurate characterization of different Rab5:GDI complex interaction modes (Figure 6.3). These were 1) the intermolecular distance d_{COM} between the centers of mass of the Rab5 G domain and GDI, 2) the distance $d_{GGpocket}$ between the GG-Cys Cα atoms and the GDI prenyl binding pocket innermost residue Met^{132}, and 3) the angle θ formed between helices H1 and H2 that opened the GG chain binding pocket. While d_{COM} provided a measure for the protein-protein association, the other two parameters focused on the characterization of the hydrophobic binding cavity regarding GG chain insertion depth and binding pocket opening.

Protein binding pockets play an important role in molecular recognition and are responsible for protein-protein or protein-ligand binding. Durrant *et al.* developed the POVME (POcket Volume

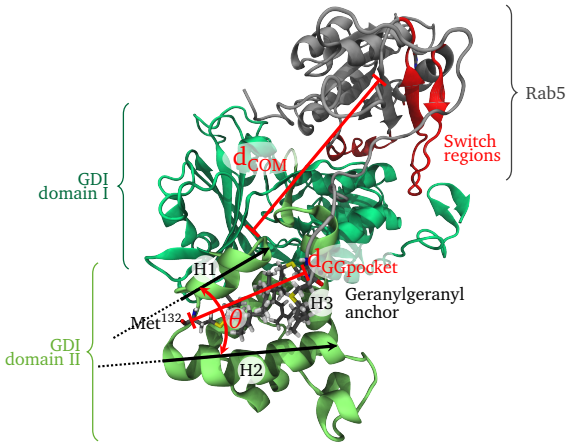


Figure 6.3: Rab5:GDI complex interaction modes were characterized by different geometrical parameters: 1) the intermolecular distance d_{COM} between the centers of mass of Rab5 and GDI, 2) the distance $d_{GGpocket}$ between the GG-Cys Cα atoms and GDI Met^{132}, the innermost residue of the prenyl binding pocket, and 3) the angle θ between helices H1 and H2 which form the prenyl binding pocket.[279]

MEasurer) algorithm[285] that allows determining pocket volumes and shapes. In the present work the revised and recently published POVME 2.0 tool was used.[286] Simply put, a potential pocket-encompassing region is filled with equidistant points, from which these points that are within a particular distance from any protein atom are removed. The remaining points are used to calculate the pocket volume and characterize the pocket shape. In the present work, pocket volume calculations were performed for the binding region of GDI which accommodates the Rab5 GG anchor when the protein complex is located in the cytoplasm.

6.2.3 Comparison of the human Rab5:GDI complex with its template structures

The model for human GDI (hereafter hGDI) was built using bovine GDI (bGDI)[281] as a template structure and yeast GDI (yGDI) complexed to Ypt1[282] as a template for the protein-protein complex interactions. Superposition of hGDI with bGDI and yGDI revealed RMSD values of 0.015 nm and 0.187 nm, respectively (Figure 6.4).[279]

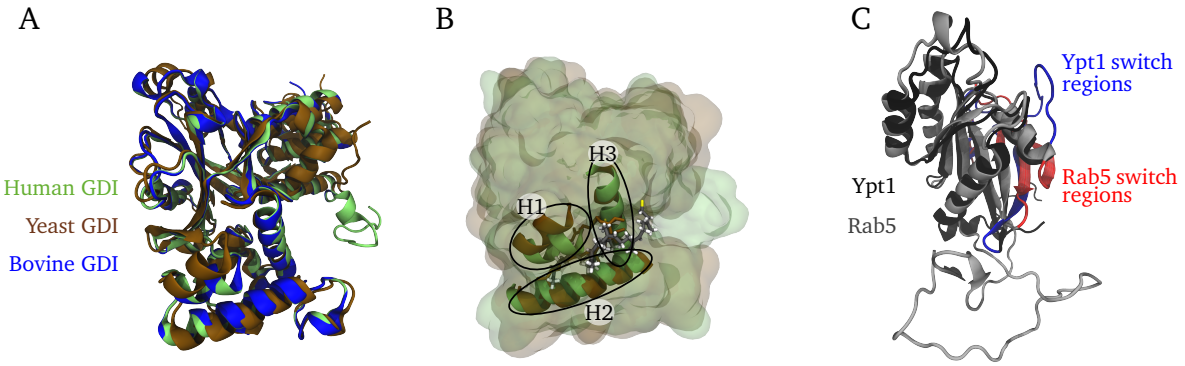
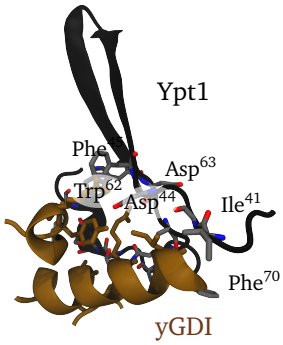


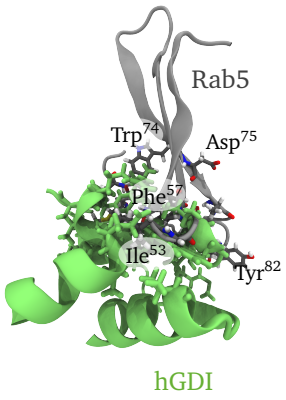
Figure 6.4: A) Superposition of GDI structures from human (modeled, green), yeast (PDB entry 2BCG, brown) and bovine (PDB entry 1LV0, blue). B) Three helices H1, H2, and H3 form the prenyl binding cavity of GDI (green: hGDI, brown: yGDI) that shields hydrophobic geranylgeranyl from water. C) Superposition of human Rab5 (gray) and yeast Ypt1 (black) with their switch regions highlighted in red and blue, respectively.[279]

Regions with highest structural deviations between hGDI and yGDI were found encompassing the GG chain binding pocket in the GDI domain II, especially within helix 1 (H1, hGDI residues Thr^{122}-Ser^{129}, Figure 6.4B) and at the rather flexible C-terminus. The former was caused by differences between bGDI and yGDI related to prenyl binding, namely the co-crystallization with a single Cys-GG ligand in case of bGDI that was localized to a prenyl binding pocket within the GDI domain I.[281] This caused the helices H1 and H2 in the GDI domain II to move closer in order to screen hydrophobic residues from the solvent. The predicted binding pocket in the bovine complex was later assumed to be incorrect.[282] Yeast GDI was crystallized with two prenyl groups located within the hydrophobic pocket opened by helices H1, H2 and H3, which was later identified by X-ray diffraction studies as the most probable binding pocket.[282,287] Since the hGDI model was derived from bGDI, the helices H1, H2 and H3 were not in the 'open' conformation. The GG chain positions as well as the complex interaction site were modeled according to the yeast Ypt1:yGDI complex. Superposition of Rab5(GDP) and yeast Ypt1 yielded a G domain RMSD of 0.479 nm (Figure 6.4C). The most prominent structural differences between Rab5 and Ypt1 occurred within the switch I region (Rab5 residues Gln^{44}-Asp^{66}). These originated from the chosen Rab5(GDP) model, which was built using Rab5 complexed with Rabaptin5 from pdb entry 1TU4 (chain A)[73] as a template. The selected Rab5 model exhibited an additional β-strand within switch I, whereas Ypt1 displayed a strongly disordered switch I region. The yeast homologue of the Rab5 C-terminal HVR (Rab5 residues Pro^{190}-Asn^{215}) was not included in the yeast Ypt1:yGDI complex crystal structure. Contacts between the GDI Rab binding platform (RBP) and the Rab binding epitope at the Rab5 switch regions represented one protein-protein interaction site within the complex.[283] The corresponding interaction forming residues in yeast and human prior to MD simulation are displayed in Figure 6.5. The interaction site exhibited many amino acid residues in switch I and switch II which were conserved between yeast and human small GTPases. Those mainly non-polar residues constituted the Rab binding epitope and exhibited almost identical physical properties in Ypt1 and human Rab5. Due to the additional β-strand in the switch I region of the initial Rab5 model, interaction with GDI comprised more contacts compared to the Ypt1:yGDI complex, which were refined during the MD simulations.

A



B



C

Ypt1		Rab5
Ile^{41}		Ile^{53}
Gly^{42}	switch I	Gly^{54}
Asp^{44}		Ala^{56}
Phe^{45}		Phe^{57}
Trp^{62}		Trp^{74}
Asp^{63}		Asp^{75}
Ala^{65}		Ala^{77}
Gln^{67}		Gln^{79}
Phe^{70}	switch II	Tyr^{82}
Thr^{72}		Ser^{84}
Thr^{74}		Ala^{86}
Ser^{75}		Pro^{87}
Ser^{76}		Met^{88}
Arg^{79}		Arg^{91}

Figure 6.5: Interactions between the GDI Rab binding platform (RBP) and residues of the Rab binding epitope located within the switch regions in A) yeast and B) humans prior to MD simulation. C) The Rab binding epitope is formed by residues of the switch I and switch II regions conserved between yeast and human.

6.2.4 Membrane-bound and membrane-associated Rab5:GDI complex

The structure and dynamics of the membrane-associated Rab5:GDI complex were investigated. Interactions between GDI and Rab5 anchored to the membrane ($_{insert}$Runs) were elucidated which represents the first step of the membrane extraction procedure. Furthermore, the Rab5:GDI complex was positioned close above the bilayer surface with the GG chains occupying the GDI prenyl binding pocket ($_{assoc}$Runs). This setup corresponds to an intermediate step when the Rab5 GG anchor is already extracted but the protein-protein complex still membrane-associated. The respective final conformations after 250 ns MD simulation are exemplarily shown in Figure 6.6.

Two additional simulations of membrane-anchored Rab5:GDI complex were set up, where steric clashes with bilayer lipids could only be avoided by placing Rab5 and GDI in a larger distance from each other, thus distorting the protein-protein interface. During these simulations Rab5 and GDI diffused from each other (see appendix, Figure B27), indicating that weak interactions between the GDI RBP and the Rab5 binding epitope were not sufficient to stably associate both proteins.

Protein-protein and protein-membrane interactions of membrane-bound Rab5:GDI complex were investigated in three MD simulations, hereafter called $_{insert}$Run1, $_{insert}$Run2, and $_{insert}$Run3. Contacts between Rab5 and GDI were initially limited to the GDI RBP and the Rab5 binding epitope in the switch regions, since the structure and positions of the Rab5 HVR and GG chains originated from the full-length GDI-free Rab5 simulations (see section 4.4.5) and were involved into bilayer contacts. Membrane-bound Rab5:GDI approached the membrane surface predominately via positively charged residues in GDI domain II (Figure 6.7A). Most interacting residues were provided by two parallel helices (Glu^{123}-Ser^{129}, Gly^{133}-Asn^{149}) as well as by an adjacent loop (Asp^{151}-Phe^{158}). Intermolecular protein-protein interactions between the GDI RBP and the Rab5 switch regions were formed mainly by a set of basic amino acids close to Arg^{68}, Arg^{70} and Arg^{98}, Tyr^{99} in GDI domain I together with residues around $helix^{GDI}$, namely Tyr^{229} and Arg^{240} (Figure 6.7B).

Besides the GDI RBP and the incorporation of the GG chains in the GDI binding pocket in cytoplasm, a third complex interaction site is provided by the C-terminus coordinating region (CCR) of GDI. This region is located at the interface of GDI domains I and II and known to interact with residues of the small GTPase HVR.[283] As a consequence of the membrane-anchored Rab5 GG chains, interactions involving the HVR were not as numerous as expected for the cytoplasmic complexes. Only a few residues from the GDI CCR, e.g. Val^{104}, Lys^{221}, formed contacts with the HVR in the membrane-

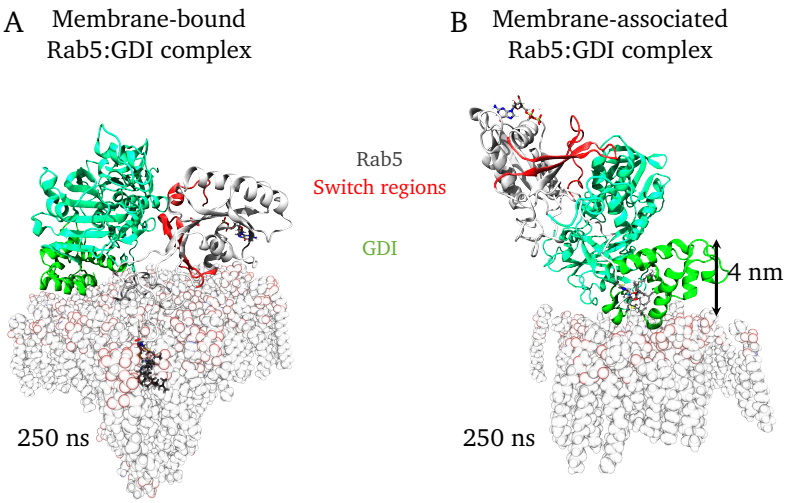


Figure 6.6: Exemplary final conformations of the Rab5:GDI complex A) with membrane-anchored Rab5 GG-chains, $_{insert}$Run2 and B) associated with the membrane surface, $_{assoc}$Run1, after 250 ns MD simulation.

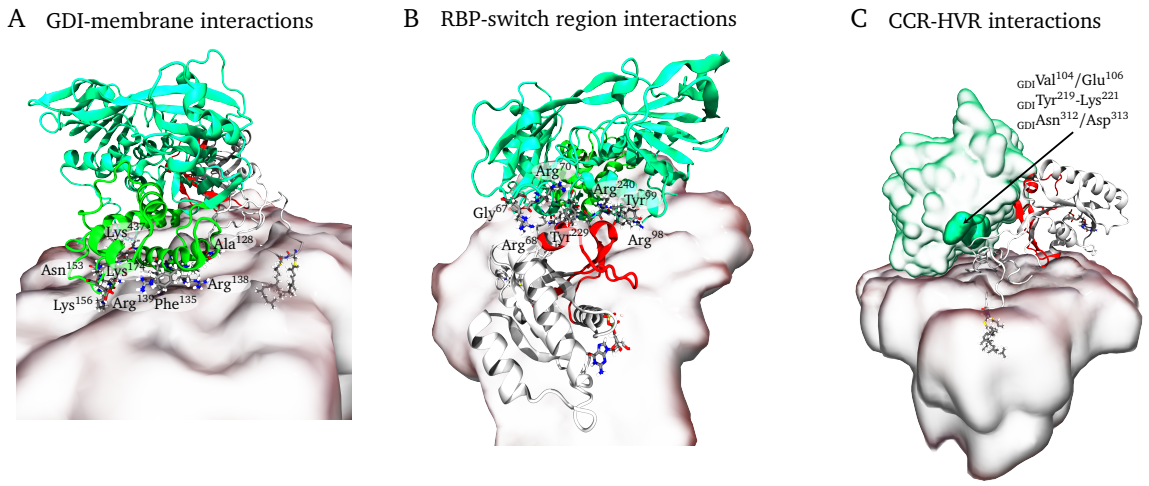


Figure 6.7: Final conformation of membrane-bound Rab5:GDI complex from insertRun2 with highlighted interactions between A) GDI domain II and membrane lipids, B) GDI RBP and Rab5 switch regions, and C) GDI CCR and Rab5 HVR. Rab5 and GDI are colored in gray/red and green, respectively.

bound Rab5:GDI complex (Figure 6.7C). However, stable GDI to membrane contacts were not established in insertRun1, for which average GDI-membrane distances of ≈4.4 nm were observed.[279]

The intermolecular distance and the prenyl binding pocket angle were displayed as functions of the GDI-membrane distance to check whether the greater GDI-membrane distances affected the relative GDI orientation with respect to Rab5 and the behavior of the prenyl binding pocket (Figure 6.8). As already mentioned, the distance between GDI domain II, which included the prenyl binding pocket and several basic residues for forming protein-membrane interactions, and the bilayer surface was largest in insertRun1 and smallest in insertRun2. Stable GDI-membrane contacts were formed in insertRun2 and insertRun3. The GDI-membrane distances were uncorrelated with the distance d_{inter} between

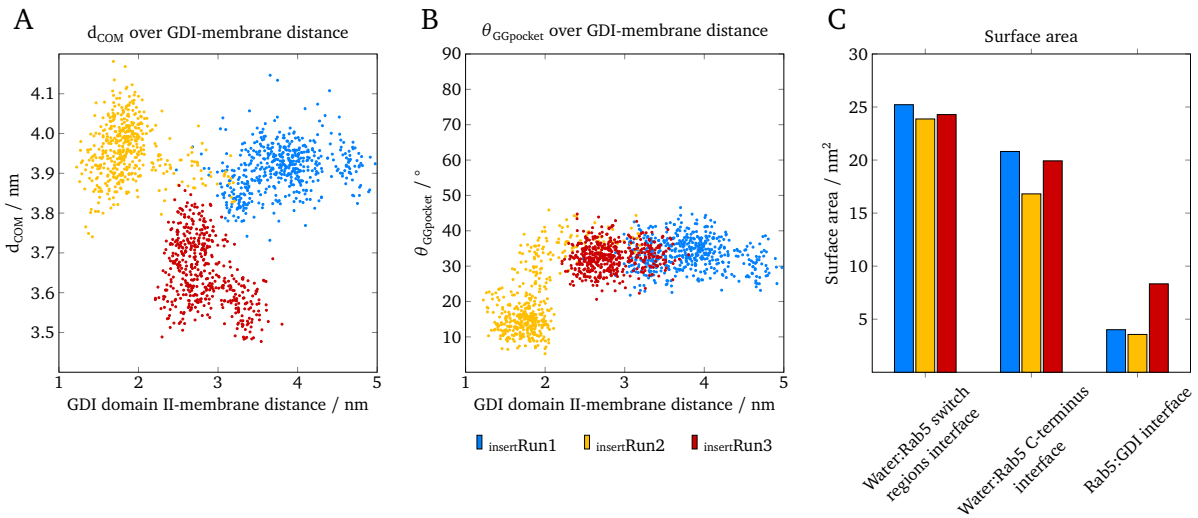


Figure 6.8: The relative GDI orientation and position with respect to Rab5 and to the membrane surface was characterized by investigation of A) the intermolecular distance d_{COM} between the Rab5 G domain and GDI as a function of the GDI domain II to membrane surface distance, B) the pocket angle θ as a function of the GDI domain II – membrane distance. C) SASA of the Rab5 switch regions and HVR as well as the Rab5:GDI interaction interface. Data from insertRun1, insertRun2, and insertRun3 are colored in blue, yellow, and red, respectively.

the Rab5 G domain and GDI, which was similar in insertRun1 and insertRun2, but smaller in insertRun3 (Figure 6.8A). Surprisingly, the close distance to the membrane surface in insertRun2 led to a smaller angle θ between helices H1 and H2 forming the prenyl binding pocket (Figure 6.8B). It may be speculated that the multitude of interactions between hydrophobic residues at the outside of the binding pocket and the bilayer lipids forced helix H2, which was closest to the membrane surface, into an almost parallel orientation relative to H1. Along with this, the SASA of the Rab5 HVR was smallest in insertRun2, indicating that contacts between the GDI CCR and the HVR contributed significantly to protein-protein interactions in this state (Figure 6.8C). In contrast, the HVR was prominently more solvent exposed in insertRun1 and insertRun3. However, besides a weakened screening of the HVR by GDI residues, a higher solvent accessibility may also be caused by less contacts with the bilayer. Compared to the other runs, the overall Rab5:GDI interaction interface was considerably larger in insertRun3; apparently mainly protein-protein interactions off the Rab5 switch regions caused a tight approach of the G domain toward GDI yielding small distances d_{COM}.

When the Rab5:GDI complex was placed above the six-component membrane with the GG chains inserted into the binding pocket all three individual simulations (hereafter called assocRun1, assocRun2, and assocRun3) predicted stable interactions between GDI and the bilayer lipids. Regions most frequently involved in forming GDI to membrane contacts were identified by analysis of the GDI-membrane distances (Figure 6.9). GDI residues that came closest to the bilayer surface were within region Pro^{120}-Gly^{220}, which represents the entire GDI domain II responsible for forming the hydrophobic cavity that accommodates the GTPase GG chains. Moreover, GDI residues between Ile^{300}-Leu^{370} closely approached the membrane surface in assocRun1 and assocRun2. More precisely, regions responsible for GDI-membrane interactions comprised a set of positively charged residues within GDI domain II, e.g. Arg^{138}, Arg^{141} and Lys^{142}, as well as within GDI domain I, namely Lys^{115} and Lys^{309} (Figure 6.10A). This was expected, since the role of basic residues for membrane contact formation is well described in literature[232,236] and was also illustrated for GDI-free full-length Rab5 (see section 4.4.5). The close approaching of GDI domain II including the prenyl binding pocket represented a plausible orientation suited for bringing the binding pocket into close proximity to the bilayer.

Interactions between the GDI RBP and the Rab5 binding epitope within the switch regions were preferentially formed by the helix^{GDI} (residues Gly^{230}-Gly^{246}) along with GDI residues at the protein N-and C-termini (Figure 6.10B). The Rab5 HVR was mainly coordinated by GDI CCR residues, e.g. Lys^{89} and Ser^{215}-Gly^{220}, as well as by Ser^{121}-Glu^{123} and rather C-terminal Lys^{435}-Gln^{438} (Figure 6.10C). Early crystallization studies of bovine GDI identified a mobile loop located in domain II (centered around hGDI residues Leu^{216}, Arg^{218} and Tyr^{219}), the so-called mobile effector loop (MEL), that was proposed to be essential for GDI membrane interaction[288] in un-complexed GDI. The present MD simulations revealed significant contributions of the MEL in contacting the Rab5

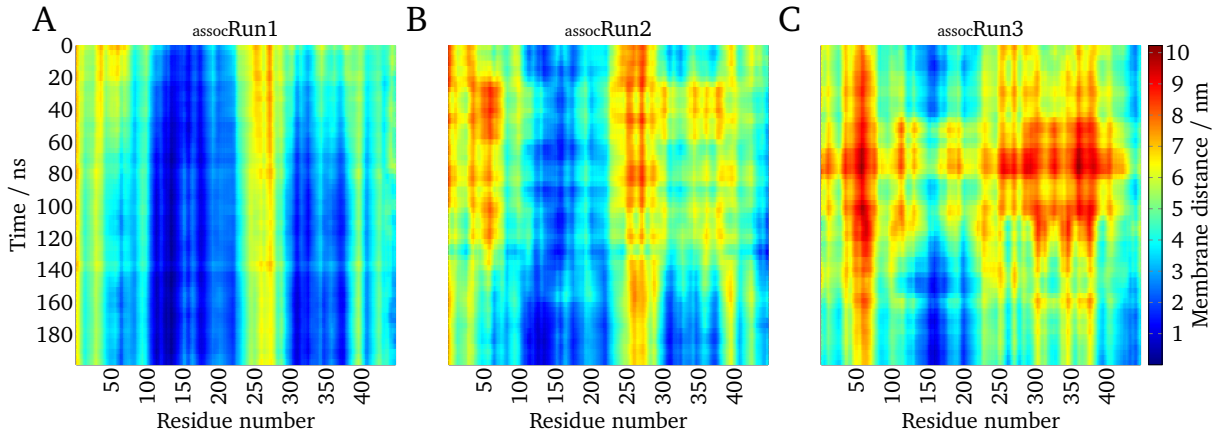


Figure 6.9: The z distances between each GDI amino acid and the membrane surface were monitored over 200 ns in A) assocRun1, B) assocRun2, and C) assocRun3.

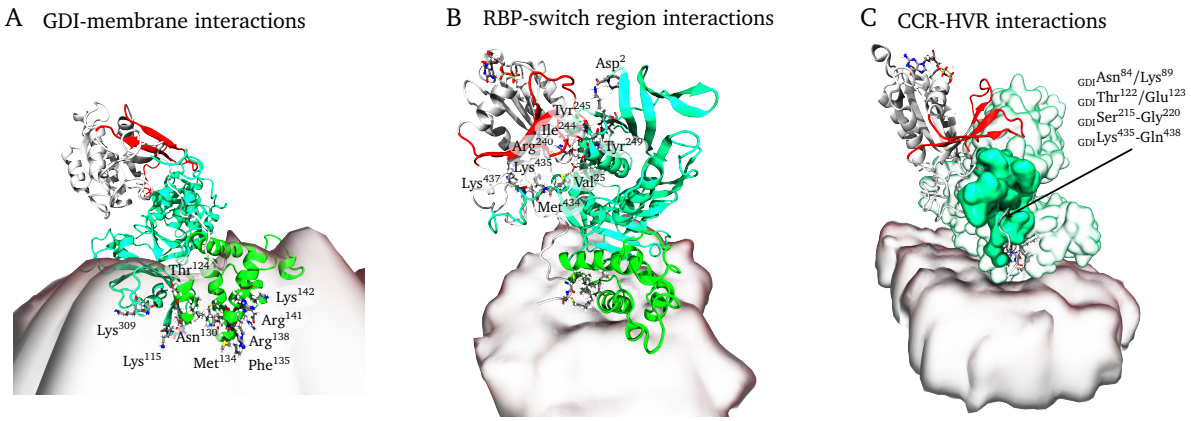


Figure 6.10: Final structure of the membrane-associated Rab5:GDI complex from $_{assoc}$Run1 with highlighting interactions between A) GDI domain II and membrane lipids, B) GDI RBP and Rab5 switch regions, and C) GDI CCR and Rab5 HVR. Rab5 and GDI are colored in gray/red and green, respectively.

HVR, which explains its need for GDI membrane targeting and the extraction of Rab5 from the membrane.

6.2.5 Rab5:GDI interactions in the cytoplasm

The final conformations of the three cytoplasmic Rab5:GDI complex simulations (hereafter called $_{cyt}$Run1, $_{cyt}$Run2, $_{cyt}$Run3) revealed two distinct Rab5:GDI binding modes with regard to the relative orientation of the Rab5 G domain toward GDI (Figure 6.11). Taking the G domain-GDI distance and the number of interactions between the Rab5 binding epitope and the GDI RBP into account, tightly bound and loosely bound states could be distinguished.

However, in addition to the G domain orientation, the complexes differed regarding their GG chain position within the prenyl binding pocket (Figure 6.12). The interatomic distances (Figure 6.12A) confirmed the assignment of the complex structures from $_{cyt}$Run1 and $_{cyt}$Run3 to the tightly bound state, whereas the complex from $_{cyt}$Run2 rather represented the loosely bound state. The GG anchor insertion depths in the binding pocket indicated three different states (Figure 6.12B) that may illustrate transition points in the process of prenyl chain extraction. The GG chain position within the hydrophobic cavity seemed to be independent of the G domain-GDI orientation as the tightly bound states revealed GG chains which were either deeply inserted ($_{cyt}$Run1) or almost extracted from the binding pocket ($_{cyt}$Run3).

As expected, interactions between the Rab5 binding epitope and the GDI RBP were more frequent in the tightly bound state (Figure 6.13). Residues of the Rab5 switch I loop connecting the β-strands were mostly involved in contact formation with the GDI N- and C-termini as well as with the helix forming the largest part of the GDI RBP (helix^{GDI}, GDI residues Gly^{230}-Gly^{246}). Rab5 switch II region residues formed numerous contacts with a short GDI loop (GDI residues Pro^{37}-Tyr^{39}), the entire helix^{GDI} and an affiliated extended region (GDI residues Gly^{247}-Leu^{251}). In the loosely bound state, interactions were restricted to contacts between both Rab5 switch regions and helix^{GDI} and a short loop region (GDI residues Arg^{68}-Asn^{73}).[279]

The question why GDI exclusively binds inactive Rab(GDP) and not Rab(GTP) may be answered with regard to the GDI RBP to Rab binding epitope interaction site. As seen above (see section 4.4.5), in Rab(GDP), the switch regions were rather disordered and more flexible compared to Rab(GTP), resulting in a better fit to the GDI RBP.[278]

A distinction between different interaction modes was not only possible with respect to the Rab5 G domain orientation but also regarding the incorporation of the GG chains in the GDI binding pocket.

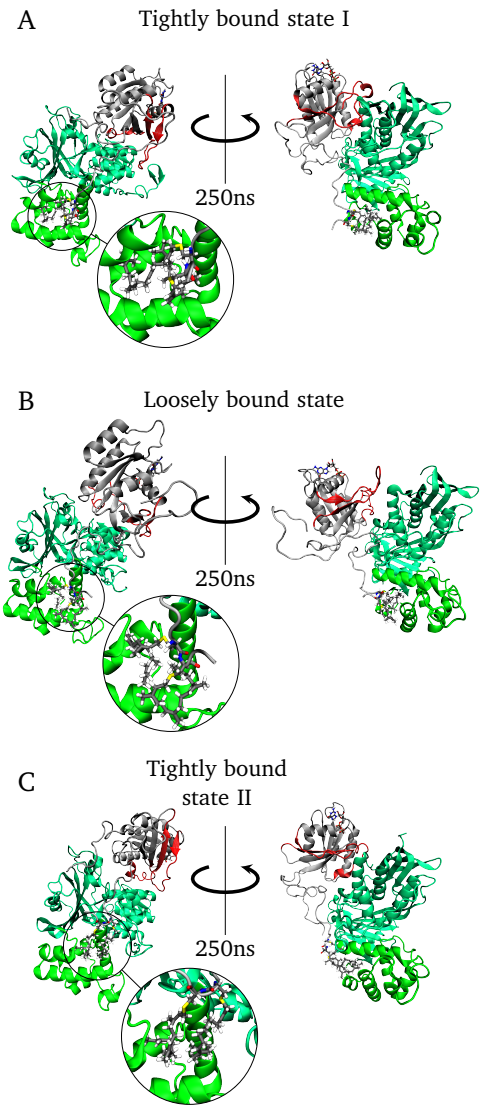


Figure 6.11: Final conformations of cytoplasmic Rab5:GDI complex after 250 ns of MD simulation displayed in front view and rotated by 180°. Conformations were derived from three independent simulations, A) $_{cyt}$Run1, B) $_{cyt}$Run2, and C) $_{cyt}$Run3.

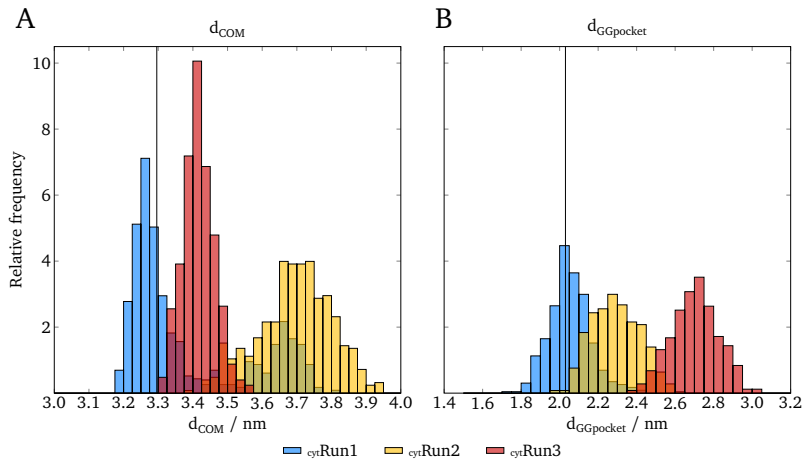


Figure 6.12: A) The intermolecular distance d_{COM} and B) the distance $d_{GGpocket}$ between the GG-Cys Cα atoms and GDI Met^{132} characterized the Rab5:GDI binding modes.[279] The vertical lines correspond to the Ypt1:yGDI experimental value.[282]

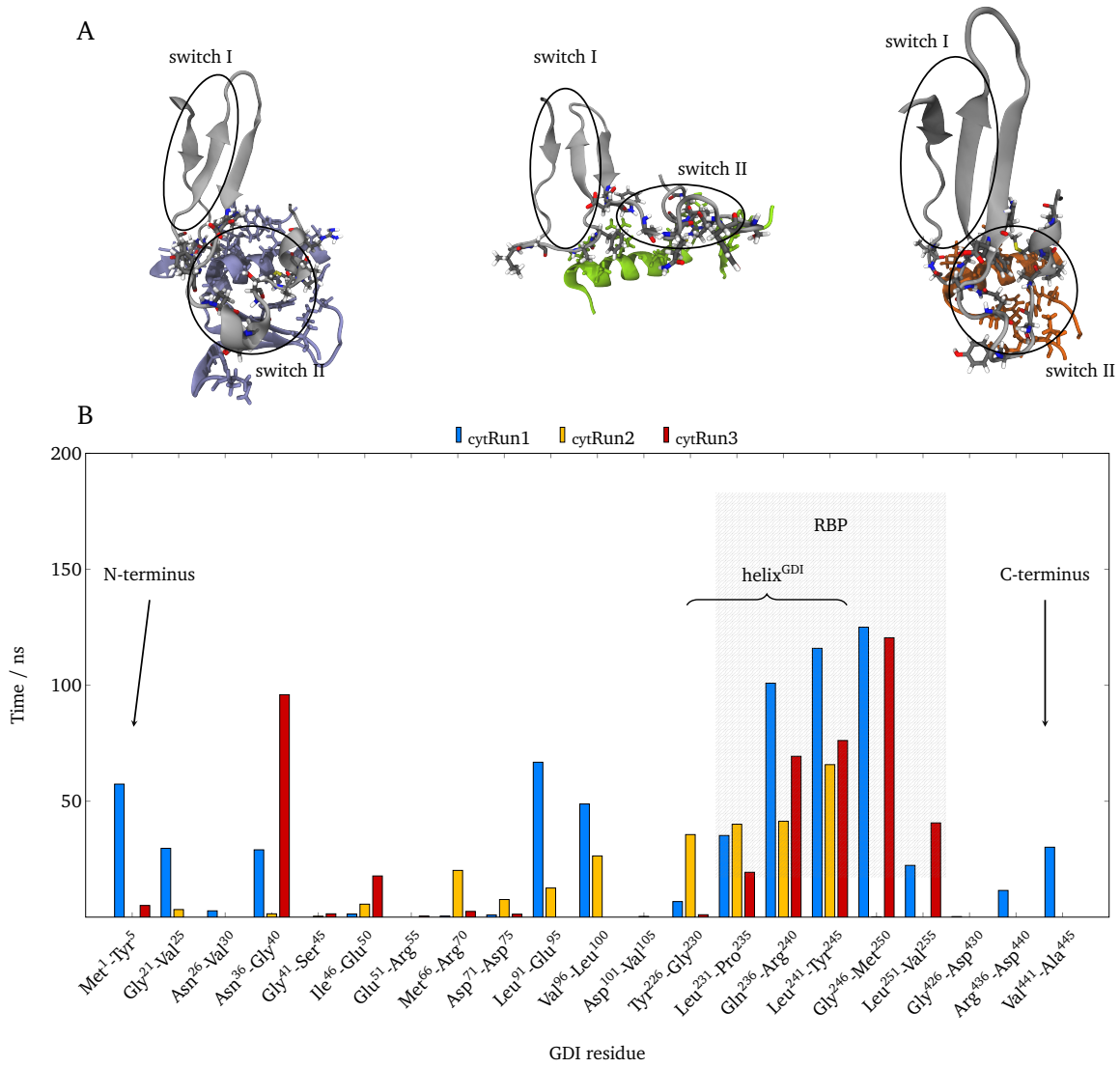


Figure 6.13: Interactions between the GDI RBP and the Rab5 binding epitope in the cytoplasmic complex. A) Representation of selected amino acid residues of the Rab5 switch regions interacting with GDI in the different binding modes. B) Average residence time of GDI RBP residues within 0.3 nm around the Rab5 binding epitope during 200 ns of MD simulation.[279]

The GG chain binding pocket was formed by the GDI domain II helices H1, H2 and H3 which included several hydrophobic leucine residues, e.g. Leu^{127}, Leu^{131}, Leu^{144}, Leu^{189}, and Leu^{216}, together with Phe^{140} and Val^{147}. Meeting the specification of the template Ypt1:yGDI complex, the initial Rab5:GDI model exhibited one GG chain inside the narrow prenyl binding pocket while the other chain was located nearby at the binding pocket surface. During the MD simulation, two binding pocket states 'open' vs. 'close' were observed for the different cytoplasmic complexes and could be characterized by the angle θ between helices H1 and H2 as well as by the pocket volume calculated by POVME (Figure 6.14).[286] The binding pocket 'open' state was characterized by a rearrangement of H1 to open the cavity and an associated angle θ>40°. In this case, the GG chains were deeply inserted (d_{GGpocket}≈2 nm) which is in excellent agreement with the results of the experimental Ypt1:yGDI complex (d_{GGpocket}=2.03 nm). The 'open' state was mainly observed for complex structures from $_{\text{cyt}}$Run1 with an average pocket volume of 692.1 Å which agreed well with a value of 656.6 Å in yGDI. In contrast, the 'close' binding pocket state was predominately found for the $_{\text{cyt}}$Run3 complex characterized by an almost complete displacement of the GG chains from the hydrophobic cavity with large distances d_{GGpocket}, small angles θ and a pocket volume of 169.8 Å. This is in accordance

with bGDI from PDB entry 1LV0 for which a pocket volume of 108.6 Å was calculated and which, thus, exhibited a 'close' binding pocket. Interestingly, despite its less close interactions between the Rab5 G domain and the GDI RBP, the Rab5:GDI complex from $_{cyt}$Run2 represented a 'semi-open' binding pocket state. This was marked by GG chains which were farther from the inner pocket core, combined with intermediate distances $d_{GGpocket}$ and angles θ. The corresponding pocket volume was 396.4 Å. In the present study, GG chains were counted to be inside the hGDI binding pocket for Met^{132} to head carbon atom distances <1.5 nm. This indicated that during most of the time the GDI prenyl binding pocket was either occupied by both GG chains or both GG chains were displaced from the hydrophobic cavity and associated with neighboring residues at the GDI surface (Figure 6.14C). The orientations of the GG chains in the hGDI prenyl binding pocket were similar to the positions of the GG chains in the doubly prenylated yeast Ypt1:yGDI complex.[282] In yeast, both GG moieties were rather extended and accommodated within the hydrophobic cavity. One chain was deeply inserted into the core of the binding pocket whereas the second chain was forming a lid on top of the first moiety aligned between helices H1 and H2.

The volume of the hGDI prenyl binding pocket was compared to other prenyl-binding proteins in the author's research article.[279] RabGDI selectively binds prenylated Rab proteins, however, to understand its unique function it may be intriguing to compare its structure to different prenyl binding proteins. RhoGDIs control the cellular localization as well as the activation state of Rho GTPases and represent the Rho-associated counterpart of RabGDI.[289] In complex with a Rho protein, RhoGDI shields the GTPase isoprenoid moiety from water exposure. In contrast to RabGDI, the prenyl-binding domain of RhoGDI adopts an immunoglobulin-like fold, which consists of antiparallel β-strands forming two β-sheets. Calculation of the volumes of the prenyl-binding pockets of mammalian RhoGDI gave values approximately half of the pocket volume in RabGDI.[279] Since Rho GTPases are modified by either one farnesyl or one GG chain, this pocket volume is sufficient to include one hydrophobic isoprenoid group. Another group of proteins known to bind GG moieties are the geranylgeranyl

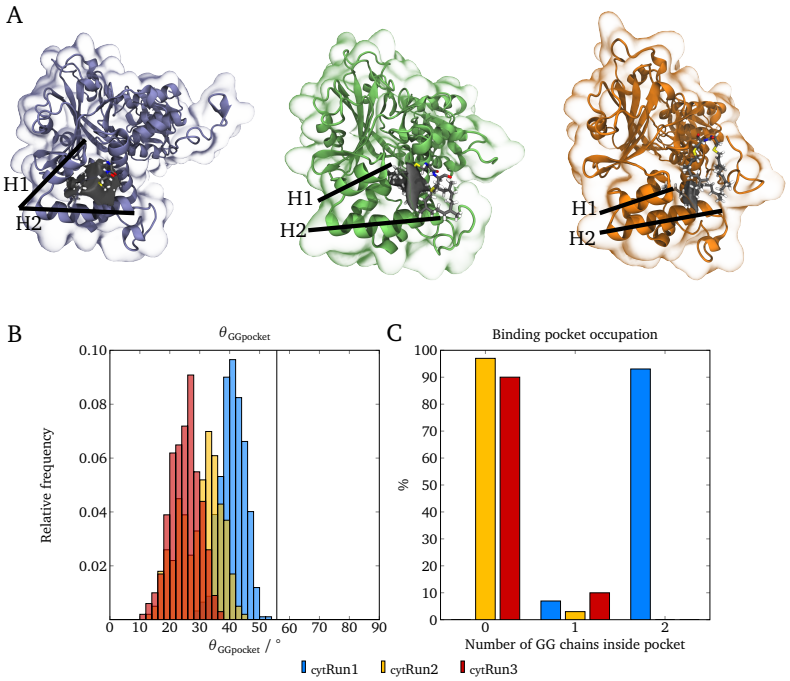


Figure 6.14: A) Representation of the prenyl binding pocket of cytoplasmic Rab5:GDI complex for $_{cyt}$Run1 (left), $_{cyt}$Run2 (middle), and $_{cyt}$Run3 (right). Volume of the binding pocket that persisted in 75 % of the simulation time is colored in gray. B) Distribution of the GG-binding pocket angle θ spanned between helices H1 and H2; the experimental value from yGDI[282] is marked by a vertical line. C) Number of GG chains that were defined as 'inside' the pocket, i.e. for Met^{132} to head GG carbon atom distances <1.5 nm.[279]

transferases (GGTases). They are responsible for the post-translational modification of proteins by transferring GG chains in a consecutive one-chain-transfer reaction.[290] GGTase type-I and RabGGTase are distinguished, the first transferring GG to a cysteine residue in a CAAX motif (C represents cysteine, A an aliphatic amino acid, and X a terminal amino acid), whereas RabGGTases exclusively transfer GG to a double-cysteine motif in Rab proteins presented by Rab escort protein (REP). However, both types of GGTases are known to bind only one GG group at once in an α-α barrel lined with hydrophobic residues.[243] Pocket volumes calculated for crystal structures of GGTase type-I and RabGGTase were smaller compared to the pocket volumes in RhoGDI and RabGDI[279] which may be due to the co-crystallization with GG and a short peptide not able to form interactions with the binding pocket.[291] After the first prenylation reaction the peptide may leave the GG binding pocket to give way for the subsequent second prenylation. Consequently, the pocket volume is sufficiently large to accommodate one GG moiety but no associated peptide like in RhoGDI or RabGDI.

REP and RabGDI are structurally related and share multiple conserved regions.[243] Consistent with GDI, REP comprises two domains with the highest structural conservation found in the RBP, the C-terminal binding region of domain I and the helices H1 and H2 of domain II. The REP RBP was shown to be invariant independent of binding to prenylated or unprenylated Rab7[292] which illustrates the fact that REP is obliged to bind unprenylated Rab proteins to enable the prenylation by RabGGTases. However, crystal structures of Rab7:REP-1 and Ypt1:yGDI complexes revealed hardly any differences regarding hydrophobicity and thus could not explain the preference of GDI to bind prenylated Rab proteins. In contrast to REP, GDI is unable to bind RabGGTases. The REP:RabGGTase interaction site includes the helices H1 and H2 in domain II, especially REP residue Phe^{279} which is absent in GDI and which induces a significant flexibility decrease in specific RabGGTase regions upon complex formation. Furthermore, residue Val^{287} within helix H2 is conserved in the REP family but substituted by a bulky phenylalanine residue in GDI, resulting in steric clashes along the complex interaction site.[243]

The formation of contacts between the long flexible Rab5 HVR and the GDI CCR in cytoplasm were observed during all MD simulations (Figure 6.15). Only part of the GDI CCR residues which were described to interact with the GTPase in the yeast homologous Ypt1:yGDI complex[287,293] were found to strongly affect binding in human Rab5:GDI. These residues corresponded to residues Leu^{92}, Thr^{97} in domain I, and Tyr^{219} in domain II of hGDI. In $_{cyt}$Run1 the HVR residues Pro^{204}-Thr^{208} preferentially interacted with GDI CCR residues Val^{88}, Thr^{97}, Phe^{102} of domain I and Tyr^{219}, Leu^{225} of domain II. In $_{cyt}$Run2 another patch of HVR residues, namely residues Arg^{197}-Asp^{200}, formed contacts with CCR residues Leu^{92}, Thr^{97}, Arg^{98} (domain I) and Arg^{218} (domain II) as well as with the GDI C-terminus. In $_{cyt}$Run3, a third HVR patch, namely residues Arg^{209}-Gln^{211}, was involved in most of the intermolecular interactions to GDI CCR residues Lys^{103}, Val^{104} (domain I) and Arg^{218}, Tyr^{219} (domain II) as well as to the extreme GDI C-terminus. Additionally, several GDI residues between Ile^{116} and Phe^{150}, that is beyond the CCR, formed contacts with Rab5 HVR residues, especially in the tightly bound state derived from $_{cyt}$Run1 (Figure 6.15D).

These results show that distinct hydrophobic and polar amino acid patches at the GDI CCR preferentially interacted with a number of Rab5 HVR residues. The exact counterparts within the Rab5 HVR strongly depended on the position and orientation of the HVR with respect to the GDI surface. The Rab5:GDI complexes exhibiting a 'close' or 'semi-open' GG chain binding pocket interacted with the GDI CCR via residue patches at the C-terminal or N-terminal part of the HVR, respectively. In case of the 'close' state this type of binding caused a strongly water-exposed arrangement of the HVR (see SASA values in the appendix, Figure B28) and thus, a weak interaction between GDI CCR and Rab5 HVR. On the other hand, in case of the 'semi-open' state, binding of the N-terminal part of the HVR decreased the motional freedom of the HVR, resulting in a reorientation of the G domain away from the GDI RBP and consequently, impairing of the protein-protein interactions which led to the loosely bound state. In contrast, the 'open' prenyl binding pocket complex formed contacts to the CCR via a residue patch located at the central region of the HVR. This represented an optimal balance between providing sufficient space to fully insert the GG chains within the prenyl binding pocket, forming

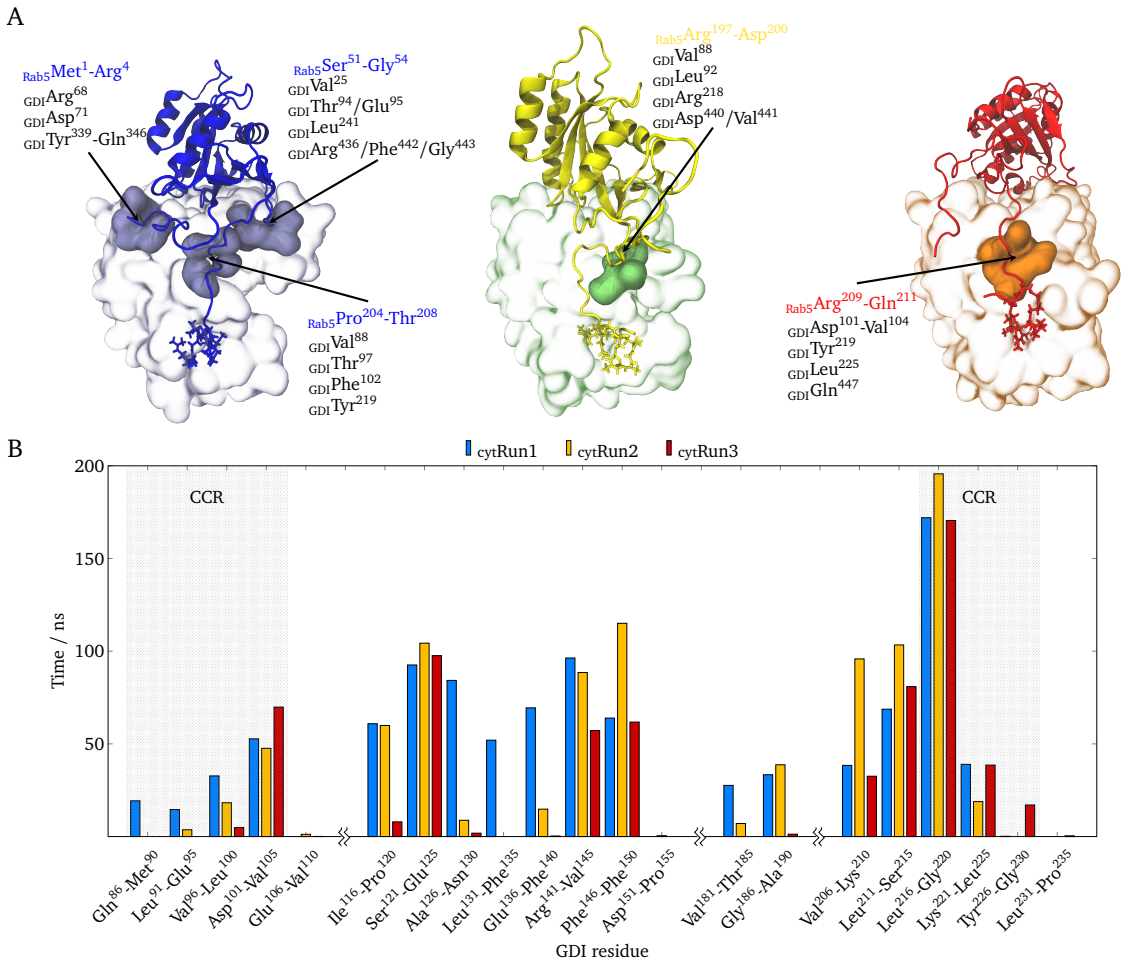


Figure 6.15: A) GDI accommodated terminal Rab5 residues in transient cavities. B) Average residence time of GDI CCR residues within 0.3 nm around the Rab5 HVR (res. 190-215) during 200 ns of MD simulation.[279]

numerous contacts between the HVR and GDI CCR and allowing the G domain to arrange in an ideal orientation relative to the GDI RBP suitable for strong intermolecular interactions. Thus, binding of the GDI CCR to a specific region of the Rab5 HVR seemed to be highly relevant in the cytoplasmic complex. However, in the membrane-associated complex, which was investigated in the following section, these interactions may be established based on the G domain orientation relative to GDI since prior to complex formation, the Rab5 HVR is coordinated by the bilayer-inserted GG chains.

In order to elucidate Rab5:GDI complex formation and the influence of the membrane on specific protein-protein interactions it was expedient to compare selected protein structural properties between membrane-bound GDI-free Rab5(GDP) (hereafter $_{\text{GDI-free}}$Rab5), cytoplasmic Rab5:GDI complex ($_{\text{cyt}}$Rab5:GDI), membrane-associated as well as membrane-bound Rab5:GDI ($_{\text{assoc}}$Rab5:GDI and $_{\text{insert}}$Rab5:GDI, respectively). The Rab5(GDP) and GDI RMSD and RMSF are provided in the appendix, Figure B29. The N- and C-terminal regions were most flexible in $_{\text{GDI-free}}$Rab5 but significantly stabilized by interactions with GDI ($_{\text{cyt}}$Rab5:GDI, $_{\text{assoc}}$Rab5:GDI) or with membrane lipids ($_{\text{insert}}$Rab5:GDI). The switch I region was most flexible in $_{\text{cyt}}$Rab5:GDI followed by $_{\text{assoc}}$Rab5:GDI. Thus, it may be suggested that membrane-anchoring had a greater stabilizing effect on this region than binding to GDI. This did not come as a surprise since the switch I region provided only a few residues for Rab5:GDI interaction. In contrast, the switch II region flexibility was highest in $_{\text{GDI-free}}$Rab5, indicating that tight interactions in the Rab5:GDI complexes involved major parts of the switch II, thus stabilizing them. The GDI RMSF was very similar for all complex simulations, exhibiting regions of higher flexibility between residues Gln^{57}-Ser^{65}, Asp^{151}-Thr^{165}, Glu^{358}-Lys^{364}, and C-terminal Gly^{426}-Gln^{447}. These

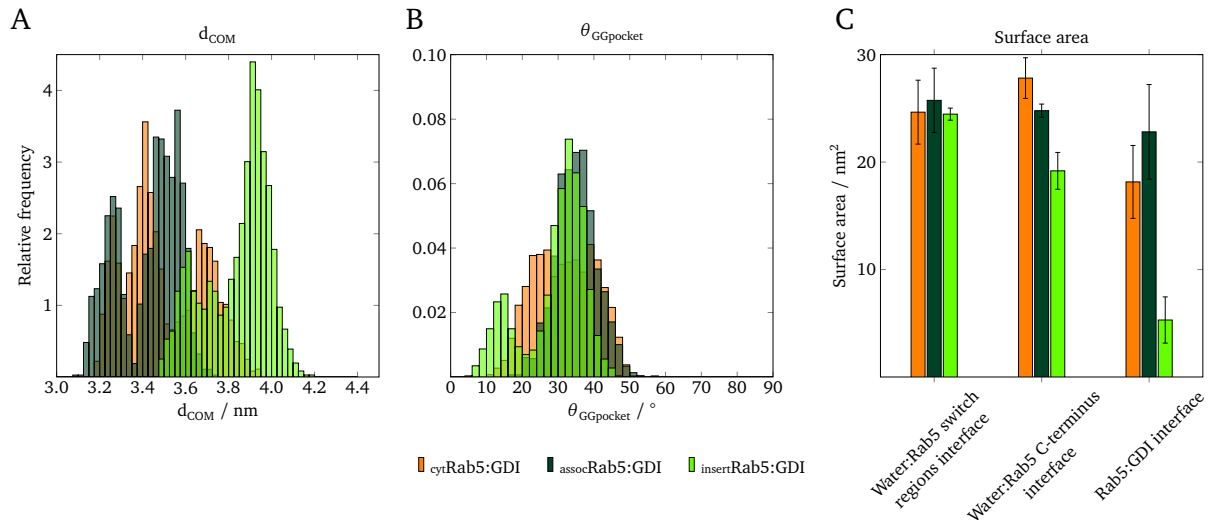


Figure 6.16: Distribution of A) intermolecular distances and B) the prenyl binding pocket angle θ in the Rab5:GDI complex in cytoplasm (orange), membrane-associated (dark green) and membrane-bound (light green). C) The SASA of the Rab5 switch regions and HVR as well as the total Rab5:GDI interaction interface were averaged over 3x 200 ns of MD simulation for each complex.

regions were hardly involved in binding to Rab5 or any bilayer lipids and rather solvent-exposed. Structural Rab5:GDI complex-related properties were compared between $_{cyt}$Rab5:GDI, $_{assoc}$Rab5:GDI and $_{insert}$Rab5:GDI and are displayed in Figure 6.16. The intermolecular distances were distributed in a bimodal manner, representing data from all three simulations, respectively. In general, d_{COM} was smallest in $_{assoc}$Rab5:GDI and largest in $_{insert}$Rab5:GDI. Fewer interactions between the GDI CCR and the Rab5 HVR in $_{insert}$Rab5:GDI rendered the RBP to Rab5 G domain contacts the predominant contribution to protein-protein interactions. Moreover, less tight interactions between the G domain and GDI were a manifestation of a slight re-orientation of GDI to enable the prenyl binding pocket to approach the bilayer surface and, subsequently, to include the GG chains.

The opening angle θ of helices H1 and H2 was similar in $_{cyt}$Rab5:GDI, $_{assoc}$Rab5:GDI and $_{insert}$Rab5:GDI. As seen above, only a small fraction of membrane-bound states was characterized by a small $\theta \approx 15°$ corresponding to $_{insert}$Run2. Solvent-exposure of the Rab5 switch regions was similar in all complexes, but greatest in $_{assoc}$Rab5:GDI with $_{cyt}$Rab5:GDI and $_{insert}$Rab5:GDI exhibiting slightly smaller values. The SASA of the Rab5 HVR was largest in $_{cyt}$Rab5:GDI followed by $_{assoc}$Rab5:GDI and $_{insert}$Rab5:GDI. Less HVR solvent contacts in the two last named systems were also due to HVR-lipid interactions. The greatest overall Rab5:GDI interaction interface was observed for $_{assoc}$Rab5:GDI, slightly reduced in $_{cyt}$Rab5:GDI and significantly smaller in $_{insert}$Rab5:GDI. As mentioned before, the reduction of the protein-protein interface in the membrane-bound complex was caused by the bilayer insertion of the GG chains which, thus, did not contribute to the overall Rab5:GDI interactions.

To identify the specific regions responsible for Rab5:GDI interactions, GDI residues that formed contacts with the Rab5 binding epitope (Figure 6.17A) and with the Rab5 HVR (Figure 6.17B) were analyzed for the different Rab5:GDI complexes. The first aspect that caught the eye was that residues involved in protein-protein contact formation were almost identical in $_{cyt}$Rab5:GDI and $_{assoc}$Rab5:GDI. In most cases, only the period of time for which those residues were found within 0.3 nm of the respective Rab5 region differed and was longer in case of $_{assoc}$Rab5:GDI. The second striking characteristic was that the number of interacting GDI residues in $_{insert}$Rab5:GDI was significantly smaller and interactions lasted for a shorter period of time. In contrast to the other systems, contacts with the Rab5 binding epitope were in the $_{insert}$Rab5:GDI complex primarily formed by Met^{66}-Asp^{75}, Val^{96}-Val^{105}, and Tyr^{226}-Arg^{240}. The focus of interacting residues in $_{cyt}$Rab5:GDI and $_{assoc}$Rab5:GDI was essentially between residues Leu^{231}-Val^{255} within the GDI RBP identified in the yeast Ypt1:yGDI

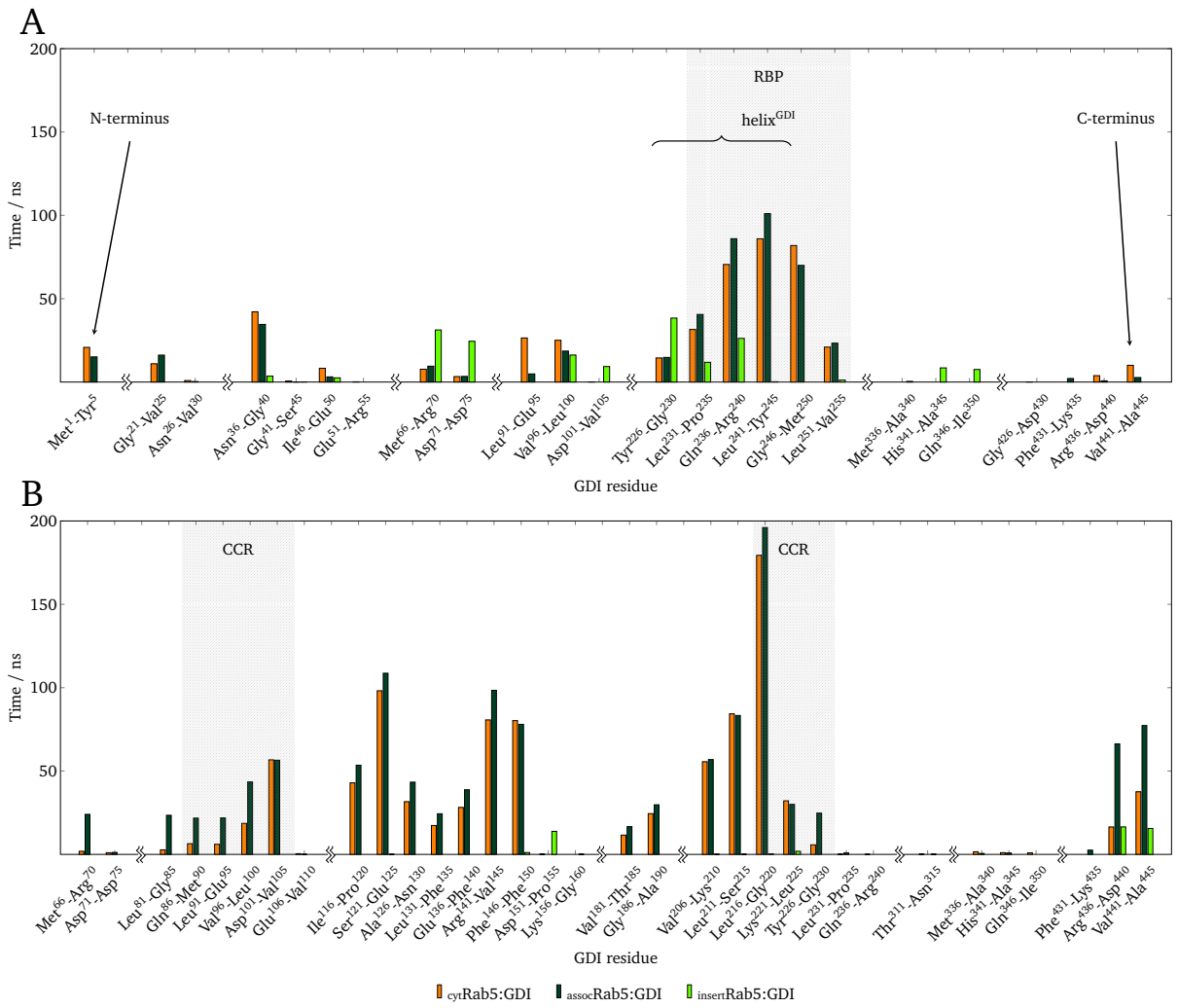


Figure 6.17: Average residence time of the A) GDI RBP residues found within a 0.3 nm cutoff of the Rab5 binding epitope and B) GDI CCR residues within a 0.3 nm cutoff around the Rab5 HVR during the last 200 ns of MD simulation. Data are averaged over three simulations for cytoplasmic (orange), membrane-associated (dark green) and membrane-bound (light green) Rab5:GDI, respectively.

complex.[283] Only very few GDI residues were found to interact with the Rab5 HVR in $_{insert}$Rab5:GDI, namely Asp^{151}-Pro^{155} and C-terminal Arg^{436}-Ala^{445}. The corresponding interactions in $_{cyt}$Rab5:GDI and $_{assoc}$Rab5:GDI were spread out across the GDI CCR (residues Gln^{86}-Val^{105}, Leu^{216}-Gly^{230}) as well as large patch of residues between Ile^{116}-Phe^{150} and at the GDI C-terminus. Especially in cytoplasm, Rab5:GDI interactions were found to be very tight forming high-affinity complexes. This raised the question of how complex dissociation and delivery to the corresponding target membrane may be realized in cells. GDI displacement factors (GDFs) have been identified as key elements responsible for the regulation of Rab protein dissociation from GDI.[278] Mammalian integral Ypt-interacting protein 3 (Yip3) was identified to be a GDF which selectively segregates endosomal Rab proteins like Rab5, Rab7 or Rab9 from GDI in a catalytic manner and conveys them to the respective membrane.[294] However, to date the structural basis for GDF interaction with Rab:GDI complexes remains obscure. Two further mechanisms are discussed that are able to influence the equilibrium between free Rab protein and the Rab:GDI complex by reducing the affinity of Rab for GDI, namely the activation of Rab proteins by GEFs[295] and Rab post-translational modifications.[296] Both methods are unable to actively displace Rab from GDI, but depend on the spontaneous intrinsic Rab:GDI complex dissociation and serve the recruitment of Rab from the complex to membranes. As seen above, compared to the cytoplasmic Rab5:GDI complex a significantly lower number of tight interactions was formed in the membrane-bound complex. To obtain a qualitative estimation of interaction strengths, the Gibbs free

energies of membrane-bound Rab5:GDI complex segregation and GG chain bilayer-to-water transfer were determined and related to each other in the following section.

6.2.6 Free energy profiles and the Rab5 membrane extraction process

To gain insight into the strengths of the protein-protein interactions in the membrane-bound Rab5:GDI complex SMD simulations were performed to determine the potential of mean force. Thus, a force was applied to the GDI residues from the final conformation of $_{\text{insert}}$Run1 to pull them in +z direction which resulted in a separation of GDI from membrane-bound Rab5 (Figure 6.18). During the separation process the Gibbs free energy increased until no more protein-protein interactions occurred and GDI was completely dissociated from Rab5. This was the case for GDI-membrane distances >8 nm yielding a Gibbs free energy difference ΔG of approximately 78 kJ mol^{-1}.

Starting from the same initial conformation but applying the +z directed force to Rab5 residues 5-210 caused an extreme extension of the HVR that culminated in the extraction of the GG anchor from the six-component membrane (Figure 6.19). During the whole process interactions between Rab5 and GDI were relocated and distorted, however, the complex remained intact. The by SMD 'enforced' GG chain extraction process from the membrane could be subdivided into several subsequent events. It is speculated that the order of these steps was predefined by their energetic costs, i.e. the energy increase associated with them. 1) The orientation of the Rab5 G domain changed from the tilted, virtually membrane parallel position toward an orientation almost perpendicular to the bilayer surface. 2) The previously coiled HVR was extended to its greatest expansion. Especially when approaching the fully extended state, this was accompanied by a severe Gibbs free energy increase of approximately 280 kJ mol^{-1} between z distances 6.5 nm-11.0 nm. 3) In a last step, the GG chains were removed from the bilayer hydrophobic core region and water exposed.

The entire process, that is starting from the preferential tilted Rab5 orientation at the membrane to the fully solvent-exposed state with an extended HVR, exposed GG chains and distorted Rab5:GDI complex interactions, was associated with a Gibbs free energy increase of 535 kJ mol^{-1}. As discussed in chapter 4, section 4.4.4, the method used here was not appropriate to determine accurate absolute Gibbs free energies. However, it allowed an energetic classification of the involved steps, i.e. orientation changes, HVR extension and GG chain membrane extraction. Accordingly, the disassembly of the Rab5:GDI complex was energetically preferred over the GG chain extraction. It may be specu-

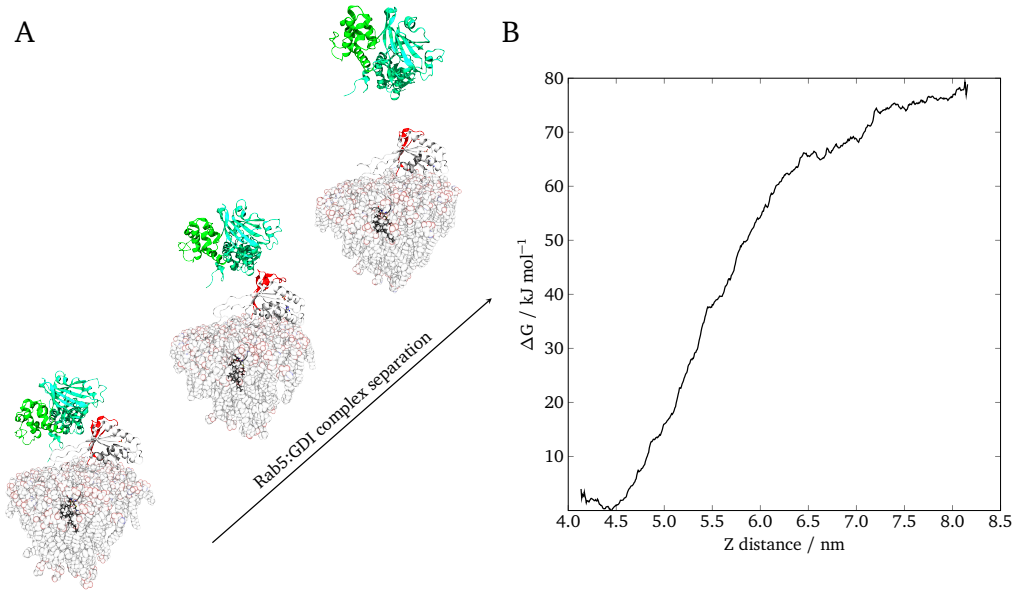


Figure 6.18: A) Separation of GDI from membrane-bound Rab5 and B) associated free energy profile.

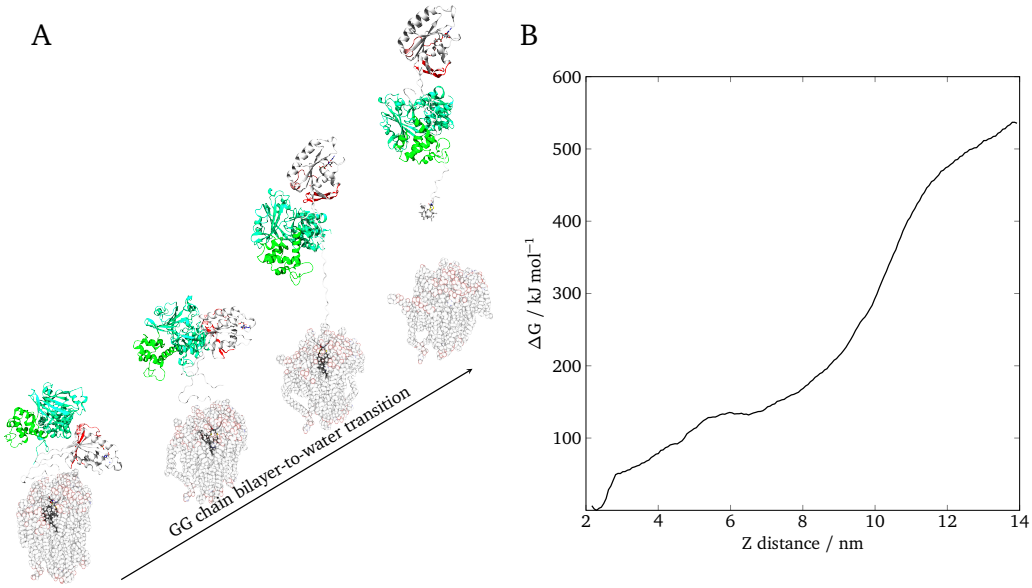


Figure 6.19: A) Membrane extraction of the GG anchor by applying a force to the Rab5 residues 5-210 occurred without Rab5:GDI complex disassembly. B) Bilayer-to-water free energy of the GG chains involved an extreme extension of the HVR.

lated that this was the case since the interactions between the GDI RBP and the Rab5 binding epitope in insertRun1 were not as tight as in insertRun3 or in the cytoplasmic complex. The Gibbs free energy difference of Rab5:GDI complex segregation of 78 kJ mol^{-1} was by far smaller than the bilayer-to-water transition free energy determined for the HVR release from the six-component membrane which yielded 124 kJ mol^{-1} (see section 4.4.4).

The final membrane-extracted Rab5:GDI complex structure was subjected to further 250 ns of MD simulation in cytoplasm to probe whether the structural changes, e.g. dislocation of the protein-protein interface and extension of the HVR, were reversible (Figure 6.20). Indeed, the HVR adopted a rather coiled structure to enable the GG chains to adsorb to the hydrophobic exterior of the GDI

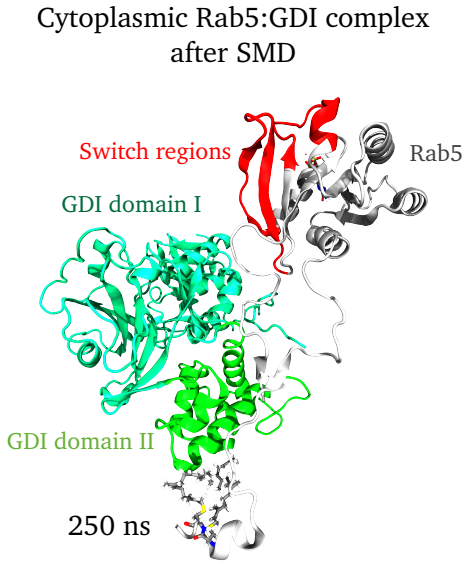


Figure 6.20: Final conformation of cytoplasmic Rab5:GDI complex after 250 ns MD simulation which was previously extracted from the membrane by SMD. The Rab5 GG chains spontaneously associated with hydrophobic residues flanking the GDI prenyl binding pocket.

prenyl binding pocket. The binding pocket itself remained in the 'close' state. However, the distortion of the protein-protein interactions between the GDI RBP and the Rab5 binding epitope were found to be irreversible. The association of the GG chains with the hydrophobic cavity occurred spontaneously, but at a position not suitable for deep insertion into the binding pocket. Performing further independent MD simulations may provide a better understanding of how the GG chains insert into the cavity and which interactions cause structural rearrangements to open the binding pocket. However, it should be kept in mind that the SMD-enforced GG chain extraction did not represent the natural mechanism in cells and that the artificial structures derived in this process may be unsuitable for the prediction of Rab5:GDI interactions.

The high energy barrier associated with the prenyl chain bilayer-to-water transfer arose the question how the GDI-mediated Rab5 membrane extraction process takes place in detail and how the extremely unfavorable GG chain retrieval from the bilayer is realized within the cell. The extraction procedure started with membrane-bound Rab5(GDP) in its tilted orientation (Figure 6.21). GDI initially recognized the small GTPase via long-range electrostatic interactions between the RBP and the Rab5 switch regions. These contacts were sufficient to stably associate GDI to Rab5. Subsequently, a slight GDI orientation change brought the GDI domain II encompassing the prenyl binding pocket closer to the bilayer surface and enabled the formation of short-range interactions between the GDI CCR and the membrane-coordinated HVR. This re-orientation was also associated with a commencing opening of the prenyl binding pocket to prepare the extraction and uptake of the hydrophobic GG chains. Simultaneously, the HVR membrane interactions were reduced. Binding of the CCR to the Rab5 HVR was shown to be guided by hydrophobic residues but was rather unspecific regarding the exact residues. Different Rab proteins vary significantly regarding their HVR sequences. Thus, the non-specific CCR-to-HVR binding may explain the fact that the membrane extraction process of a variety of Rab proteins is mediated by only two human RabGDI isoforms.[279] The present study showed that rather the length of the HVR, or more exactly the distance between the HVR residues bound by the CCR and the G domain, determined the orientation of the GTPase G domain and thus the strengths of interactions between the GDI RBP and the switch regions. The next step in the membrane extraction process, namely the incorporation of the GG chains in the GDI prenyl binding pocket could not be observed in MD simulations. It is supposed that hydrophobic residues flanking the binding pocket were located in very close proximity to the bilayer and enabled the transfer of the GG chains to the hydrophobic cavity surface. This step is speculated to be driven by the energetic advantage of shielding the GDI hydrophobic cavity from the solvent by adsorbing the likewise

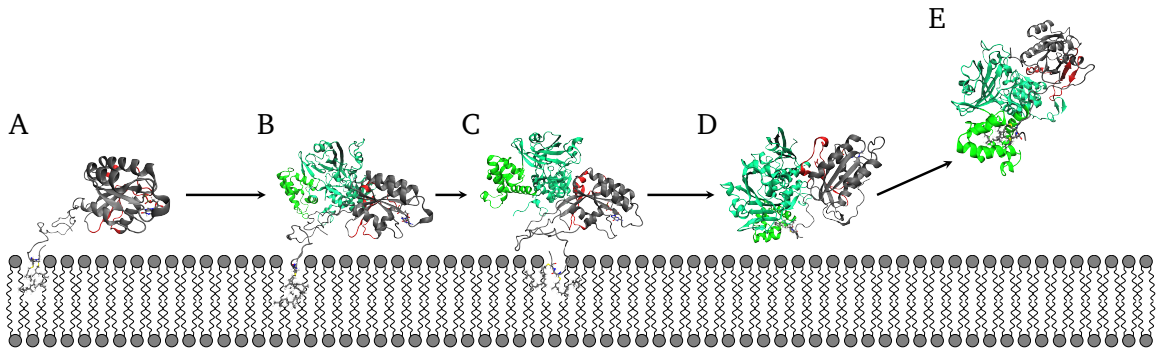


Figure 6.21: The Rab5:GDI extraction process from the EE membrane proceeded via sequential recognition and stabilization steps. A) GDI-free Rab5(GDP) tilted to the bilayer surface with partially solvent screened switch regions. B) First Rab5:GDI recognition occurred via the GDI RBP and the Rab5 binding epitope, followed by C) a slight GDI orientation change, which allowed formation of GDI CCR - Rab5 HVR interactions and a close approaching of the GDI prenyl binding pocket toward the membrane surface. D) Accommodation of the Rab5 GG chains in the binding pocket induced detachment of the Rab5:GDI complex from the membrane and E) formation of a stable soluble complex.[279]

hydrophobic GG anchor. Thus, the extremely energetically unfavorable solvent exposures of the GG chains and the hydrophobic pocket residues were bypassed. The large positive Gibbs free energy differences calculated for the SMD 'enforced' GG chain extraction would not occur in the GDI-mediated process since the direct GG chain transfer from the bilayer hydrophobic core to the prenyl binding pocket circumvented a fully solvent-exposed GG chain state. With inserted GG chains the Rab5:GDI complex dissociated from the membrane surface, probably due to free diffusion. The cytoplasmic complex was characterized by tight interactions between 1) the GDI RBP and the Rab5 binding epitope, 2) the GDI CCR and the HVR, and 3) the GDI prenyl binding pocket and deeply inserted GG chains.

6.3 Summary

In this chapter protein-protein interactions of human Rab5(GDP):RabGDI complex in cytoplasm as well as associated with and bound to a complex negatively charged model membrane were investigated. The focus was on addressing the following issues:

- Identification of human Rab5:GDI interaction sites in cytoplasm and in the membrane-bound complex
- Determination of the Gibbs free energy profiles of Rab5:GDI complex binding and Rab5 membrane extraction
- Proposal of a GDI-mediated Rab5 extraction mechanism from the EE membrane

Similar to the yeast complex, three main sites of protein-protein interactions were identified in the cytoplasmic human Rab5:GDI complex. These included contact formation between 1) the RBP comprising residues of GDI domain I and the Rab5 binding epitope encompassing the switch regions, 2) the GDI CCR and the Rab5 HVR, and 3) the GDI prenyl binding pocket formed by helices H1, H2, and H3 and the Rab5 GG chains. Based on the interactions between the RBP and the Rab5 switch regions two distinct binding modes have been identified. The tightly bound state was characterized by a great number of protein-protein contacts and a small Rab5:GDI intermolecular distance. The loosely bound state exhibited fewer RBP to Rab5 binding epitope contacts and a larger overall protein-protein distance. The orientation of the G domain and, thus, the strength of interactions was found to depend on the CCR to HVR binding mode. Hydrophobic interactions between the CCR and different HVR residues were rather un-specific regarding the residue position along the HVR sequence. Shorter distances between the CCR-interacting HVR residues and the Rab5 G domain led to a re-orientation of the G domain away from the GDI RBP and less tight binding. Interaction with residues at the extreme HVR C-terminus caused a drag of the GG chains out of the prenyl binding pocket. An optimal G domain orientation with regard to GDI binding properties as well as translational and rotational degrees of freedom was obtained when the GDI CCR interacted with the Rab5 HVR residues Pro^{204}-Thr^{208}.

Depending on the position of the GG chains being either inserted within or outside of the GDI hydrophobic cavity, the binding pocket adopted an 'open' or a 'close' state, respectively. Basic residues formed and surrounded the GDI prenyl binding pocket which had two consequences for the Rab5:GDI complex:

First, cytoplasmic GDI approached the membrane and established electrostatic interactions via amino acids residues in proximity to the GG chain binding pocket which kept the binding pocket stably in close contact with the bilayer. Initial contacts between GDI and Rab5 were previously formed via the GDI RBP and the Rab5 switch regions. A slight GDI re-orientation shifted the prenyl binding pocket closer to the membrane surface and allowed the formation of further interactions between the GDI CCR and the Rab5 HVR.

Second, basic residues around the GDI hydrophobic cavity allowed the transfer of the GG chains

from the membrane interior to the hydrophobic GDI domain II surface and subsequently to the inner prenyl binding pocket.

The energy barrier associated with the GDI-mediated membrane extraction and transfer of the GG chains to the prenyl binding pocket was supposed to be significantly smaller than the GG chain bilayer-to-water transition determined from SMD simulations. A complete solvent exposure of the prenyl chains is not expected to take place in cells. Thus, the SMD-enforced extraction procedure revealed artificial steps like a full extension of the Rab5 HVR and allowed an estimation of the corresponding energy contributions. Furthermore, SMD led to a segregation of the membrane-bound Rab5:GDI complex only with substantial energetic costs which indicated that an intact RBP to Rab5 switch region interaction site was sufficient to stably associate GDI to Rab5 under native conditions.

7 Summary and conclusions

This thesis intends to contribute to the understanding of the membrane-associated behavior of peripheral small GTPases taking Rab5 as an example. Small GTPases belong to the membrane-located G proteins which contain a phosphate-binding loop and serve as molecular switches shuttling between the inactive GDP-bound and active GTP-bound states with distinct nucleotide-dependent conformations. Members of the Ras superfamily of small GTPases regulate various cellular functions ranging from vesicular and non-vesicular transport, cell shape and motility control to cell signaling. The catalytically active G domain is conserved between different small GTPases and connected to the membrane surface via a protein-specific, long and flexible C-terminal domain, termed the HVR. Two GG chains are covalently attached to the HVR extreme C-terminus and serve membrane anchoring of the Rab5 GTPase.

Rab5 is the key regulator of early endosomal trafficking and thus, directly and indirectly via binding partners, involved in a tremendous spectrum of cellular processes and signaling cascades. The great flexibility of the HVR has two important implications which are addressed in the present thesis:

1. To date, it was not possible to determine a crystal structure of full-length Rab5 by experimental techniques. Thus, membrane-protein and protein-protein interactions involving the HVR were limited to studies of endocytosis using cell assays and tissue and did not provide atomic resolution.
2. The orientation and position of the Rab5 G domain at the membrane are variable and able to adapt in response to specific signals, e.g. approaching the bilayer or certain binding proteins.

Regarding Rab GTPases, these challenges are hardly addressed in literature. However, the importance of the closely related Ras proteins for cell signaling and their involvement in different types of human cancer, has motivated research and led to some insight in the last years.[233–235] In order to probe whether the Ras-specific behavior may also be valid for Rab GTPases, it was decided to construct a full-length model of Rab5 by comparative modeling and subject it to full-atomistic MD simulations. The protein active and inactive states were considered in order to reveal nucleotide-dependent conformations, which may represent starting points for activation state-specific therapies in treating Rab5-related dysfunctions. Furthermore, since Rab5 is targeted to the early endosome membrane with a defined mole fraction of certain lipids, the influence of the membrane composition was determined using MD. In addition to protein-membrane interactions, which are expected to significantly influence Rab5 orientation and dynamics, protein-protein interactions represent potent modifiers of protein dynamics and function. A multitude of experimental studies exists, that thoroughly investigate Rab5(GTP) G domain interactions with regulatory GAPs[297] or downstream effector proteins like Rabaptin5[73] or EEA1.[298] Since binding partner recognition occurs via residues of the switch regions located within the G domain, contributions of the HVR are usually ignored in those studies. The full-length Rab5 model derived in the present thesis was used to probe and elucidate homotypic Rab5 dimerization. Membrane-associated dimerization is described for related Ras proteins as a prerequisite for effector protein activation which also act as dimers. Chemical cross-linking assays provided evidence that Rab5 dimers indeed exist, however, the structural basis of the dimerization as well as its physiological role in the cell remain elusive to date.[256]

The Rab5 nucleotide-dependent activation cycle is coupled to a spatial shuttling between the early endosome membrane and the cytoplasm. While Rab5(GTP) is exclusively membrane-associated, Rab5(GDP) is extracted from the bilayer by GDI, kept soluble in cytoplasm, and recycled back to donor membranes, where it passes a new catalytic cycle. The intermolecular interactions underlying

Rab5:GDI complex formation have been investigated in the analogous yeast Ypt1:GDI complex.[282] However, that crystal structure is missing the long GTPase HVR, hence, full-length human Rab5 complexed to GDI was comprehensively characterized in the present study using MD.

Major findings and results of the present thesis are summarized in the following section.

The Rab5 C-terminal HVR$^{206-215}$

Full-atomistic and coarse-grained topologies for the Rab5 GG anchor were adapted from existing force fields and verified by comparison to quantum mechanically optimized structures. The truncated Rab5 HVR$^{206-215}$ including two GG chains was investigated in membrane models of different compositions in order to compare its dynamics to findings from studies of short Ras peptides described in literature.[169–171] It was possible to characterize the influence of the HVR$^{206-215}$ on membrane structure and dynamics. The overall effect on the membrane structure was only minor. Lipid order and the related membrane thickness were slightly reduced within a radius of 0.5 nm around the GG chains. To a greater extent, the GG anchor adapted to the membrane environment by adjusting the insertion depth and chain order to avoid a spatial mismatch between the hydrophobic membrane interior and the GG chains. The Rab5 HVR$^{206-215}$adopted a very flexible, intrinsically disordered structure without forming any secondary structure elements. HVR dynamics were found to be membrane composition-dependent. Persistent electrostatic interactions between the HVR basic Arg^{209} and the negatively-charged signaling lipid PI(3)P in the model bilayer mimicking the early endosome membrane were observed. These interactions increased the HVR-membrane free binding energy compared to the uncharged membrane models, prevented a deeper intrusion of the HVR side chains into the lipid head group region and thus, enabled a greater flexibility of the terminal HVR residues in the early endosome membrane. The interaction between the HVR and PI(3)P was confirmed in long-term coarse-grained MD simulations, which indicated an accumulation of PI(3)P in close proximity to the Rab5 peptide. The observed protein-lipid interactions may lay the foundation for the formation of Rab5 and PI(3)P-enriched microdomains, which serve as spatially segregated platforms controlling distinct signaling processes in the cell.[156]

Full-length monomeric membrane-bound Rab5

Full-length models of Rab5(GDP) and Rab5(GTP) were generated by a comparative modeling approach based on crystal structures of the catalytic G domain deposited in the protein data bank.[72,73] Depending on the membrane composition, two distinct membrane-associated protein orientations were detected in case of double-GG Rab5. Starting from an initial perpendicular orientation, the G domain adapted a tilted conformation close to the bilayer surface during the MD simulations. This tilting was most significant in the charged bilayer model mimicking the early endosome membrane. The tilting in the charged membrane was accompanied by a nucleotide state-dependent G domain orientation with respect to the bilayer surface. The GDP-bound state was characterized by a 'syn' conformation of the switch regions, which are essential for binding partner recognition. In the 'syn' conformation, the switch regions were on the proximal protein side, partly buried between G domain and bilayer and hardly effector accessible. In contrast, the Rab5(GTP) was found in the 'anti' conformation with the switch regions on the G domain side distal to the membrane surface and fully solvent-exposed.[8] The reason for the exclusively GDP-bound state related 'syn' orientation was suspected to be the more disordered structure of the switch regions compared to the GTP-bound state. In Rab5(GDP), the switch regions were less stabilized by interactions with the nucleotide or G domain interior, and more likely to be involved in protein-lipid contacts via long-range electrostatic interactions.

In contrast to the membrane composition- and nucleotide state-dependent multiple membrane-binding modes observed in double-GG Rab5, mono-GG Rab5 was shown to remain in the initial perpendicular conformation. This behavior was explained by a faster translational diffusion of the

single GG chain anchored Rab5 compared to the double-GG protein, which in turn impeded the formation of protein-lipid electrostatic interactions.

Besides the GG anchor, the HVR was found to establish protein-lipid contacts and may serve as an auxiliary binding site for effectors and regulatory proteins. Thus, the role of the HVR is not only to be a simple linker between membrane and catalytic G domain, but rather a highly flexible and adaptable interaction tool allowing the G domain a high degree of rotational and translation freedom as well as high amplitude motions.

Homotypic Rab5 dimerization

The initial Rab5 monomer relative orientations were chosen using the dimerization interfaces observed in GTP-bound K-Ras dimers as a model. Rab5 dimerization occurred predominately via residues of three predominant dimerization sites which comprised the switch I β-interface, the switch II α2-helix as well as the α3-helix. The helical interfaces were involved in Rab5 dimerization independent of the nucleotide state, but were more frequently found in the GTP-loaded states. In contrast, dimerization via the β-interface, which included residues of the β2'-strand that is solely found in Rab5(GDP), was limited to the GDP-bound dimers. Due to the rather disordered structure of the switch I in the GDP-bound states, a large surface area was buried upon dimerization via the β-interface. Consequently, determination of the binding energies including an interface size-dependent favorable surface tension energy contribution, indicated a more stable formation of Rab5(GDP) dimers compared to Rab5(GTP) dimers. The preference of a certain Rab5 nucleotide state for dimerization via the β- or helical interface has important implications for the availability of the switch regions for binding partner recognition and interaction. In Rab5(GDP) dimers, which preferentially interacted via the switch I β-interface, residues in the switch II region remained rather accessible. On the other hand, in Rab5(GTP) dimers, where dimerization occurred via the helical interfaces α2 and α3, the switch I region as well as residues of the invariant hydrophobic triad essential for effector recognition, were readily accessible. Furthermore, an asymmetric arrangement of the Rab5(GTP) dimers was observed, with one monomer being tilted toward the membrane surface while the second monomer was in a perpendicular orientation rather distant from the bilayer. This orientation improved the accessibility of the switch regions in the Rab5(GTP) monomer which was more distant from the membrane surface.

Rab5(GDP):GDI complex formation and Rab5 membrane extraction

Based on the interaction sites in the yeast Ypt1:GDI complex, a human Rab5(GDP):GDI complex was modeled and simulated in cytoplasm as well as associated to the charged early endosome membrane. The simulations were designed to allow a characterization of important conformations representing steps in the GDI-mediated Rab5 membrane extraction process. In a first step, free GDI was expected to approach membrane-bound Rab5(GDP) tilted toward the bilayer surface. At this stage, protein-protein contacts occurred predominately via the GDI RBP and the Rab5 switch regions.[279] SMD simulations revealed that interactions between the RBP and the switch regions were sufficient to stably associate both proteins with each other. Electrostatic interactions between the GDI domain II, which harbors the GG chain binding pocket, were shown to result in a slight re-orientation of GDI, thereby shifting the binding pocket closer to the membrane surface. Subsequently, additional interactions between the GDI CCR and the Rab5 HVR further stabilized the protein-protein complex. In a next step, which could not be captured by MD simulations, the Rab5 chains were transferred from the bilayer to the GDI prenyl binding pocket. This transfer was promoted by basic residues flanking the prenyl binding pocket. It is speculated that shielding the hydrophobic cavity from solvent by adsorption of the GG anchor provides the energy required for the GG chain membrane extraction step. Finally, the Rab5:GDI complex dissociated from the membrane and formed a tightly interacting cytoplasmic complex. Interactions occurred via three binding sites, namely, 1) the GDI RBP and the

Rab5 switch regions, predominately residues of the switch II, 2) the GDI CCR and the HVR, and 3) the GDI prenyl binding pocket and GG chains. A tightly bound and a loosely bound binding mode were identified in the cytoplasmic complex, which were determined by the number of RBP-switch contacts and the intermolecular distance between both protein centers. The orientation of the Rab5 G domain and thus, the binding mode was affected by the CCR to HVR interaction. The most stable Rab5:GDI complex with the highest overall number of interactions was achieved when Rab5 HVR residues 204-208 were bound by the CCR. In this case, the G domain was able to rotate in a way that simultaneously optimized RBP to switch and prenyl binding pocket to GG contacts. Binding the HVR at an extreme N- or C-terminal position, resulted in either a distortion of the RBP-switch interface or an extraction of the GG chains from the GDI prenyl binding pocket, respectively.

The MD simulations performed in this thesis shed light on Rab5 dynamics and interactions in the early endosome. They contribute to the profound understanding of membrane-associated Rab5 by taking both, active Rab5(GTP) and inactive Rab5(GDP) into account with explicitly considering the C-terminal HVR. The HVR serves as a flexible, multifunctional unit that connects the catalytic G domain to the bilayer, influences correct membrane targeting, contributes to interactions with lipids and binding partners, and enables the G domain to adopt a large variety of orientations. It is the prerequisite for the tilting of the monomeric Rab5 G domain toward the bilayer surface in the early endosome membrane. This orientation has important implications in effector signaling downstream of Rab5. Rab5(GTP) recruits effector proteins to the membrane via its switch regions, where they exert their biological function. One example is the PI(3)-kinase Vps34 which generates PI(3)P in the early endosome membrane. Furthermore, the effector EEA1 simultaneously binds PI(3)P and Rab5(GTP) via its C-terminal FYVE domain. The Rab5 G domain tilting allows to bring binding partners in very close proximity to the bilayer surface. In contrast to the largely accessible switch regions in GTP-bound Rab5, Rab5(GDP) adopted a conformation which is characterized by partly inaccessible switch regions. This behavior is mostly caused by the rather disordered structure of the Rab5(GDP) switch regions and represents a mechanism that may serve the activation state-dependent specificity in binding partner recognition. However, the partly buried switch conformation is challenging for Rab5(GDP) interacting proteins like GDI. Dimerization, which was experimentally shown for Ras proteins,[257] may represent a concept to prevent the tilting of the monomeric G domain in Rab5. Supporting this assumption, dimerization was found to be more stable in Rab5(GDP) compared to Rab5(GTP) in terms of buried interfacial area. Since dimerization in Rab5(GDP) preferentially occurred via the switch I, residues of the switch II remained largely accessible. The switch II was shown to be the first and one of the major interaction sites in the Rab5(GDP):GDI complex. Thus, approaching GDI may recognize and bind dimeric Rab5(GDP) via the RBP in order to form a stable complex and finally extract the GTPase from the membrane. Hence, the physiological role of Rab5(GDP) dimerization may be a membrane-associated inactive reservoir which is able to be recycled by GDI, possibly in a Rab5 concentration-dependent manner. The biological relevance of Rab5(GTP) dimers remains elusive. Similar to K-Ras in the mitogen-activated protein kinase pathway,[247] it may be an essential prerequisite for effector dimerization. However, an involvement of Rab5(GTP) dimers in downstream signaling within the early endosome has not been described in literature so far.

8 Outlook

Full-atomistic MD simulations represent a valuable tool for investigation of protein dynamics and interactions. However, they easily suffer from the 'sampling problem', which means that the sampled conformations are not sufficient to allow a proper statistical analysis or not all relevant conformations are even sampled. A prolongation of the simulation time is straightforward and improves sampling by the expense of computational costs. Unfortunately, it does not yield new conformations and insights in cases the system is 'caught' in a local energy minimum. It is more promising to run a large number of short replicate simulations starting from slightly different initial positions. This approach may be useful for the Rab5 dimers, since it may reveal whether the identified interaction sites are the most likely ones. A completely different relative monomer orientation was observed in a crystal structure of the dimeric K-Ras G domain.[268] Therein, both G domains were asymmetrically complexed with their HVR bases pointing in opposite directions. In theory, the long and flexible HVR would enable such an orientation. Whether this represents a realistic conformation in Rab5 dimers should be elucidated by further simulations.

Rab5 is the key regulator of vesicular transport and signaling in the early endosome. As such, it is involved in a large variety of protein-protein interactions apart from dimerization and complex formation with GDI. Investigation of the interactions with regulatory GEFs and GAPs is beyond the scope of the present thesis. However, it may be instructive to computationally determine whether the tilting of monomeric Rab5 toward the membrane surface or the dimerization has implications on the Rab5 catalytic cycle.

Computational studies like MD simulations contribute to the understanding of protein structure and dynamics and assist characterizing cellular processes. However, experimental studies are recommended and highly appreciated to verify theoretical findings. X-ray crystallography, nuclear magnetic resonance and cryo-electron microscopy are the standard techniques for protein structure determination. The X-ray structures of the Rab5 G domain in its GDP- and GTP-bound states were used as structural templates in the present study. The thesis results and conclusions may be confirmed on the basis of appropriate experimental investigations. In the field of membrane-anchored small GTPases, attenuated total reflectance Fourier transform infrared (ATR-FTIR) and Förster resonance energy transfer (FRET) measurements are frequently used to analyze the orientation of α helices and thus, G domain orientation.[248] Provided that geranylgeranylated Rab5 is available, ATR-FTIR and FRET experiments would be suitable to confirm the monomeric Rab5 tilted conformations observed in the present thesis. Cross-linking/mass spectrometry (CLMS) represents an experimental technique able to investigate the structure of single proteins and protein complexes in their native, physiological environment.[62,67] It is based on the concept that certain amino acid pairs in close proximity to each other are chemically cross-linked by reactive groups connected by a spacer of defined length. The cross-linked proteins are digested by proteases, enriched by liquid chromatography, and the peptides' masses and fragmentation spectra are analyzed by mass spectrometry. Linkers with only one bound residue indicate regions of high solvent accessibility, internally linked peptides reflect local structure elements, and cross-linked peptides provide long-distance information. Thus, CLMS is a valuable tool to investigate Rab5 dimers and Rab5:GDI complexes and able to support the theoretical findings of the present thesis. The switch I Cys^{63} was shown to be close to the dimerization site in Rab5(GDP) and may represent a suitable target for cross-linking. Another experimental setup introduces single point mutations in the protein amino acid sequence in order to probe the relevance of specific residues in complex formation. Mutations of distinct residues in the Rab5 switch regions may confirm the identified protein-protein or protein-lipid interactions.

Its prominent role in endocytosis renders Rab5 a sore point regarding dysfunctions caused by pathogens or mutations. However, interference with Rab5 signaling by small molecule drugs or specific therapies is high-risk due to the need to preserve physiological function. In contrast to Ras proteins, there is no evidence that mutated Rab5 directly causes tumor growth and dissemination. The Rab5 role in malfunctions is more subtle. Constitutively active Rab5 (Q79L) leads to the formation of anomalously enlarged endosomes and distortion of Rab5 membrane extraction.[299] Downregulation of Rab5 results in an impairment of endosome maturation as well as protein internalization and degradation in lysosomes.[3] Mutated dysfunctional Rab5 may be defanged by interference with the C-terminal prenylation motif, thus hindering correct membrane targeting and promoting immediate elimination via degradative pathways. The nucleotide-specific distinct membrane-associated orientations of monomeric Rab5 may represent a potential starting point for selectively targeting Rab5(GDP) or Rab5(GTP), respectively. It needs to be tested, whether the switch regions contain high-affinity binding sites for small drug-like molecules, able to interfere with Rab5 function. The development of inhibitors which compete with the nucleotide for the binding pocket is a common motif in drug development.[300] Although Rab5 dimerization retains the opportunity of binding interaction partners, binding was nevertheless impeded due to less accessible switch regions. Thus, promoting dimerization, e.g. by mutating amino acids involved in protein-protein interactions, may lower Rab5 activity, either by reducing Rab5(GDP) recycling by GDI or by complicating Rab5(GTP)-effector binding. Targeting protein–protein interactions is a common approach in drug development. In case of the highly interwoven Rab5 signaling, it provides the opportunity to selectively perturb one specific interaction thereby reducing unintended side effects.

Bibliography

(1) Li, G. P.; Stahl, P. D. *Journal of Biological Chemistry* **1993**, *268*, 24475–24480.

(2) Stenmark, H.; Valencia, A.; Martinez, O.; Ullrich, O.; Goud, B.; Zerial, M. *Embo Journal* **1994**, *13*, 575–583.

(3) Zerial, M.; McBride, H. *Nature Reviews Molecular Cell Biology* **2001**, *2*, 107–117.

(4) Christoforidis, S.; McBride, H. M.; Burgoyne, R. D.; Zerial, M. *Nature* **1999**, *397*, 621–625.

(5) Pyrzynska, B.; Pilecka, I.; Miaczynska, M. *Molecular Oncology* **2009**, *3*, 321–338.

(6) Miaczynska, M. *Cold Spring Harbor Perspectives in Biology* **2013**, *5*, 20.

(7) Howe, C. L.; Mobley, W. C. *Journal of Neurobiology* **2004**, *58*, 207–216.

(8) Edler, E.; Stein, M. *Journal of Enzyme Inhibition and Medicinal Chemistry* **2017**, *32*, 434–443.

(9) Prada-Delgado, A.; Carrasco-Marin, E.; Pena-Macarro, C.; del Cerro-Vadillo, E.; Fresno-Escudero, M.; Leyva-Cobian, F.; Alvarez-Dominguez, C. *Traffic* **2005**, *6*, 252–265.

(10) Mallo, G. V.; Espina, M.; Smith, A. C.; Terebiznik, M. R.; Aleman, A.; Finlay, B. B.; Rameh, L. E.; Grinstein, S.; Brumell, J. H. *Journal of Cell Biology* **2008**, *182*, 741–752.

(11) Sarantis, H.; Balkin, D. M.; De Camilli, P.; Isberg, R. R.; Brumell, J. H.; Grinstein, S. *Cell Host & Microbe* **2012**, *11*, 117–128.

(12) Coyne, C. B.; Le, S.; Turner, J. R.; Bergelson, J. M. *Cell Host & Microbe* **2007**, *2*, 181–192.

(13) Colacurcio, D. J.; Pensalfini, A.; Jiang, Y.; Nixon, R. A. *Free Radical Biology and Medicine* **2018**, *114*, 40–51.

(14) Selkoe, D. J.; Hardy, J. *Embo Molecular Medicine* **2016**, *8*, 595–608.

(15) Xu, W.; Fang, F.; Ding, J. Q.; Wu, C. B. *Traffic* **2018**, *19*, 253–262.

(16) Cataldo, A. M.; Barnett, J. L.; Pieroni, C.; Nixon, R. A. *Journal of Neuroscience* **1997**, *17*, 6142–6151.

(17) Ginsberg, S. D.; Alldred, M. J.; Counts, S. E.; Cataldo, A. M.; Neve, R. L.; Jiang, Y.; Wuu, J.; Chao, M. V.; Mufson, E. J.; Nixon, R. A.; Che, S. L. *Biological Psychiatry* **2010**, *68*, 885–893.

(18) Lin, D. C.; Quevedo, C.; Brewer, N. E.; Bell, A.; Testa, J. R.; Grimes, M. L.; Miller, F. D.; Kaplan, D. R. *Molecular and Cellular Biology* **2006**, *26*, 8928–8941.

(19) Miaczynska, M.; Christoforidis, S.; Giner, A.; Shevchenko, A.; Uttenweiler-Joseph, S.; Habermann, B.; Wilm, M.; Parton, R. G.; Zerial, M. *Cell* **2004**, *116*, 445–456.

(20) Hupalowska, A.; Pyrzynska, B.; Miaczynska, M. *Journal of Cell Science* **2012**, *125*, 4090–4102.

(21) Bohdanowicz, M.; Balkin, D. M.; De Camilli, P.; Grinstein, S. *Molecular Biology of the Cell* **2012**, *23*, 176–187.

(22) Ryu, J. Y. et al. *Cell Reports* **2014**, *7*, 1227–1238.

(23) Fang, F.; Yang, W. L.; Florio, J. B.; Rockenstein, E.; Spencer, B.; Orain, X. M.; Dong, S. X.; Li, H. Y.; Chen, X. Q.; Sung, K. J.; Rissman, R. A.; Masliah, E.; Ding, J. Q.; Wu, C. B. *Scientific Reports* **2017**, *7*, 13.

(24) Veleri, S.; Punnakkal, P.; Dunbar, G. L.; Maiti, P. *Neuromolecular Medicine* **2018**, *20*, 18–36.

(25) Ravikumar, B.; Imarisio, S.; Sarkar, S.; O'Kane, C. J.; Rubinsztein, D. C. *Journal of Cell Science* **2008**, *121*, 1649–1660.

(26) Torres, V. A.; Stupack, D. G. *Current Protein & Peptide Science* **2011**, *12*, 43–51.

(27) Mendoza, P.; Ortiz, R.; Diaz, J.; Quest, A. F. G.; Leyton, L.; Stupack, D.; Torres, V. A. *Journal of Cell Science* **2013**, *126*, 3835–3847.

(28) Mendoza, P.; Diaz, J.; Torres, V. A. *Current Molecular Medicine* **2014**, *14*, 235–245.

(29) Hajduk, P. J.; Huth, J. R.; Tse, C. *Drug Discovery Today* **2005**, *10*, 1675–1682.

(30) Owens, J. *Nature Reviews Drug Discovery* **2007**, *6*, 187–187.

(31) Cox, A. D.; Fesik, S. W.; Kimmelman, A. C.; Luo, J.; Der, C. J. *Nature Reviews Drug Discovery* **2014**, *13*, 828–851.

(32) Bottaro, S.; Lindorff-Larsen, K. *Science* **2018**, *361*, 355–360.

(33) Alder, B. J.; Wainwright, T. E. *Journal of Chemical Physics* **1957**, *27*, 1208–1209.

(34) Alder, B. J.; Wainwright, T. E. *Journal of Chemical Physics* **1959**, *31*, 459–466.

(35) McCammon, J. A.; Gelin, B. R.; Karplus, M. *Nature* **1977**, *267*, 585–590.

(36) Cisneros, G. A.; Karttunen, M.; Ren, P. Y.; Sagui, C. *Chemical Reviews* **2014**, *114*, 779–814.

(37) Baker, C. M. *Wiley Interdisciplinary Reviews-Computational Molecular Science* **2015**, *5*, 241–254.

(38) McGibbon, R. T.; Taube, A. G.; Donchev, A. G.; Siva, K.; Hernandez, F.; Hargus, C.; Law, K. H.; Klepeis, J. L.; Shaw, D. E. *Journal of Chemical Physics* **2017**, *147*, 15.

(39) Bernardi, R. C.; Melo, M. C. R.; Schulten, K. *Biochimica Et Biophysica Acta-General Subjects* **2015**, *1850*, 872–877.

(40) Durrant, J. D.; McCammon, J. A. *BMC Biology* **2011**, *9*, 9.

(41) Stone, J. E.; Phillips, J. C.; Freddolino, P. L.; Hardy, D. J.; Trabuco, L. G.; Schulten, K. *Journal of Computational Chemistry* **2007**, *28*, 2618–2640.

(42) Stone, J. E.; Hardy, D. J.; Ufimtsev, I. S.; Schulten, K. *Journal of Molecular Graphics & Modelling* **2010**, *29*, 116–125.

(43) Shaw, D. E.; Maragakis, P.; Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Eastwood, M. P.; Bank, J. A.; Jumper, J. M.; Salmon, J. K.; Shan, Y. B.; Wriggers, W. *Science* **2010**, *330*, 341–346.

(44) Shaw, D. E. et al. In *Sc14: International Conference for High Performance Computing, Networking, Storage and Analysis*; International Conference for High Performance Computing Networking Storage and Analysis; Ieee: New York, 2014, pp 41–53.

(45) Shirts, M.; Pande, V. S. *Science* **2000**, *290*, 1903–1904.

(46) Husic, B. E.; Pande, V. S. *Journal of the American Chemical Society* **2018**, *140*, 2386–2396.

(47) Perricone, U.; Gulotta, M. R.; Lombino, J.; Parrino, B.; Cascioferro, S.; Diana, P.; Cirrincione, G.; Padova, A. *MedChemComm* **2018**, *9*, 920–936.

(48) Seco, J.; Luque, F. J.; Barril, X. *Journal of Medicinal Chemistry* **2009**, *52*, 2363–2371.

(49) Bakan, A.; Nevins, N.; Lakdawala, A. S.; Bahar, I. *Journal of Chemical Theory and Computation* **2012**, *8*, 2435–2447.

(50) Ferruz, N.; Doerr, S.; Vanase-Frawley, M. A.; Zou, Y. Z.; Chen, X. M.; Marr, E. S.; Nelson, R. T.; Kormos, B. L.; Wager, T. T.; Hou, X. J.; Villalobos, A.; Sciabola, S.; De Fabritiis, G. *Scientific Reports* **2018**, *8*, 10.

(51) Makala, H.; Ulaganathan, V. *Journal of Receptors and Signal Transduction* **2018**, *38*, 12–19.

(52) Tan, Y. S.; Spring, D. R.; Abell, C.; Verma, C. S. *Journal of Chemical Theory and Computation* **2015**, *11*, 3199–3210.

(53) Votapka, L. W.; Jagger, B. R.; Heyneman, A. L.; Amaro, R. E. *Journal of Physical Chemistry B* **2017**, *121*, 3597–3606.

(54) Amaro, R. E.; Mulholland, A. J. *Nature Reviews Chemistry* **2018**, *2*, 12.

(55) Lodola, A.; Mor, M.; Rivara, S.; Christov, C.; Tarzia, G.; Piomelli, D.; Mulholland, A. J. *Chemical Communications* **2008**, 214–216.

(56) Lonsdale, R.; Houghton, K. T.; Zurek, J.; Bathelt, C. M.; Foloppe, N.; de Groot, M. J.; Harvey, J. N.; Mulholland, A. J. *Journal of the American Chemical Society* **2013**, *135*, 8001–8015.

(57) Lonsdale, R.; Rouse, S. L.; Sansom, M. S. P.; Mulholland, A. J. *Plos Computational Biology* **2014**, *10*, 16.

(58) Katiyar, R. S.; Jha, P. K. *Wiley Interdisciplinary Reviews-Computational Molecular Science* **2018**, *8*, 18.

(59) Alper, H. E.; Stouch, T. R. *Journal of Physical Chemistry* **1995**, *99*, 5724–5731.

(60) Krarup, L. H.; Christensen, I. T.; Hovgaard, L.; Frokjaer, S. *Pharmaceutical Research* **1998**, *15*, 972–978.

(61) Warren, D. B.; King, D.; Benameur, H.; Pouton, C. W.; Chalmers, D. K. *Pharmaceutical Research* **2013**, *30*, 3238–3253.

(62) Schneider, M.; Belsom, A.; Rappsilber, J. *Trends in Biochemical Sciences* **2018**, *43*, 157–169.

(63) McPherson, A. *Methods* **2004**, *34*, 254–265.

(64) Caffrey, M. *Journal of Structural Biology* **2003**, *142*, 108–132.

(65) Bai, X. C.; McMullan, G.; Scheres, S. H. W. *Trends in Biochemical Sciences* **2015**, *40*, 49–57.

(66) Korasick, D. A.; Tanner, J. J. *Protein Science* **2018**, *27*, 814–824.

(67) Rappsilber, J. *Journal of Structural Biology* **2011**, *173*, 530–540.

(68) Ginalski, K. *Current Opinion in Structural Biology* **2006**, *16*, 172–177.

(69) Simons, K. T.; Bonneau, R.; Ruczinski, I.; Baker, D. *Proteins-Structure Function and Bioinformatics* **1999**, 171–176.

(70) Woolf, T. B.; Roux, B. *Proceedings of the National Academy of Sciences of the United States of America* **1994**, *91*, 11631–11635.

(71) Edholm, O.; Berger, O.; Jahnig, F. *Journal of Molecular Biology* **1995**, *250*, 94–111.

(72) Terzyan, S.; Zhu, G. Y.; Li, G. P.; Zhang, X. J. C. *Acta Crystallographica Section D-Biological Crystallography* **2004**, *60*, 54–60.

(73) Zhu, G. Y.; Zhai, P.; Liu, J.; Terzyan, S.; Li, G. P.; Zhang, X. J. C. *Nature Structural & Molecular Biology* **2004**, *11*, 975–983.

(74) Brunsveld, L.; Kuhlmann, J.; Alexandrov, K.; Wittinghofer, A.; Goody, R. S.; Waldmann, H. *Angewandte Chemie-International Edition* **2006**, *45*, 6622–6646.

(75) Wittinghofer, A., *Ras Superfamily Small G Proteins: Biology and Mechanisms 1 : General Features, Signaling*; Springer-Verlag Wien: Vienna, 2014.

(76) Shih, T. Y.; Weeks, M. O.; Young, H. A.; Scolnick, E. M. *Virology* **1979**, *96*, 64–79.

(77) Takai, Y.; Sasaki, T.; Matozaki, T. *Physiological Reviews* **2001**, *81*, 153–208.

(78) Reuther, G. W.; Der, C. J. *Current Opinion in Cell Biology* **2000**, *12*, 157–165.

(79) Zhen, Y.; Stenmark, H. *Journal of Cell Science* **2015**, *128*, 3171–3176.

(80) Hodge, R. G.; Ridley, A. J. *Nature Reviews Molecular Cell Biology* **2016**, *17*, 496–510.

(81) Jackson, C. L. In *Ras Superfamily Small G Proteins: Biology and Mechanisms 2: Transport*, Wittinghofer, A., Ed.; Springer International Publishing: Cham, 2014, pp 151–180.

(82) Stewart, M. In *Ras Superfamily Small G Proteins: Biology and Mechanisms 2: Transport*, Wittinghofer, A., Ed.; Springer International Publishing: Cham, 2014, pp 109–124.

(83) Jewell, J. L.; Guan, K.-L. In *Ras Superfamily Small G Proteins: Biology and Mechanisms 2: Transport*, Wittinghofer, A., Ed.; Springer International Publishing: Cham, 2014, pp 277–292.

(84) Kim, J.; Kim, E. *Amino Acids* **2016**, *48*, 915–928.

(85) Brunsveld, L.; Waldmann, H.; Huster, D. *Biochimica Et Biophysica Acta-Biomembranes* **2009**, *1788*, 273–288.

(86) Vetter, I. R. In *Ras Superfamily Small G Proteins: Biology and Mechanisms 1: General Features, Signaling*, Wittinghofer, A., Ed.; Springer-Verlag Wien: Vienna, 2014, pp 25–50.

(87) Stenmark, H.; Olkkonen, V. M. *Genome Biology* **2001**, *2*, 7.

(88) Stenmark, H. *Nature Reviews Molecular Cell Biology* **2009**, *10*, 513–525.

(89) Galvez, T.; Gilleron, J.; Zerial, M.; O'Sullivan, G. A. *Cell* **2012**, *151*, 234–U244.

(90) Tokarev, A. A.; Alfonso, A.; Segev, N. In *Trafficking Inside Cells: Pathways, Mechanisms and Regulation*; Springer New York: New York, NY, 2009, pp 3–14.

(91) Lim, J. P.; Gleeson, P. A. *Immunology and Cell Biology* **2011**, *89*, 836–843.

(92) Schwartz, S. L.; Cao, C.; Pylypenko, O.; Rak, A.; Wandinger-Ness, A. *Journal of Cell Science* **2007**, *120*, 3905–3910.

(93) Stein, M.; Pilli, M.; Bernauer, S.; Habermann, B. H.; Zerial, M.; Wade, R. C. *PLoS One* **2012**, *7*, e34870.

(94) Baron, R. A.; Seabra, M. C. *Biochemical Journal* **2008**, *415*, 67–75.

(95) Liang, P. H.; Ko, T. P.; Wang, A. H. J. *European Journal of Biochemistry* **2002**, *269*, 3339–3354.

(96) Klebe, C.; Prinz, H.; Wittinghofer, A.; Goody, R. S. *Biochemistry* **1995**, *34*, 12543–12552.

(97) Chin, H. F.; Cai, Y. Y.; Menon, S.; Ferro-Novick, S.; Reinisch, K. M.; De La Cruz, E. M. *Journal of Molecular Biology* **2009**, *389*, 275–288.

(98) Goody, R. S.; Hofmann-Goody, W. *European Biophysics Journal with Biophysics Letters* **2002**, *31*, 268–274.

(99) Guo, Z.; Ahmadian, M. R.; Goody, R. S. *Biochemistry* **2005**, *44*, 15423–15429.

(100) Cherfils, J.; Zeghouf, M. *Physiological Reviews* **2013**, *93*, 269–309.

(101) Van Meer, G.; Voelker, D. R.; Feigenson, G. W. *Nature Reviews Molecular Cell Biology* **2008**, *9*, 112–124.

(102) Edidin, M. *Nature Reviews Molecular Cell Biology* **2003**, *4*, 414–418.

(103) Gorter, E.; Grendel, F. *Journal of Experimental Medicine* **1925**, *41*, 439–443.

(104) Danielli, J. F.; Davson, H. *Journal of Cellular and Comparative Physiology* **1935**, *5*, 495–508.

(105) Chapman, D. *Quarterly Reviews of Biophysics* **1975**, *8*, 185–235.

(106) Singer, S. J.; Nicolson, G. L. *Science* **1972**, *175*, 720–731.

(107) Simons, K.; Ikonen, E. *Nature* **1997**, *387*, 569–572.

(108) Fahy, E. et al. *Journal of Lipid Research* **2005**, *46*, 839–861.

(109) Chiantia, S.; London, E. *Handbook of experimental pharmacology* **2013**, 33–55.

(110) Degroote, S.; Wolthoorn, J.; van Meer, G. *Seminars in Cell & Developmental Biology* **2004**, *15*, 375–387.

(111) Van Meer, G. *Embo Journal* **2005**, *24*, 3159–3165.

(112) Ingolfsson, H. I.; Melo, M. N.; van Eerden, F. J.; Arnarez, C.; Lopez, C. A.; Wassenaar, T. A.; Periole, X.; de Vries, A. H.; Tieleman, D. P.; Marrink, S. J. *Journal of the American Chemical Society* **2014**, *136*, 14554–14559.

(113) Huotari, J.; Helenius, A. *Embo Journal* **2011**, *30*, 3481–3500.

(114) Shin, H. W.; Hayashi, M.; Christoforidis, S.; Lacas-Gervais, S.; Hoepfner, S.; Wenk, M. R.; Modregger, J.; Uttenweiler-Joseph, S.; Wilm, M.; Nystuen, A.; Frankel, W. N.; Solimena, M.; De Camilli, P.; Zerial, M. *Journal of Cell Biology* **2005**, *170*, 607–618.

(115) Zhang, Z.; Zhang, T. L.; Wang, S. S.; Gong, Z.; Tang, C.; Chen, J. Y.; Ding, J. P. *Elife* **2014**, *3*, 45.

(116) Lybrand, T. P. In *Reviews in Computational Chemistry*; John Wiley & Sons, Inc.: 1990, pp 295–320.

(117) Lewars, E. G. In *Computational Chemistry: Introduction to the Theory and Applications of Molecular and Quantum Mechanics*; Springer Netherlands: Dordrecht, 2011, pp 45–83.

(118) Karplus, M.; McCammon, J. A. *Nature Structural Biology* **2002**, *9*, 646–652.

(119) Paquet, E.; Viktor, H. L. *BioMed Research International* **2015**, 18.

(120) Verlet, L. *Physical Review* **1967**, *159*, 98–103.

(121) Petrenko, R.; Meller, J. In *eLS*; John Wiley & Sons, Ltd: 2001.

(122) Darden, T.; York, D.; Pedersen, L. *Journal of Chemical Physics* **1993**, *98*, 10089–10092.

(123) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *Journal of Chemical Physics* **1995**, *103*, 8577–8593.

(124) Ewald, P. P. *Annalen Der Physik* **1921**, *64*, 253–287.

(125) Benkovic, S. J.; Hammes-Schiffer, S. *Science* **2003**, *301*, 1196–1202.

(126) Kandt, C.; Monticelli, L. In *Membrane Protein Structure Determination: Methods and Protocols*, Lacapere, J. J., Ed.; Methods in Molecular Biology, Vol. 654; Humana Press Inc: Totowa, 2010, pp 423–440.

(127) Toxvaerd, S.; Heilmann, O. J.; Dyre, J. C. *Journal of Chemical Physics* **2012**, *136*, 8.

(128) Holian, B. L.; Hoover, W. G.; Posch, H. A. *Physical Review Letters* **1987**, *59*, 10–13.

(129) Calvo, F. *Physical Review E* **1999**, *60*, 2771–2778.

(130) Dellago, C.; Posch, H. A., *Realizing Boltzmann's dream: computer simulations in modern statistical mechanics*; Boltzmann's Legacy; European Mathematical Soc: Zurich, 2008, pp 171–202.

(131) Grebogi, C.; Hammel, S. M.; Yorke, J. A.; Sauer, T. *Physical Review Letters* **1990**, *65*, 1527–1530.

(132) Kastner, J. *Wiley Interdisciplinary Reviews-Computational Molecular Science* **2011**, *1*, 932–942.

(133) Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M. *Journal of Computational Chemistry* **1992**, *13*, 1011–1021.

(134) Souaille, M.; Roux, B. *Computer Physics Communications* **2001**, *135*, 40–57.

(135) Marrink, S. J.; Risselada, H. J.; Yefimov, S.; Tieleman, D. P.; de Vries, A. H. *Journal of Physical Chemistry B* **2007**, *111*, 7812–7824.

(136) Marrink, S. J.; de Vries, A. H.; Mark, A. E. *Journal of Physical Chemistry B* **2004**, *108*, 750–760.

(137) Marrink, S. J.; Tieleman, D. P. *Chemical Society Reviews* **2013**, *42*, 6801–6822.

(138) De Jong, D. H.; Singh, G.; Bennett, W. F. D.; Arnarez, C.; Wassenaar, T. A.; Schafer, L. V.; Periole, X.; Tieleman, D. P.; Marrink, S. J. *Journal of Chemical Theory and Computation* **2013**, *9*, 687–697.

(139) Ramadurai, S.; Holt, A.; Schafer, L. V.; Krasnikov, V. V.; Rijkers, D. T. S.; Marrink, S. J.; Killian, J. A.; Poolman, B. *Biophysical Journal* **2010**, *99*, 1447–1454.

(140) Sali, A.; Blundell, T. L. *Journal of Molecular Biology* **1993**, *234*, 779–815.

(141) Jo, S.; Kim, T.; Iyer, V. G.; Im, W. *Journal of Computational Chemistry* **2008**, *29*, 1859–1865.

(142) Wu, E. L.; Cheng, X.; Jo, S.; Rui, H.; Song, K. C.; Davila-Contreras, E. M.; Qi, Y. F.; Lee, J. M.; Monje-Galvan, V.; Venable, R. M.; Klauda, J. B.; Im, W. *Journal of Computational Chemistry* **2014**, *35*, 1997–2004.

(143) Humphrey, W.; Dalke, A.; Schulten, K. *Journal of Molecular Graphics & Modelling* **1996**, *14*, 33–38.

(144) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. *Journal of Computational Chemistry* **2005**, *26*, 1781–1802.

(145) Michaud-Agrawal, N.; Denning, E. J.; Woolf, T. B.; Beckstein, O. *Journal of Computational Chemistry* **2011**, *32*, 2319–2327.

(146) Manual for Package PGFPLOTS, 2D/3D Plots in Latex, Version 1.5.1 DEV, Dr. Christian Feuersänger, April 22, 2012.

(147) Andersen, H. C. *Journal of Chemical Physics* **1980**, *72*, 2384–2393.

(148) Berendsen, H. J. C.; Postma, J. P. M.; Vangunsteren, W. F.; Dinola, A.; Haak, J. R. *Journal of Chemical Physics* **1984**, *81*, 3684–3690.

(149) Kandt, C.; Ash, W. L.; Tieleman, D. P. *Methods* **2007**, *41*, 475–488.

(150) Van der Bliek, A. M. *Nature Cell Biology* **2005**, *7*, 548–550.

(151) Tall, G. G.; Barbieri, M. A.; Stahl, P. D.; Horazdovsky, B. F. *Developmental Cell* **2001**, *1*, 73–82.

(152) Lanzetti, L.; Palamidessi, A.; Areces, L.; Scita, G.; Di Fiore, P. P. *Nature* **2004**, *429*, 309–314.

(153) Frittoli, E. et al. *Journal of Cell Biology* **2014**, *206*, 307–328.

(154) Prakash, P.; Gorfe, A. A. *Biochimica Et Biophysica Acta-General Subjects* **2013**, *1830*, 5211–5218.

(155) Kalli, A. C.; Sansom, M. S. P. *Biochemical Society Transactions* **2014**, *42*, 1418–1424.

(156) Edler, E.; Schulze, E.; Stein, M. *Biochimica Et Biophysica Acta-Biomembranes* **2017**, *1859*, 1335–1349.

(157) Webb, B.; Sali, A. In *Protein Structure Prediction, 3rd Edition*, Kihara, D., Ed.; Methods in Molecular Biology, Vol. 1137; Humana Press Inc: Totowa, 2014, pp 1–15.

(158) Shen, M. Y.; Sali, A. *Protein Science* **2006**, *15*, 2507–2524.

(159) Jorgensen, W. L.; Maxwell, D. S.; TiradoRives, J. *Journal of the American Chemical Society* **1996**, *118*, 11225–11236.

(160) Vetter, I. R.; Wittinghofer, A. *Science* **2001**, *294*, 1299–1304.

(161) Jo, S.; Lim, J. B.; Klauda, J. B.; Im, W. *Biophysical Journal* **2009**, *97*, 50–58.

(162) Best, R. B.; Zhu, X.; Shim, J.; Lopes, P. E. M.; Mittal, J.; Feig, M.; MacKerell, A. D. *Journal of Chemical Theory and Computation* **2012**, *8*, 3257–3273.

(163) Pastor, R. W.; MacKerell, A. D. *Journal of Physical Chemistry Letters* **2011**, *2*, 1526–1532.

(164) Weigend, F.; Ahlrichs, R. *Physical Chemistry Chemical Physics* **2005**, *7*, 3297–3305.

(165) Perdew, J. P. *Physical Review B* **1986**, *33*, 8822–8824.

(166) Becke, A. D. *Physical Review A* **1988**, *38*, 3098–3100.

(167) Schafer, A.; Klamt, A.; Sattel, D.; Lohrenz, J. C. W.; Eckert, F. *Physical Chemistry Chemical Physics* **2000**, *2*, 2187–2193.

(168) Tanizaki, S.; Feig, M. *Journal of Chemical Physics* **2005**, *122*, 13.

(169) Huster, D.; Vogel, A.; Katzka, C.; Scheidt, H. A.; Binder, H.; Dante, S.; Gutberlet, T.; Zschornig, O.; Waldmann, H.; Arnold, K. *Journal of the American Chemical Society* **2003**, *125*, 4070–4079.

(170) Gorfe, A. A.; Pellarin, R.; Caflisch, A. *Journal of the American Chemical Society* **2004**, *126*, 15277–15286.

(171) Reuther, G.; Tan, K. T.; Kohler, J.; Nowak, C.; Pampel, A.; Arnold, K.; Kuhlmann, J.; Waldmann, H.; Huster, D. *Angewandte Chemie-International Edition* **2006**, *45*, 5387–5390.

(172) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *Journal of Chemical Physics* **1983**, *79*, 926–935.

(173) Beglov, D.; Roux, B. *Journal of Chemical Physics* **1994**, *100*, 9050–9063.

(174) Hatcher, E. R.; Guvench, O.; MacKerell, A. D. *Journal of Chemical Theory and Computation* **2009**, *5*, 1315–1327.

(175) Klauda, J. B.; Venable, R. M.; Freites, J. A.; O'Connor, J. W.; Tobias, D. J.; Mondragon-Ramirez, C.; Vorobyov, I.; MacKerell, A. D.; Pastor, R. W. *Journal of Physical Chemistry B* **2010**, *114*, 7830–7843.

(176) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; MacKerell, A. D. *Journal of Computational Chemistry* **2010**, *31*, 671–690.

(177) Hart, K.; Foloppe, N.; Baker, C. M.; Denning, E. J.; Nilsson, L.; MacKerell, A. D. *Journal of Chemical Theory and Computation* **2012**, *8*, 348–362.

(178) Lim, J. B.; Rogaski, B.; Klauda, J. B. *Journal of Physical Chemistry B* **2012**, *116*, 203–210.

(179) Venable, R. M.; Sodt, A. J.; Rogaski, B.; Rui, H.; Hatcher, E.; MacKerell, A. D.; Pastor, R. W.; Klauda, J. B. *Biophysical Journal* **2014**, *107*, 134–145.

(180) Foloppe, N.; MacKerell, A. D. *Journal of Computational Chemistry* **2000**, *21*, 86–104.

(181) Pavelites, J. J.; Gao, J. L.; Bash, P. A.; Mackerell, A. D. *Journal of Computational Chemistry* **1997**, *18*, 221–239.

(182) Van der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. *Journal of Computational Chemistry* **2005**, *26*, 1701–1718.

(183) Wassenaar, T. A.; Ingolfsson, H. I.; Bockmann, R. A.; Tieleman, D. P.; Marrink, S. J. *Journal of Chemical Theory and Computation* **2015**, *11*, 2144–2155.

(184) De Jong, D. H.; Baoukina, S.; Ingolfsson, H. I.; Marrink, S. J. *Computer Physics Communications* **2016**, *199*, 1–7.

(185) Parrinello, M.; Rahman, A. *Journal of Applied Physics* **1981**, *52*, 7182–7190.

(186) Monticelli, L.; Kandasamy, S. K.; Periole, X.; Larson, R. G.; Tieleman, D. P.; Marrink, S. J. *Journal of Chemical Theory and Computation* **2008**, *4*, 819–834.

(187) Kabsch, W.; Sander, C. *Biopolymers* **1983**, *22*, 2577–2637.

(188) Siu, S. W. I.; Vacha, R.; Jungwirth, P.; Bockmann, R. A. *Journal of Chemical Physics* **2008**, *128*, 12.

(189) Douliez, J. P.; Ferrarini, A.; Dufourc, E. J. *Journal of Chemical Physics* **1998**, *109*, 2513–2518.

(190) De Berg, M.; Cheong, O.; van Kreveld, M.; Overmars, M. In *Computational Geometry: Algorithms and Applications*; Springer Berlin Heidelberg: Berlin, Heidelberg, 2008, pp 147–171.

(191) Allen, W. J.; Lemkul, J. A.; Bevan, D. R. *Journal of Computational Chemistry* **2009**, *30*, 1952–1958.

(192) Codling, E. A.; Plank, M. J.; Benhamou, S. *Journal of the Royal Society Interface* **2008**, *5*, 813–834.

(193) Almeida, P. F. F.; Vaz, W. L. C., *Lateral diffusion in membranes*; Handbook of Biological Physics, Vol. 1; Structure and dynamics of membranes: From cells to vesicles; Elsevier Science Publishers B.V.: Sara Burgerhartstraat 25, 1000 AE Amsterdam, Netherlands, 1995, pp 305–357.

(194) Saxton, M. J. *Biophysical Journal* **1997**, *72*, 1744–1753.

(195) Grunwald, D.; Martin, R. M.; Buschmann, V.; Bazett-Jones, D. P.; Leonhardt, H.; Kubitscheck, U.; Cardoso, M. C. *Biophysical Journal* **2008**, *94*, 2847–2858.

(196) Siebrasse, J. P.; Veith, R.; Dobay, A.; Leonhardt, H.; Daneholt, B.; Kubitscheck, U. *Proceedings of the National Academy of Sciences of the United States of America* **2008**, *105*, 20291–20296.

(197) Sonnleitner, A.; Schutz, G. J.; Schmidt, T. *Biophysical Journal* **1999**, *77*, 2638–2642.

(198) Guigas, G.; Weiss, M. *Biophysical Journal* **2006**, *91*, 2393–2398.

(199) Weimann, L.; Ganzinger, K. A.; McColl, J.; Irvine, K. L.; Davis, S. J.; Gay, N. J.; Bryant, C. E.; Klenerman, D. *PLoS One* **2013**, *8*, 9.

(200) Dolinsky, T. J.; Nielsen, J. E.; McCammon, J. A.; Baker, N. A. *Nucleic Acids Research* **2004**, *32*, W665–W667.

(201) Dolinsky, T. J.; Czodrowski, P.; Li, H.; Nielsen, J. E.; Jensen, J. H.; Klebe, G.; Baker, N. A. *Nucleic Acids Research* **2007**, *35*, W522–W525.

(202) Baker, N. A.; Sept, D.; Joseph, S.; Holst, M. J.; McCammon, J. A. *Proceedings of the National Academy of Sciences of the United States of America* **2001**, *98*, 10037–10041.

(203) Li, H.; Robertson, A. D.; Jensen, J. H. *Proteins-Structure Function and Bioinformatics* **2005**, *61*, 704–721.

(204) Miller, S.; Janin, J.; Lesk, A. M.; Chothia, C. *Journal of Molecular Biology* **1987**, *196*, 641–656.

(205) Ichiye, T.; Karplus, M. *Proteins-Structure Function and Genetics* **1991**, *11*, 205–217.

(206) Mu, Y. G.; Nguyen, P. H.; Stock, G. *Proteins-Structure Function and Bioinformatics* **2005**, *58*, 45–52.

(207) Altis, A.; Nguyen, P. H.; Hegger, R.; Stock, G. *Journal of Chemical Physics* **2007**, *126*, 10.

(208) Sittel, F.; Jain, A.; Stock, G. *Journal of Chemical Physics* **2014**, *141*, 9.

(209) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. *Journal of Computational Chemistry* **1995**, *16*, 1339–1350.

(210) Roux, B. *Computer Physics Communications* **1995**, *91*, 275–282.

(211) Krause, M. R.; Regen, S. L. *Accounts of Chemical Research* **2014**, *47*, 3512–3521.

(212) Filippov, A.; Oradd, G.; Lindblom, G. *Biophysical Journal* **2003**, *84*, 3079–3086.

(213) Vogel, A.; Tan, K. T.; Waldmann, H.; Feller, S. E.; Brown, M. F.; Huster, D. *Biophysical Journal* **2007**, *93*, 2697–2712.

(214) Reuther, G.; Tan, K. T.; Vogel, A.; Nowak, C.; Arnold, K.; Kuhlmann, J.; Waldmann, H.; Huster, D. *Journal of the American Chemical Society* **2006**, *128*, 13840–13846.

(215) Vogel, A.; Katzka, C. P.; Waldmann, H.; Arnold, K.; Brown, M. F.; Huster, D. *Journal of the American Chemical Society* **2005**, *127*, 12263–12272.

(216) Silvius, J. R.; Lheureux, F. *Biochemistry* **1994**, *33*, 3014–3022.

(217) Shahinian, S.; Silvius, J. R. *Biochemistry* **1995**, *34*, 3813–3822.

(218) Neale, C.; Bennett, W. F. D.; Tieleman, D. P.; Pomes, R. *Journal of Chemical Theory and Computation* **2011**, *7*, 4175–4188.

(219) Neale, C.; Pomes, R. *Biochimica Et Biophysica Acta-Biomembranes* **2016**, *1858*, 2539–2548.

(220) Filipe, H. A. L.; Moreno, M. J.; Rog, T.; Vattulainen, I.; Loura, L. M. S. *Journal of Physical Chemistry B* **2014**, *118*, 3572–3581.

(221) Kempf, N.; Postupalenko, V.; Bora, S.; Didier, P.; Arntz, Y.; de Rocquigny, H.; Mely, Y. *Journal of Virology* **2015**, *89*, 1756–1767.

(222) Koldso, H.; Shorthouse, D.; Helie, J.; Sansom, M. S. P. *Plos Computational Biology* **2014**, *10*, 11.

(223) Van den Bogaart, G.; Meyenberg, K.; Risselada, H. J.; Amin, H.; Willig, K. I.; Hubrich, B. E.; Dier, M.; Hell, S. W.; Grubmuller, H.; Diederichsen, U.; Jahn, R. *Nature* **2011**, *479*, 552–555.

(224) Yeagle, P. L.; Young, J.; Rice, D. *Biochemistry* **1988**, *27*, 6449–6452.

(225) Laursen, M.; Yatime, L.; Nissen, P.; Fedosova, N. U. *Proceedings of the National Academy of Sciences of the United States of America* **2013**, *110*, 10958–10963.

(226) Cherezov, V.; Rosenbaum, D. M.; Hanson, M. A.; Rasmussen, S. G. F.; Thian, F. S.; Kobilka, T. S.; Choi, H. J.; Kuhn, P.; Weis, W. I.; Kobilka, B. K.; Stevens, R. C. *Science* **2007**, *318*, 1258–1265.

(227) Hanson, M. A.; Cherezov, V.; Griffith, M. T.; Roth, C. B.; Jaakola, V. P.; Chien, E. Y. T.; Velasquez, J.; Kuhn, P.; Stevens, R. C. *Structure* **2008**, *16*, 897–905.

(228) Liu, W.; Chun, E.; Thompson, A. A.; Chubukov, P.; Xu, F.; Katritch, V.; Han, G. W.; Roth, C. B.; Heitman, L. H.; Ijzerman, A. P.; Cherezov, V.; Stevens, R. C. *Science* **2012**, *337*, 232–236.

(229) Boesze-Battaglia, K.; Albert, A. D. *Journal of Biological Chemistry* **1990**, *265*, 20727–20730.

(230) Khelashvili, G.; Grossfield, A.; Feller, S. E.; Pitman, M. C.; Weinstein, H. *Proteins-Structure Function and Bioinformatics* **2009**, *76*, 403–417.

(231) Yeagle, P. L. *Biochimica Et Biophysica Acta-Biomembranes* **2014**, *1838*, 1548–1559.

(232) BenTal, N.; Honig, B.; Peitzsch, R. M.; Denisov, G.; McLaughlin, S. *Biophysical Journal* **1996**, *71*, 561–575.

(233) Gorfe, A. A.; Hanzal-Bayer, M.; Abankwa, D.; Hancock, J. F.; McCammon, J. A. *Journal of Medicinal Chemistry* **2007**, *50*, 674–684.

(234) Prakash, P.; Zhou, Y.; Liang, H.; Hancock, J. F.; Gorfe, A. A. *Biophysical Journal* **2016**, *110*, 1125–1138.

(235) Prakash, P.; Gorfe, A. A. *Small GTPases* **2016**, 1–10.

(236) Mulgrew-Nesbitt, A.; Diraviyam, K.; Wang, J. Y.; Singh, S.; Murray, P.; Li, Z. H.; Rogers, L.; Mirkovic, N.; Murray, D. *Biochimica Et Biophysica Acta-Molecular and Cell Biology of Lipids* **2006**, *1761*, 812–826.

(237) Rogaski, B.; Klauda, J. B. *Journal of Molecular Biology* **2012**, *423*, 847–861.

(238) Gomes, A. Q.; Ali, B. R.; Ramalho, J. S.; Godfrey, R. F.; Barral, D. C.; Hume, A. N.; Seabra, M. C. *Molecular Biology of the Cell* **2003**, *14*, 1882–1899.

(239) Murray, D.; Arbuzova, A.; Honig, B.; McLaughlin, S. In *Peptide-Lipid Interactions*, Simon, S. A., McIntosh, T. J., Eds.; Current Topics in Membranes, Vol. 52; Elsevier Academic Press Inc: San Diego, 2002, pp 277–307.

(240) Peters, C.; Wagner, M.; Volkert, M.; Waldmann, H. *Naturwissenschaften* **2002**, *89*, 381–390.

(241) Silvius, J. R. *Peptide-Lipid Interactions* **2002**, *52*, 371–395.

(242) Ali, B. R.; Wasmeier, C.; Lamoreux, L.; Strom, M.; Seabra, M. C. *Journal of Cell Science* **2004**, *117*, 6401–6412.

(243) Leung, K. F.; Baron, R.; Seabra, M. C. *Journal of Lipid Research* **2006**, *47*, 467–475.

(244) Li, F.; Yi, L.; Zhao, L.; Itzen, A.; Goody, R. S.; Wu, Y. W. *Proceedings of the National Academy of Sciences of the United States of America* **2014**, *111*, 2572–2577.

(245) Santos, E.; Nebreda, A. R.; Bryan, T.; Kempner, E. S. *Journal of Biological Chemistry* **1988**, *263*, 9853–9858.

(246) Inouye, K.; Mizutani, S.; Koide, H.; Kaziro, Y. *Journal of Biological Chemistry* **2000**, *275*, 3737–3740.

(247) Nan, X. L.; Tamgueney, T. M.; Collisson, E. A.; Lin, L. J.; Pitt, C.; Galeas, J.; Lewis, S.; Gray, J. W.; McCormick, F.; Chu, S. *Proceedings of the National Academy of Sciences of the United States of America* **2015**, *112*, 7996–8001.

(248) Guldenhaupt, J.; Rudack, T.; Bachler, P.; Mann, D.; Triola, G.; Waldmann, H.; Kotting, C.; Gerwert, K. *Biophysical Journal* **2012**, *103*, 1585–1593.

(249) Lin, W. C.; Iversen, L.; Tu, H. L.; Rhodes, C.; Christensen, S. M.; Iwig, J. S.; Hansen, S. D.; Huang, W. Y. C.; Groves, J. T. *Proceedings of the National Academy of Sciences of the United States of America* **2014**, *111*, 2996–3001.

(250) Prakash, P.; Sayyed-Ahmad, A.; Cho, K. J.; Dolino, D. M.; Chen, W.; Li, H. Y.; Grant, B. J.; Hancock, J. F.; Gorfe, A. A. *Scientific Reports* **2017**, *7*, 11.

(251) Wittmann, J. G.; Rudolph, M. G. *Febs Letters* **2004**, *568*, 23–29.

(252) Beck, R.; Sun, Z.; Adolf, F.; Rutz, C.; Bassler, J.; Wild, K.; Sinning, I.; Hurt, E.; Brugger, B.; Bethune, J.; Wieland, F. *Proceedings of the National Academy of Sciences of the United States of America* **2008**, *105*, 11731–11736.

(253) Hariri, H.; Bhattacharya, N.; Johnson, K.; Noble, A. J.; Stagg, S. M. *Journal of Molecular Biology* **2014**, *426*, 3811–3826.

(254) Zhang, B. L.; Zheng, Y. *Journal of Biological Chemistry* **1998**, *273*, 25728–25733.

(255) Kang, P. J.; Beven, L.; Hariharan, S.; Park, H. O. *Molecular Biology of the Cell* **2010**, *21*, 3007–3016.

(256) Daitoku, H.; Isida, J.; Fujiwara, K.; Nakajima, T.; Fukamizu, A. *International Journal of Molecular Medicine* **2001**, *8*, 397–404.

(257) Zhou, Y.; Prakash, P.; Gorfe, A. A.; Hancock, J. F. *Cold Spring Harbor Perspectives in Medicine* **2017**, *8*, a031831.

(258) Muratcioglu, S.; Chavan, T. S.; Freed, B. C.; Jang, H.; Khavrutskii, L.; Freed, R. N.; Dyba, M. A.; Stefanisko, K.; Tarasov, S. G.; Gursoy, A.; Keskin, O.; Tarasova, N. I.; Gaponenko, V.; Nussinov, R. *Structure* **2015**, *23*, 1325–1335.

(259) Maurer, T. et al. *Proceedings of the National Academy of Sciences of the United States of America* **2012**, *109*, 5299–5304.

(260) Prior, I. A.; Muncke, C.; Parton, R. G.; Hancock, J. F. *Journal of Cell Biology* **2003**, *160*, 165–170.

(261) Nan, X. L.; Collisson, E. A.; Lewis, S.; Huang, J.; Tamguney, T. M.; Liphardt, J. T.; McCormick, F.; Gray, J. W.; Chu, S. *Proceedings of the National Academy of Sciences of the United States of America* **2013**, *110*, 18519–18524.

(262) Wassenaar, T. A.; Pluhackova, K.; Moussatova, A.; Sengupta, D.; Marrink, S. J.; Tieleman, D. P.; Bockmann, R. A. *Journal of Chemical Theory and Computation* **2015**, *11*, 2278–2291.

(263) Pluhackova, K.; Gahbauer, S.; Kranz, F.; Wassenaar, T. A.; Bockmann, R. A. *Plos Computational Biology* **2016**, *12*, 25.

(264) Gahbauer, S.; Pluhackova, K.; Btockmann, R. A. *Plos Computational Biology* **2018**, *14*, 30.

(265) Santos, E. *Science Signaling* **2014**, *7*, 2.

(266) Lo Conte, L.; Chothia, C.; Janin, J. *Journal of Molecular Biology* **1999**, *285*, 2177–2198.

(267) Merithew, E.; Hatherly, S.; Dumas, J. J.; Lawe, D. C.; Heller-Harrison, R.; Lambright, D. G. *Journal of Biological Chemistry* **2001**, *276*, 13982–13988.

(268) Ostrem, J. M.; Peters, U.; Sos, M. L.; Wells, J. A.; Shokat, K. M. *Nature* **2013**, *503*, 548–+.

(269) Jang, H.; Muratcioglu, S.; Gursoy, A.; Keskin, O.; Nussinov, R. *Biochemical Journal* **2016**, *473*, 1719–1732.

(270) Pasqualato, S.; Senic-Matuglia, F.; Renault, L.; Goud, B.; Salamero, J.; Cherfils, J. *Journal of Biological Chemistry* **2004**, *279*, 11480–11488.

(271) Aytuna, A. S.; Gursoy, A.; Keskin, O. *Bioinformatics* **2005**, *21*, 2850–2855.

(272) Rajakulendran, T.; Sahmi, M.; Lefrancois, M.; Sicheri, F.; Therrien, M. *Nature* **2009**, *461*, 542–U114.

(273) Callaghan, J.; Simonsen, A.; Gaullier, J. M.; Toh, B. H.; Stenmark, H. *Biochemical Journal* **1999**, *338*, 539–543.

(274) Dumas, J. J.; Merithew, E.; Sudharshan, E.; Rajamani, D.; Hayes, S.; Lawe, D.; Corvera, S.; Lambright, D. G. *Molecular Cell* **2001**, *8*, 947–958.

(275) Murray, D. H.; Jahnel, M.; Lauer, J.; Avellaneda, M. J.; Brouilly, N.; Cezanne, A.; Morales-Navarrete, H.; Perini, E. D.; Ferguson, C.; Lupas, A. N.; Kalaidzidis, Y.; Parton, R. G.; Grill, S. W.; Zerial, M. *Nature* **2016**, *537*, 107–111.

(276) Kozakov, D.; Hall, D. R.; Xia, B.; Porter, K. A.; Padhorny, D.; Yueh, C.; Beglov, D.; Vajda, S. *Nature Protocols* **2017**, *12*, 255–278.

(277) De Vries, S. J.; van Dijk, M.; Bonvin, A. *Nature Protocols* **2010**, *5*, 883–897.

(278) Seabra, M. C.; Wasmeier, C. *Current Opinion in Cell Biology* **2004**, *16*, 451–457.

(279) Edler, E.; Stein, M. *Small GTPases* **2017**, 1–16.

(280) Schalk, I.; Zeng, K.; Wu, S. K.; Stura, E. A.; Matteson, J.; Huang, M. D.; Tandon, A.; Wilson, I. A.; Balch, W. E. *Nature* **1996**, *381*, 42–48.

(281) An, Y.; Shao, Y.; Alory, C.; Matteson, J.; Sakisaka, T.; Chen, W.; Gibbs, R. A.; Wilson, I. A.; Balch, W. E. *Structure* **2003**, *11*, 347–357.

(282) Pylypenko, O.; Rak, A.; Durek, T.; Kushnir, S.; Dursina, B. E.; Thomae, N. H.; Constantinescu, A. T.; Brunsveld, L.; Watzke, A.; Waldmann, H.; Goody, R. S.; Alexandrov, K. *Embo Journal* **2006**, *25*, 13–23.

(283) Ignatev, A.; Kravchenko, S.; Rak, A.; Goody, R. S.; Pylypenko, O. *Journal of Biological Chemistry* **2008**, *283*, 18377–18384.

(284) Waterhouse, A. M.; Procter, J. B.; Martin, D. M. A.; Clamp, M.; Barton, G. J. *Bioinformatics* **2009**, *25*, 1189–1191.

(285) Durrant, J. D.; de Oliveira, C. A. F.; McCammon, J. A. *Journal of Molecular Graphics & Modelling* **2011**, *29*, 773–776.

(286) Durrant, J. D.; Votapka, L.; Sorensen, J.; Amaro, R. E. *Journal of Chemical Theory and Computation* **2014**, *10*, 5047–5056.

(287) Rak, A.; Pylypenko, O.; Durek, T.; Watzke, A.; Kushnir, S.; Brunsveld, L.; Waldmann, H.; Goody, R. S.; Alexandrov, K. *Science* **2003**, *302*, 646–650.

(288) Luan, P.; Heine, A.; Zeng, K.; Moyer, B.; Greasely, S. E.; Kuhn, P.; Balch, W. E.; Wilson, I. A. *Traffic* **2000**, *1*, 270–281.

(289) Garcia-Mata, R.; Boulter, E.; Burridge, K. *Nature Reviews Molecular Cell Biology* **2011**, *12*, 493–504.

(290) Desnoyers, L.; Seabra, M. C. *Proceedings of the National Academy of Sciences of the United States of America* **1998**, *95*, 12266–12270.

(291) Guo, Z.; Wu, Y. W.; Das, D.; Delon, C.; Cramer, J.; Yu, S.; Thuns, S.; Lupilova, N.; Waldmann, H.; Brunsveld, L.; Goody, R. S.; Alexandrov, K.; Blankenfeldt, W. *Embo Journal* **2008**, *27*, 2444–2456.

(292) Rak, A.; Pylypenko, O.; Niculae, A.; Pyatkov, K.; Goody, R. S.; Alexandrov, K. *Cell* **2004**, *117*, 749–760.

(293) Gilbert, P. M.; Burd, C. G. *Journal of Biological Chemistry* **2001**, *276*, 8014–8020.

(294) Sivars, U.; Aivazian, D.; Pfeffer, S. R. *Nature* **2003**, *425*, 856–859.

(295) Schoebel, S.; Oesterlin, L. K.; Blankenfeldt, W.; Goody, R. S.; Itzen, A. *Molecular Cell* **2009**, *36*, 1060–1072.

(296) Oesterlin, L. K.; Goody, R. S.; Itzen, A. *Proceedings of the National Academy of Sciences of the United States of America* **2012**, *109*, 5621–5626.

(297) Haas, A. K.; Fuchs, E.; Kopajtich, R.; Barr, F. A. *Nature Cell Biology* **2005**, *7*, 887–U36.

(298) Merithew, E.; Stone, C.; Eathiraj, S.; Lambright, D. G. *Journal of Biological Chemistry* **2003**, *278*, 8494–8500.

(299) Bucci, C.; Parton, R. G.; Mather, I. H.; Stunnenberg, H.; Simons, K.; Hoflack, B.; Zerial, M. *Cell* **1992**, *70*, 715–728.

(300) Bull, S. C.; Doig, A. J. *PLoS One* **2015**, *10*, 44.

Appendices

A MD simulation parameter settings

Table A1: Parameter settings for full-atomistic simulations in NAMD (for details see NAMD user guide version 2.9).

NAMD parameter	Value	Description
		Minimization
minimization	on	Do conjugate gradient minimization of potential energy
		Heating phase
seed	1010	Seed the random number generator so that consecutive simulations produce the same results
numsteps	310000	Number of integration steps
temperature	0	Initial temperature for generating initial velocity distribution (in K)
reassignFreq	1	Steps between velocity reassignment
reassignIncr	0.001	ΔT by which temperature in gradually increased (in K)
reassignHold	310	Final target temperature after heating (in K)
		MD run
Time step parameters		
timestep	2	Time step of 2 fs
rigidBonds	all	Covalent bond lengths involving hydrogen atoms are restrained to value given in parameter file (SHAKE algorithm)
nonbondedFreq	1	Time steps between calculation of short-range non-bonded interactions
fullElectFrequency	2	Number of time steps between each full electrostatics evaluation
Non-bonded interactions		
cutoff	15	Distance where van der Waals interactions are truncated and electrostatics are directly calculated every time step
switching	on	Smoothing function is applied to both vdW and electrostatic forces
vdwForceSwitching	on	CHARMM force switching for vdW forces (used for CHARMM36 lipids to give correct membrane surface tension)
switchdist	10	Distance at which the switching function begins to have an effect
exclude	scaled1-4	Specifies which pairs of bonded atoms are excluded from non-bonded interactions
1-4scaling	1	Scale factor for (i, i+3) electrostatic interactions
pairlistdist	16.5	Maximum distance used for generating pair list, that contains pairs of atoms for which electrostatics and van der Waals interactions are calculated
stepspercycle	20	Time steps in one cycle before pair list is updated

PME	on	Use Particle Mesh Ewald summation for long-range electrostatics; to use with periodic boundary conditions
PMEGridSizeX/Y/Z	470	Number of grid point in x,y,z dimension; should be twofold the box size in x,y,z direction
Periodic boundary conditions (PBC)		
cellBasisVector1/2/3	155.0 0.0 0.0	Basis vector periodic boundary conditions (in Å)
cellOrigin	0.67 -0.58 28.72	Specifies center of the periodic cell (in Å)
wrapAll	on	Wrap all coordinates around periodic boundaries
Temperature and pressure control		
langevin	on	Use Langevin dynamics
langevinDamping	1	Langevin damping / coupling coefficient (in ps^{-1})
langevinTemp	310	Target temperature for Langevin dynamics (in K)
useGroupPressure	yes	Pressure is calculated using a hydrogen-group based pseudo-molecular virial and kinetic energy (needed for SHAKE algorithm)
useFlexibleCell	yes	Allows anisotropic cell fluctuations (the three orthogonal cell dimensions fluctuate independently)
useConstantArea	no	No constant area
useConstantRatio	yes	Keeps ratio of the unit cell in the x-y plane constant (semi-isotropic coupling)
langevinPiston	on	Langevin piston pressure control
langevinPistonTarget	1.01325	Target pressure in bar (1 atm)
langevinPistonPeriod	200	Specifies barostat oscillation time scale (in fs)
langevinPistonDecay	50	Specifies barostat damping time scale (in fs)
langevinPistonTemp	310	Specifies barostat noise temperature (should be equal to target temperature)

Table A2: Parameter settings for SMD, harmonic restraints and colvar-based calculations in NAMD.

NAMD parameter	Value	Description
SMD	on	Apply SMD harmonic constraint to system
SMDFile	pulling.ref	PDB file containing initial reference position for constraints; all atoms with nonzero value in occupancy column will be tagged as SMD atoms
SMDk	2.4	Harmonic constraint force constant (in $\mathrm{kcal\,mol^{-1}\,Å^{-2}}$); corresponds to $\approx$1000 $\mathrm{kJ\,mol^{-1}\,nm^{-2}}$
SMDVel	0.00001	Velocity of reference position movement (in Å/time step); corresponds to 5 $\mathrm{Å\,ns^{-1}}$
SMDDir	0 0 1	Direction of the SMD center of mass movement; here, +z direction
SMDOutputFreq	20	Frequency with which SMD data output is generated (in time steps)
Harmonic restraint parameters		
constraints	on	Calculates harmonic constraints
consexp	2	Exponent for harmonic constraint energy function
consref	constraints.ref	PDB file containing reference positions for constraints
conskfile	constraints.ref	PDB file containing force constants for harmonic constraints
conskcol	O	Column in constraints file that contains force constant value; here, the occupancy column

selectConstraints	on	Restrains only selected Cartesian components of coordinates; atoms are allowed to move in the other directions
selectConstrX	off	No restriction in x direction
selectConstrY	off	No restriction in y direction
selectConstrZ	on	Restrain z components of coordinates
Collective variables module		
colvars	on	Activates the collective variables module
colvarsConfig	US-win0.in	TCL configuration file containing the definition of all collective variables

Table A3: Non-default parameter settings for a Martini coarse-grained MD production run in GROMACS (for details see GROMACS user manual version 5.0.7).

GROMACS parameter	Value	Description
	MD production run	
Integrator		
integrator	md	Default MD integrator of Newtons equations of motion is leap-frog algorithm
dt	0.02	Time step for integration (in ps)
nsteps	1000000	Maximum number of steps to integrate/ minimize
Output control		
nstxout	0	Number of steps between writing coordinates to output trajectory
nstvout	0	Number of steps between writing velocities to output trajectory
nstfout	0	Number of steps between writing forces to output trajectory
nstlog	1000	Number of steps between writing energies to log file
nstxout-compressed	1000	Number of steps between writing position coordinates using lossy compression
compressed-x-precision	1000	Precision of compressed trajectory file
Non-bonded interactions and neighbor searching parameters		
Cutoff-scheme	Verlet	Generate pair list with buffering
nstlist	20	Frequency to update neighbor list (in steps)
ns_type	grid	Make grid and check only atoms in neighboring grid cells when constructing new neighbor list
pbc	xyz	Use periodic boundary conditions in all directions
verlet-buffer-tolerance	0.005	Maximum allowed error for pair interactions per particle caused by Verlet buffer (in $kJ\,mol^{-1}\,ps^{-1}$)
rlist	1.4	Cut-off distance for short-range neighbor list (in nm)
coulombtype	Reaction-field	Reaction field with Coulomb cut-off $r_{coulomb}$
rcoulomb	1.1	Distance for Coulomb cut-off (in ns)
epsilon_r	15	Relative dielectric constant
vdw-type	shift	Lennard-Jones potential is decreased over whole range and forces decay smoothly to zero
rvdw-switch	0.9	Distance where to start switching Lennard-Jones force (in nm)
rvdw	1.1	Distance for Lennard-Jones cut-off (in nm)
Temperature and pressure control		
t-coupl	v-rescale	Temperature coupling using velocity rescaling with a stochastic term (required for Martini)

tc-grps	Water Membrane Protein	Groups to couple separately to temperature bath
tau-t	1.0 1.0 1.0	Time constant for coupling (in ps)
ref-t	310 310 310	Reference temperature for coupling (in K)
Pcoupl	Parrinello-Rahman	Extended-ensemble pressure coupling with the box vectors are subject to an equation of motion
Pcoupltype	semiisotropic	Isotropic pressure coupling in x,y-directions; required for membrane simulations
tau-p	12.0 12.0	Time constant for coupling (in ps)
compressibility	3e-4 3e-4	Compressibility (in bar^{-1})
ref-p	1.0 1.0	Reference pressure for coupling (in bar)
refcoord-scaling	all	Reference coordinates are scaled with the scaling matrix of the pressure coupling
Bond constraints		
constraints	none	No additional constraints for water (only those in topology)
constraint-algorithm	LINCS	Linear Constraint Solver: resets bonds to correct lengths after an unconstrained update
continuation	no	Apply constraints to the start configuration and reset shells
lincs_order	8	Highest order in the expansion of the constraint coupling matrix (8 required for Martini)
lincs_iter	2	Number of iterations to correct for rotational lengthening in LINCS (2 required for Martini)
lincs_warnangle	30	Maximum angle for a bond to rotate before LINCS will complain

B Additional MD simulation results

B.1 Geranylgeranyl chain parameters for coarse-grained MD

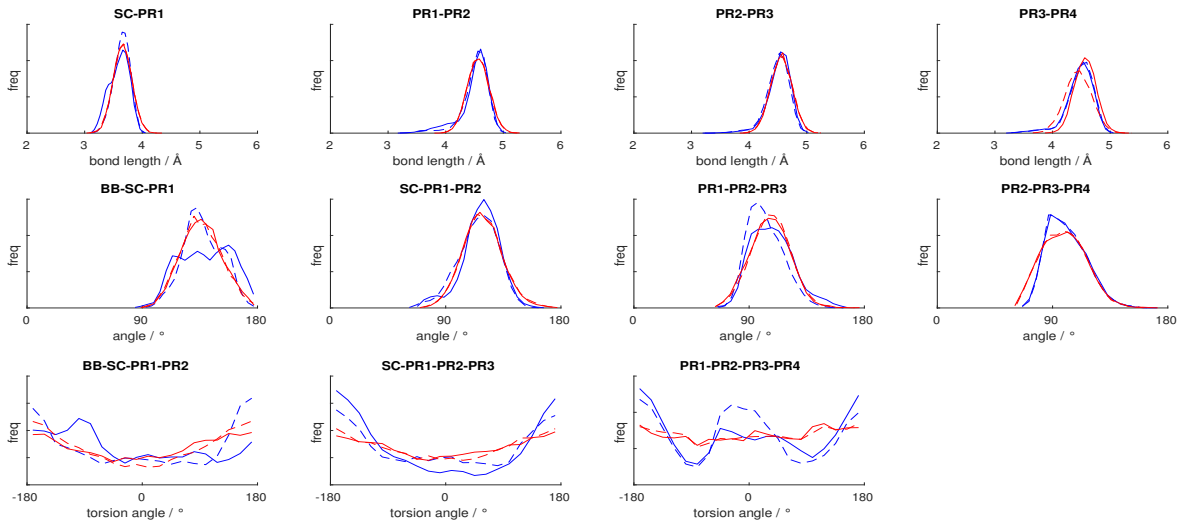


Figure B1: The bond lengths, bond angles and torsion angles of coarse-grained geranylgeranyl chains (red) were compared to full-atomistic geranylgeranyl anchor simulations (blue). Both modified cysteine residues are shown (GG-Cys^{212}: solid, GG-Cys^{213}: dashed).

B.2 Properties of the membrane models for different protein systems

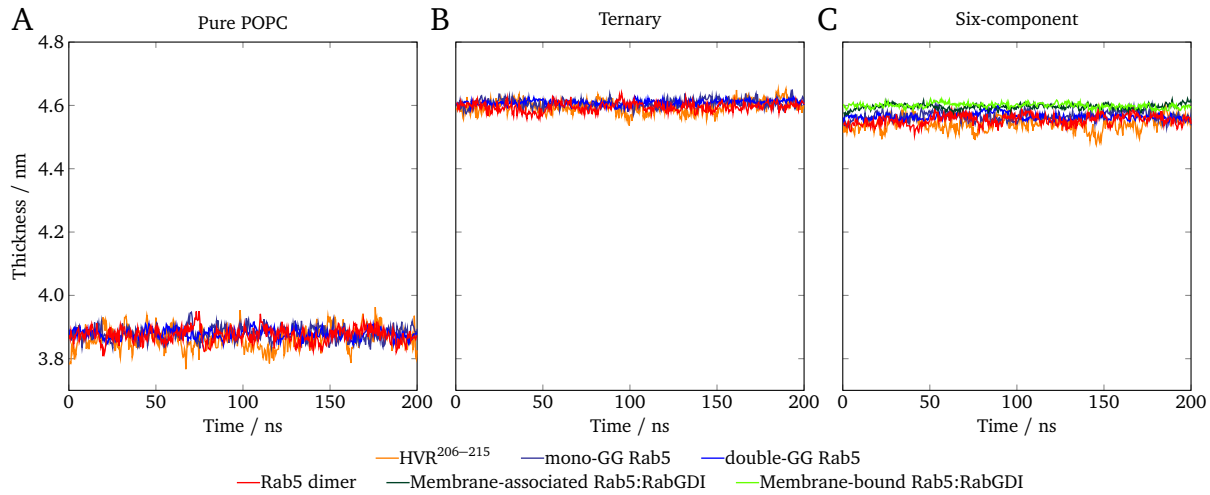


Figure B2: Global thickness of the three investigated model membranes as a function of time; A) pure POPC, B) the ternary mixture and C) the complex six-component membrane. Averaged data are shown for the different peptide or protein systems.

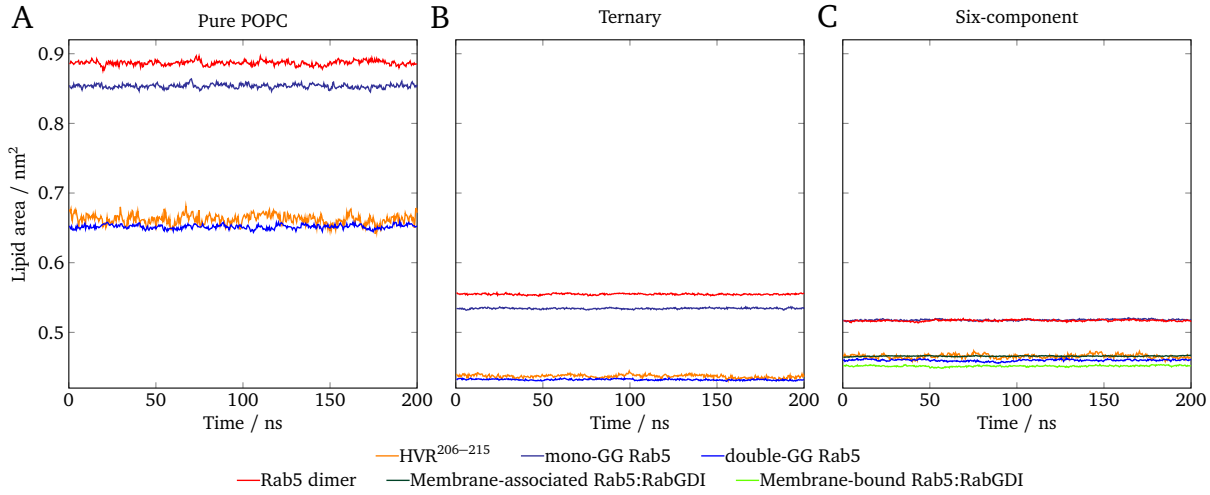


Figure B3: Global area per lipid during 200 ns full-atomistic MD simulation of A) the pure POPC membrane, B) the ternary mixture and C) the six-component membrane. Data were derived from different peptide / protein systems.

Table B1: Global thickness, lipid area and average lipid order parameter of the three model membranes when simulated with different peptide / protein systems. Average and deviation are calculated from the replicates and Rab5 GDP/GTP activation states where possible.

		Pure POPC	**Ternary**	**Six-component**
Membrane thickness / nm	$HVR^{206-215}$	3.86	4.60	4.54
	Mono-GG Rab5	3.88 ± 0.02	4.61 ± 0.01	4.57 ± 0.01
	Double-GG Rab5	3.88 ± 0.03	4.61 ± 0.02	4.56 ± 0.02
	Rab5 dimer	3.87 ± 0.02	4.60 ± 0.01	4.55 ± 0.01
	Membrane-associated Rab5:GDI			4.60 ± 0.01
	Membrane-bound Rab5:GDI			4.60 ± 0.01
Lipid area / nm^2	$HVR^{206-215}$	0.662	0.437	0.466
	Mono-GG Rab5	0.854 ± 0.006	0.534 ± 0.004	0.518 ± 0.005
	Double-GG Rab5	0.652 ± 0.006	0.432 ± 0.002	0.460 ± 0.002
	Rab5 dimer	0.887 ± 0.044	0.555 ± 0.027	0.517 ± 0.017
	Membrane-associated Rab5:GDI			0.466 ± 0.010
	Membrane-bound Rab5:GDI			0.452 ± 0.001
Lipid order parameter	$HVR^{206-215}$	0.176	0.327	0.305
	Mono-GG Rab5	0.174 ± 0.000	0.326 ± 0.002	0.309 ± 0.000
	Double-GG Rab5	0.176 ± 0.001	0.327 ± 0.001	0.309 ± 0.000
	Rab5 dimer	0.176 ± 0.001	0.325 ± 0.002	0.308 ± 0.002
	Membrane-associated Rab5:GDI			0.317 ± 0.002
	Membrane-bound Rab5:GDI			0.317 ± 0.001

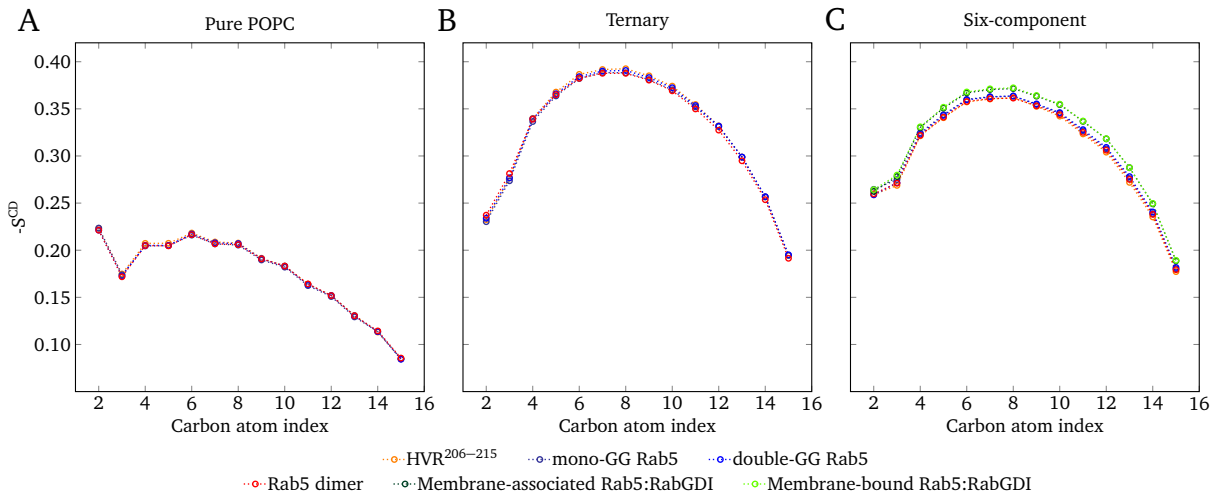


Figure B4: Acyl chain order parameters for the saturated *sn-1* palmitoyl chains of POPC, PSM, POPE and POPS in A) pure POPC, B) the ternary mixture and C) the six-component membrane. Data are averaged over all lipids within each protein system.

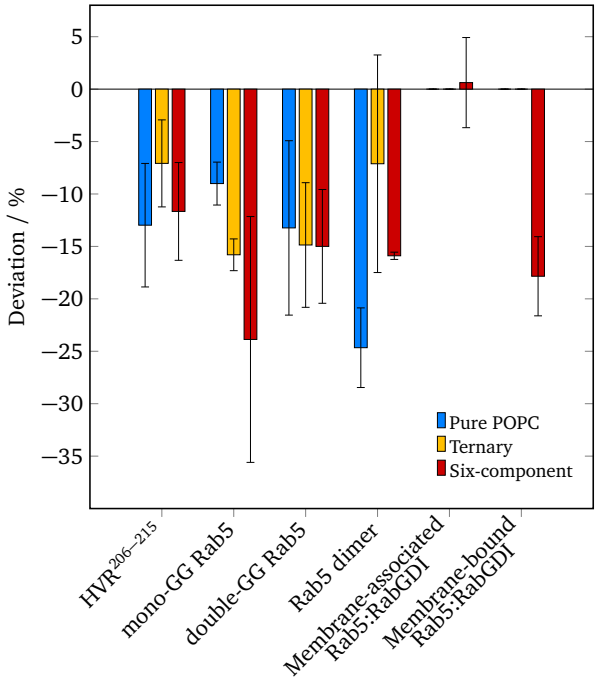


Figure B5: Local percentage variance of the average acyl chain order parameter around the GG anchor (cutoff radius = 0.5 nm) from the global average in the three investigated membrane models for different protein systems.

Table B2: Lateral lipid diffusion coefficients from full-atomistic MD simulations of the different peptide and protein systems in pure POPC, the ternary and six-component membranes.

Diffusion coefficient / 10^{-7} cm^2 s^{-1}		**POPC**	**Cholesterol**	**PSM**	**POPE**	**POPS**	**PI(3)P**
Pure POPC	$HVR^{206-215}$	1.36					
	Mono-GG Rab5	1.63 ± 0.08					
	Double-GG Rab5	1.44 ± 0.03					
	Rab5 dimer	1.73 ± 0.15					
Ternary	$HVR^{206-215}$	0.78	0.59	0.60			
	Mono-GG Rab5	0.83 ± 0.03	0.65 ± 0.05	0.72 ± 0.04			
	Double-GG Rab5	0.75 ± 0.02	0.55 ± 0.02	0.64 ± 0.02			
	Rab5 dimer	0.73 ± 0.00	0.57 ± 0.04	0.69 ± 0.02			
Six-component	$HVR^{206-215}$	0.74	0.59	0.62	0.68	0.65	0.62
	Mono-GG Rab5	0.79 ± 0.01	0.62 ± 0.02	0.66 ± 0.02	0.73 ± 0.01	0.70 ± 0.03	0.67 ± 0.03
	Double-GG Rab5	0.74 ± 0.01	0.58 ± 0.01	0.61 ± 0.01	0.67 ± 0.01	0.64 ± 0.01	0.60 ± 0.01
	Rab5 dimer	0.80 ± 0.02	0.63 ± 0.01	0.68 ± 0.01	0.73 ± 0.02	0.71 ± 0.02	0.69 ± 0.02
	Membrane-associated Rab5:GDI	0.63 ± 0.01	0.47 ± 0.01	0.54 ± 0.01	0.59 ± 0.01	0.57 ± 0.01	0.54 ± 0.01
	Membrane-bound Rab5:GDI	0.62 ± 0.00	0.47 ± 0.01	0.54 ± 0.01	0.58 ± 0.00	0.56 ± 0.00	0.53 ± 0.00

B.3 Protein properties in the different membrane models

Table B3: Lateral diffusion coefficients of the GG anchor in different membranes and peptide/ protein systems.

Diffusion coefficient / 10^{-7} cm^2 s^{-1}	HVR$^{206-215}$	Mono-GG Rab5	Double-GG Rab5	Rab5 dimer	Membrane-bound Rab5:GDI
Pure POPC	0.78 ± 0.13	1.59 ± 0.33	0.93 ± 0.09	0.79 ± 0.11	
Ternary	0.47 ± 0.10	1.31 ± 0.30	0.51 ± 0.06	0.64 ± 0.09	
Six-component	0.43 ± 0.08	1.49 ± 0.08	0.43 ± 0.05	0.56 ± 0.02	0.33 ± 0.05

B.4 Rab5 truncated HVR results in different model membranes

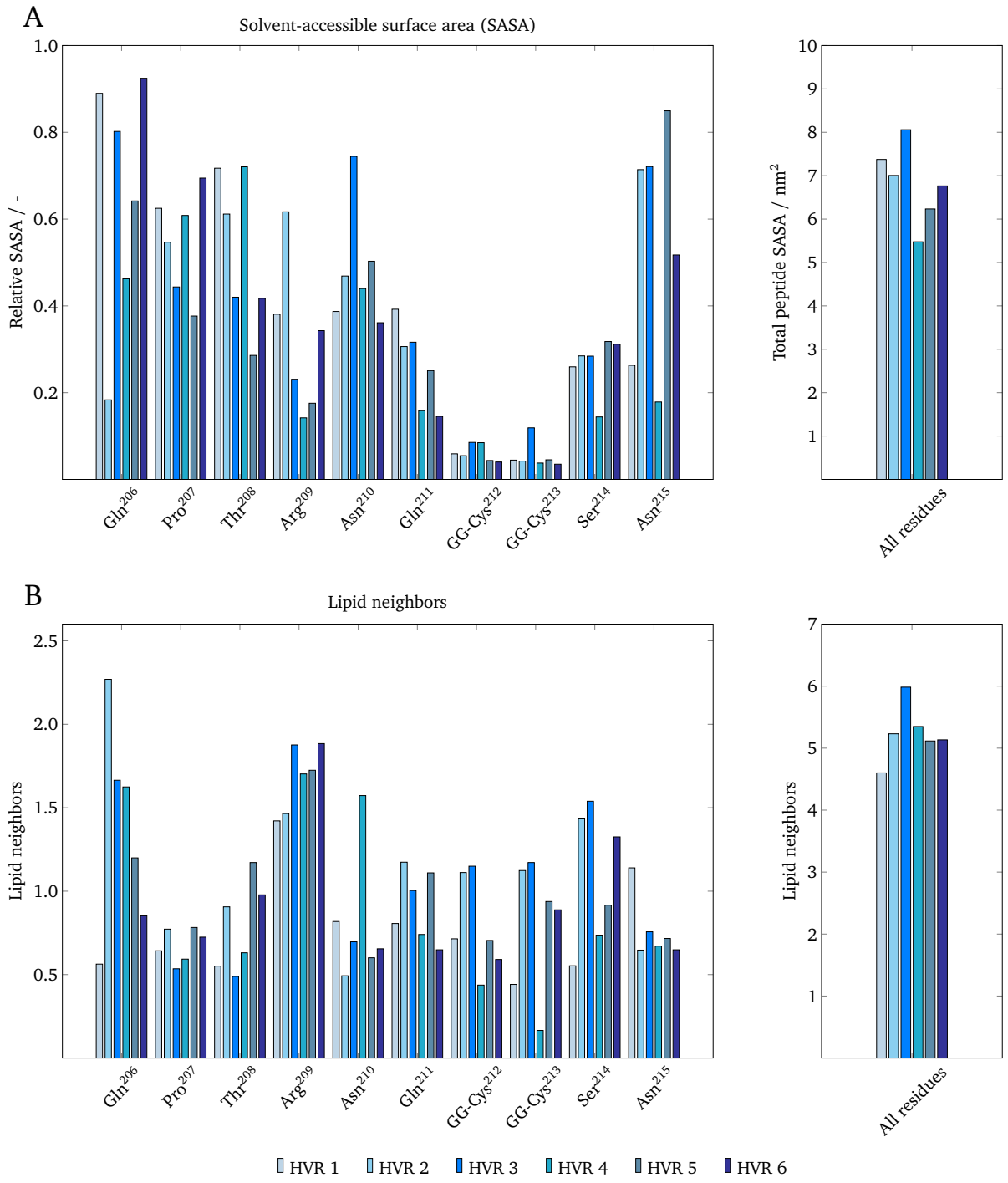


Figure B6: HVR$^{206–215}$ structural conformations in POPC were explored regarding their A) SASA and B) the number of lipids within a radius of 0.5 nm.

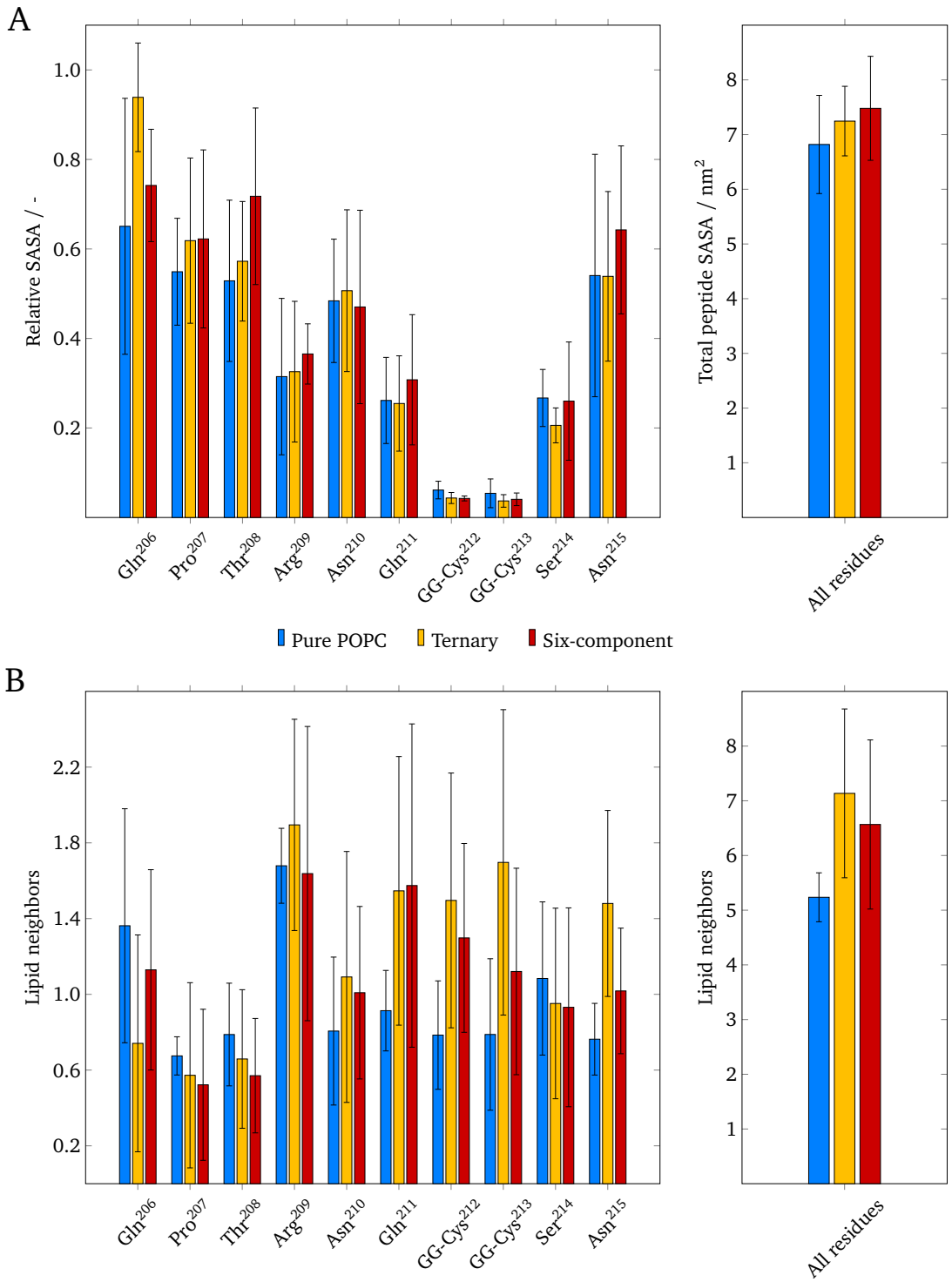


Figure B7: HVR$^{206–215}$ structural conformations were analyzed in terms of their A) SASA and B) the number of lipids within a radius of 0.5 nm. Data were averaged over 200 ns and the six HVR$^{206–215}$ replicates in pure POPC (blue), the ternary mixture (yellow) and the six-component membrane (red). The standard deviation between the replicates is represented by error bars.

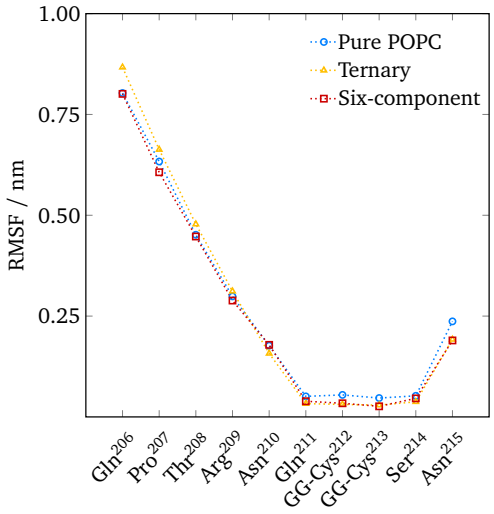


Figure B8: RMSF of the HVR backbone Cα atoms over 200 ns in pure POPC (blue), the ternary mixture (yellow) and the six-component membrane (red).

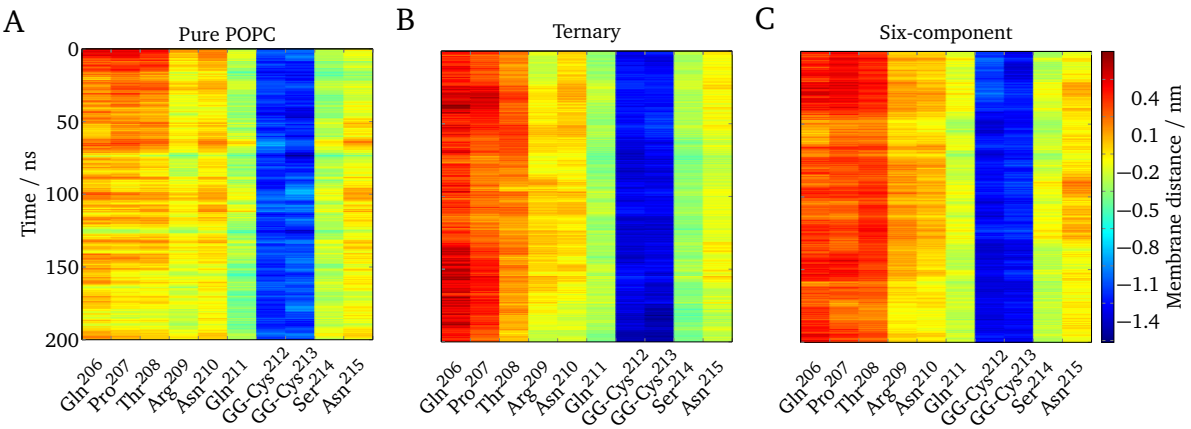


Figure B9: Residue-based membrane distance averaged over six $HVR^{206-215}$ replicates in A) pure POPC, B) the ternary mixture, and C) the six-component membrane.

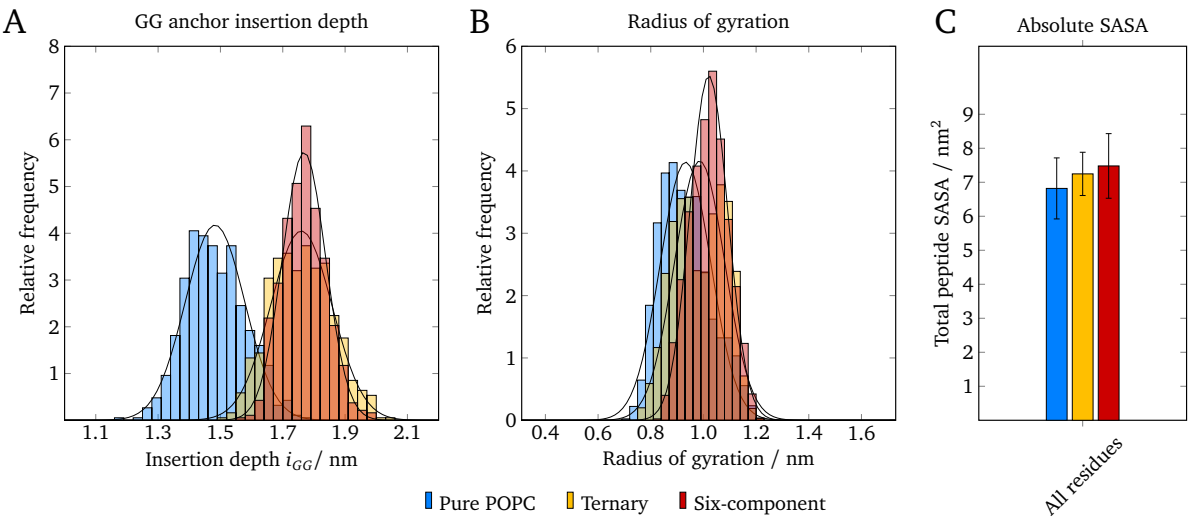


Figure B10: Peptide properties A) GG anchor insertion depth, B) gyration radius, and C) absolute SASA compared for the three investigated membrane systems.

Table B4: Properties of the individual $HVR^{206-215}$ replicates in the three model membranes. Local properties and the lipid composition were determined within a cutoff distance of 0.5 nm.

Property	HVR 1	HVR 2	HVR 3	HVR 4	HVR 5	HVR 6
	Pure POPC					
GG insertion depth / nm	1.438	1.343	1.444	1.579	1.393	1.699
Radius of gyration / nm	0.959	0.895	0.913	0.890	0.864	1.072
Diffusion coefficient / 10^{-7} cm^2 s^{-1}	0.69	1.02	0.83	0.80	0.62	0.71
Local thickness / nm	3.846	3.808	3.842	3.892	3.879	3.864
Local order parameter S^{CD}	0.160	0.143	0.161	0.135	0.159	0.161
	Ternary					
Surrounding lipids						
POPC / %	41.1	25.4	35.0	44.2	35.8	35.4
Cholesterol / %	45.6	52.7	46.0	34.6	49.1	39.3
PSM / %	13.3	21.9	19.0	21.2	15.1	25.3
GG insertion depth / nm	1.746	2.130	1.610	1.818	1.448	1.803
Radius of gyration / nm	0.922	1.085	0.854	1.021	0.956	1.075
Diffusion coefficient / 10^{-7} cm^2 s^{-1}	0.35	0.43	0.43	0.66	0.49	0.48
Local thickness / nm	4.518	4.487	4.522	4.489	4.585	4.525
Local order parameter S^{CD}	0.314	0.316	0.321	0.286	0.292	0.293
	Six-component					
Surrounding lipids						
POPC / %	28.4	7.0	15.8	16.1	4.4	16.9
Cholesterol / %	27.4	35.0	30.6	21.8	33.5	21.0
PSM / %	7.3	9.1	18.3	14.8	12.3	10.6
POPE / %	17.7	30.3	22.0	29.1	29.9	29.7
POPS / %	7.2	13.4	11.1	10.8	14.0	12.4
PI(3)P / %	12.0	5.2	2.2	7.4	5.9	9.4
GG insertion depth / nm	1.861	1.682	1.941	1.617	1.744	1.765
Radius of gyration / nm	1.038	0.981	1.059	0.940	1.013	1.09
Diffusion coefficient / 10^{-7} cm^2 s^{-1}	0.36	0.57	0.39	0.45	0.35	0.49
Local thickness / nm	4.525	4.480	4.465	4.454	4.488	4.500
Local order parameter S^{CD}	0.297	0.281	0.262	0.261	0.256	0.263

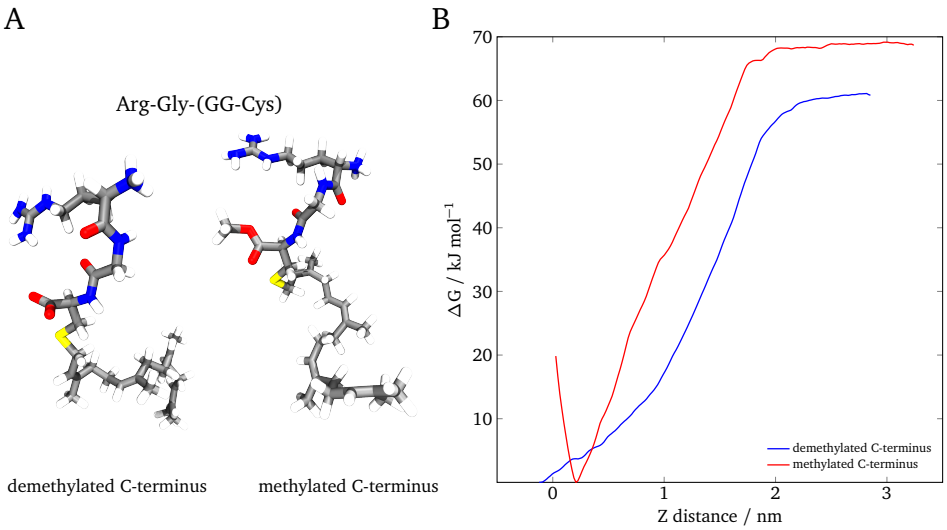


Figure B11: A) The methylated and demethylated tripeptide studied by Silvius *et al.*[216] was extracted from pure POPC in order to verify the umbrella sampling results. B) Gibbs free energy change upon membrane extraction.

B.5 Full-length Rab5 protein results in different model membranes

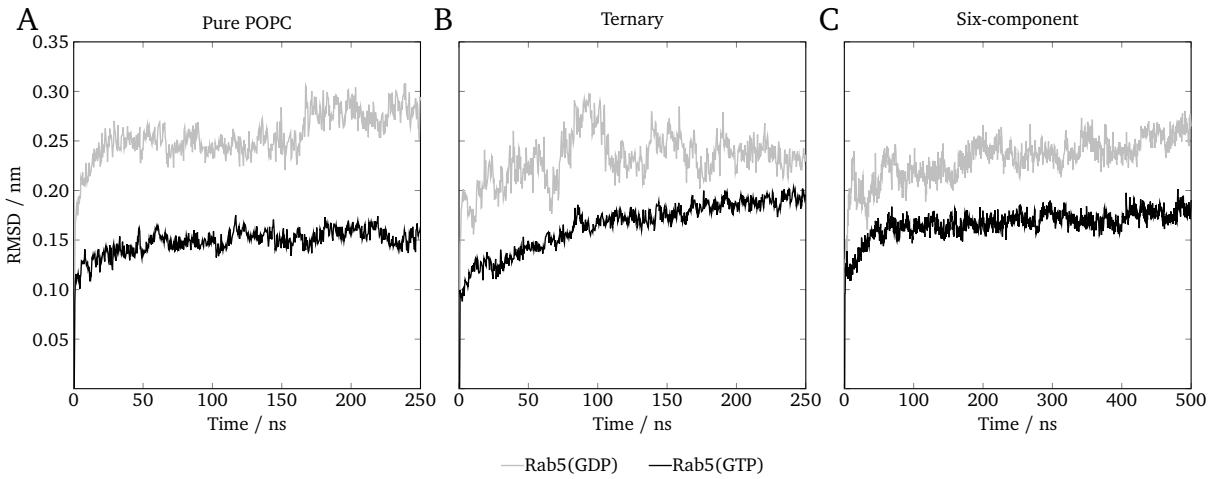


Figure B12: RMSD of the Rab5 G domain Cα atoms from the initial configuration over the complete MD simulation in A) pure POPC, B) the ternary mixture, and C) in the six-component membrane. Data are averaged over three independent simulations for Rab5(GDP) or Rab5(GTP) in each membrane system, respectively.

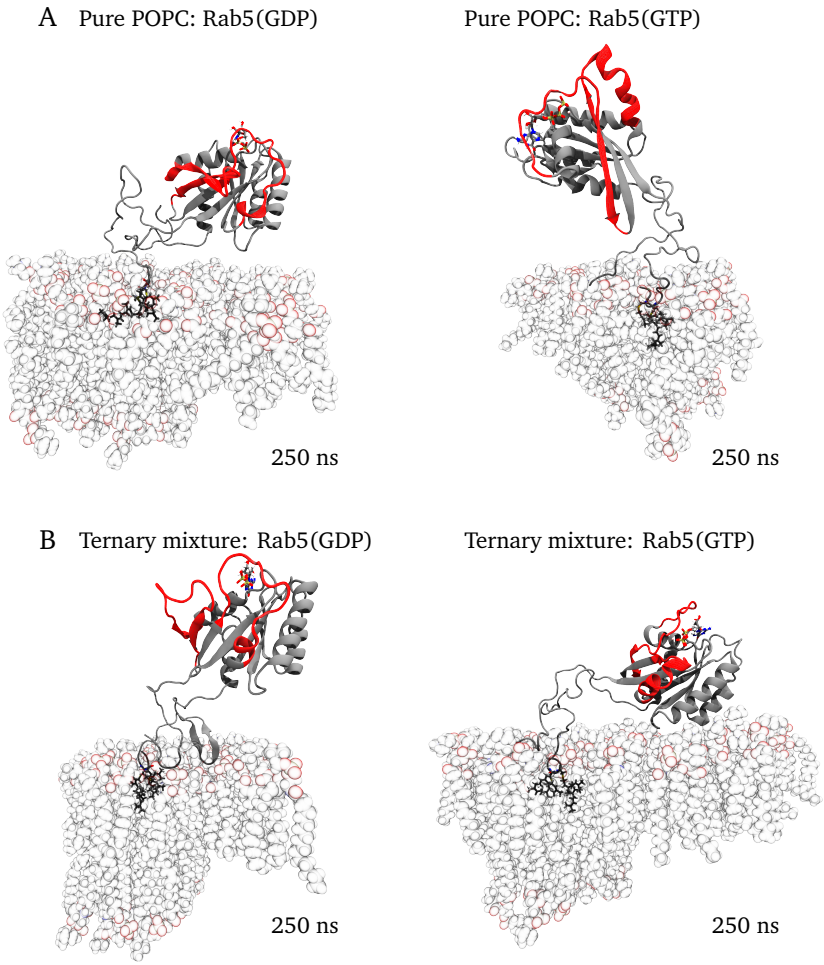


Figure B13: Final orientations of membrane-bound Rab5(GDP) (left) and Rab5(GTP) (right) in A) pure POPC and B) the ternary mixture. Regardless of the activation state, the switch regions (red) are completely solvent-exposed.

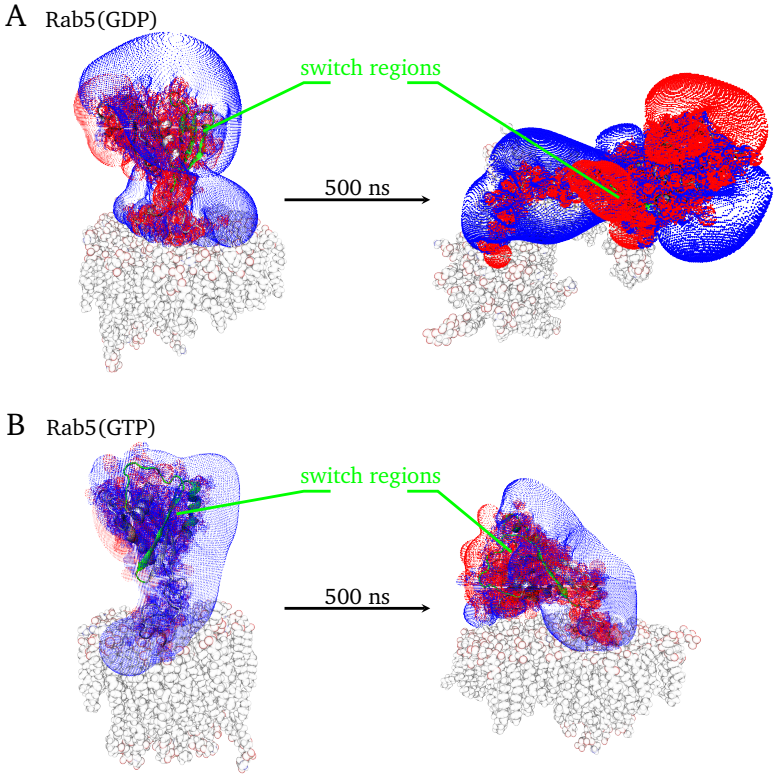


Figure B14: Electrostatic isocontour surfaces of full-length A) Rab5(GDP) and B) Rab5(GTP) bound to the six-component membrane at $\pm 1\, k_BT/e_c$. Representations are given for the initial electrostatic surfaces (left) and the final configurations when the protein was tilted toward the bilayer surface (right).

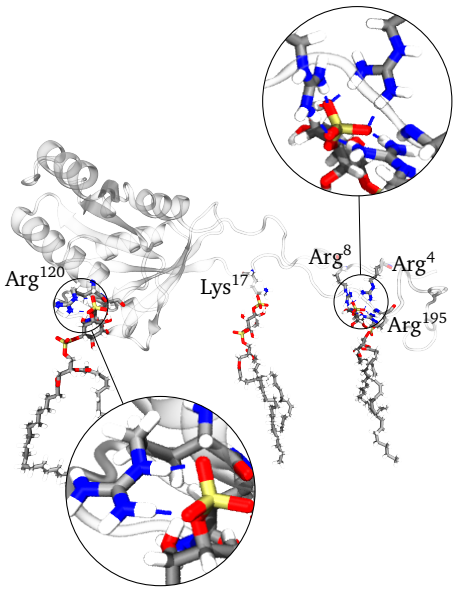


Figure B15: Hydrogen bonds (blue lines) formed between Rab5(GDP) basic residues and the PI(3)P head group phosphate in the six-component membrane.

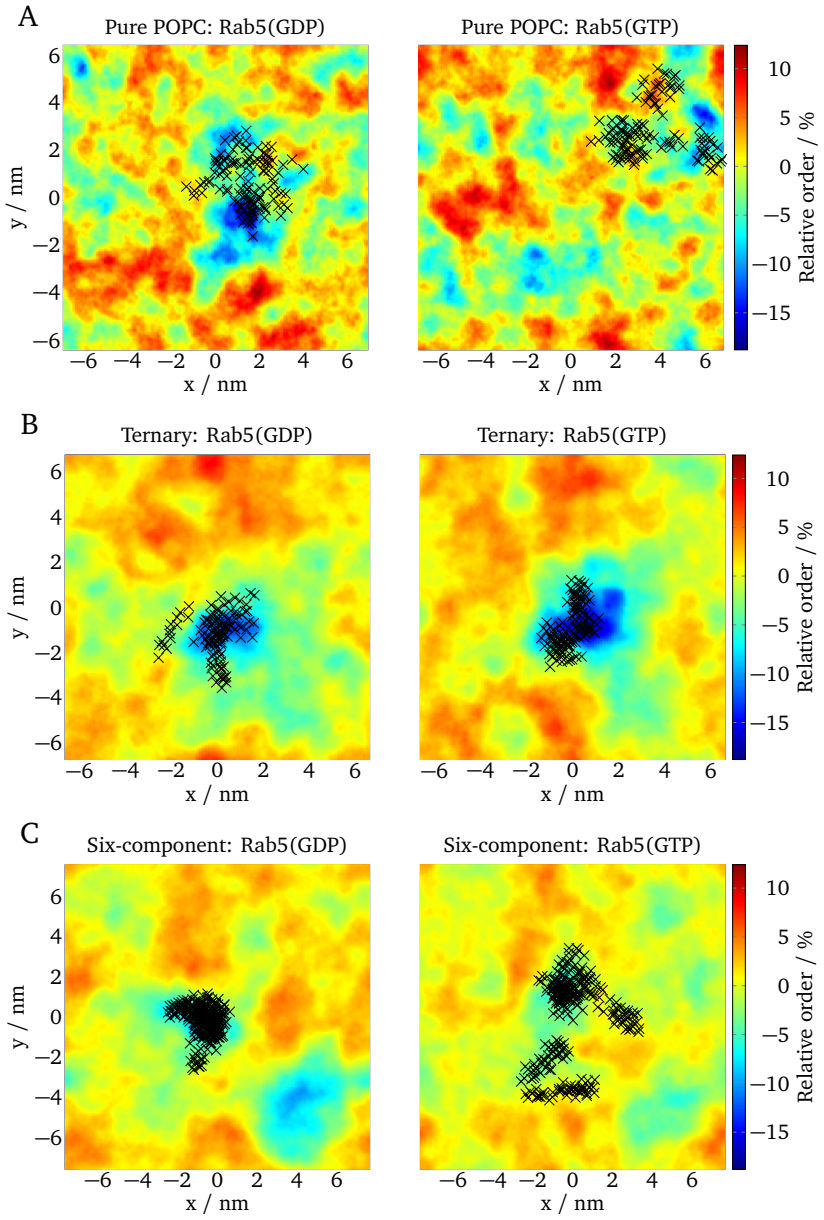


Figure B16: Local lipid order deviations in close proximity to full-length Rab5(GDP) or Rab5(GTP), respectively, in A) pure POPC, B) the ternary mixture, and C) the six-component membrane. Black crosses indicate the GG anchors positions; data are averaged over three independent MD simulations for each membrane system.

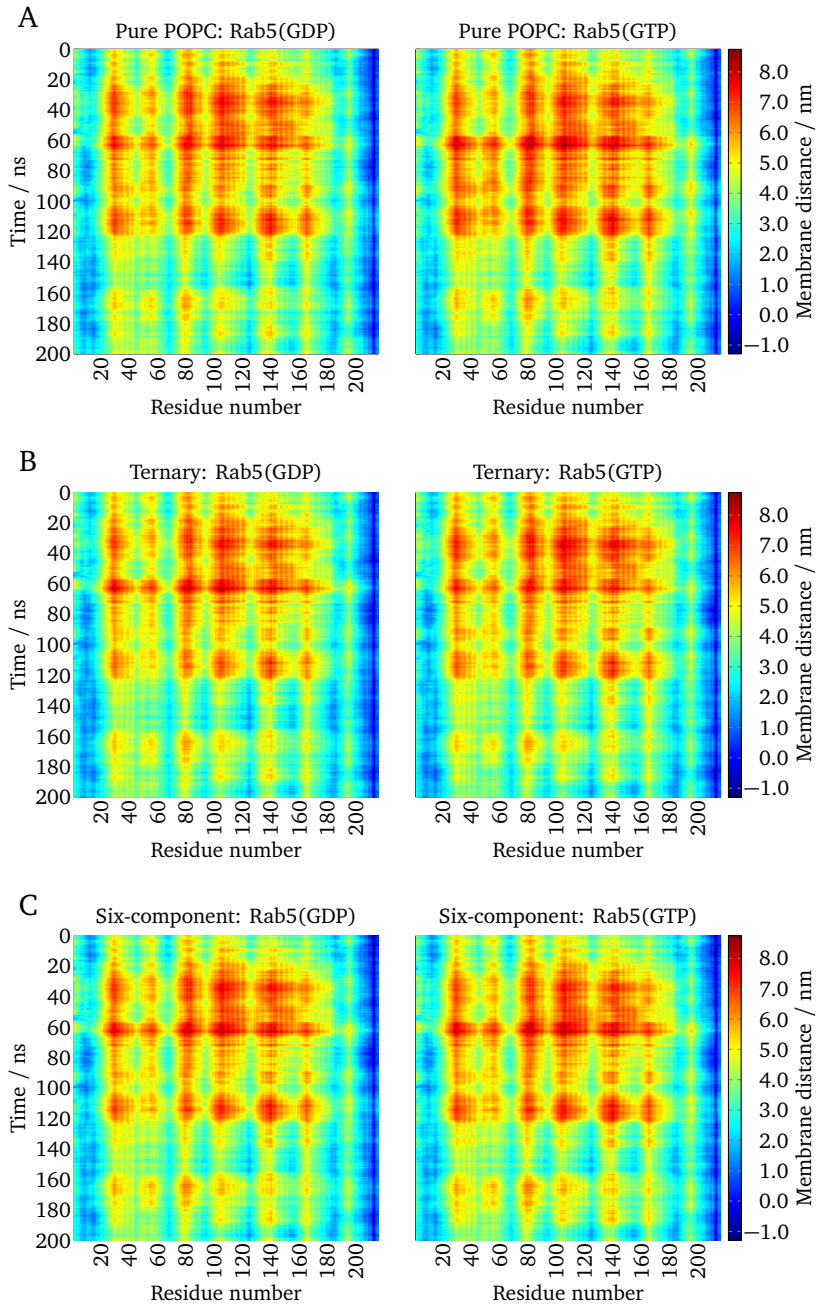


Figure B17: The z distances between each amino acid residue and the membrane surface were monitored over 200 ns in A) pure POPC and B) the ternary mixture and C) the six-component membrane. Data are shown for mono-GG Rab5(GDP) (left) and Rab5(GTP) (right).

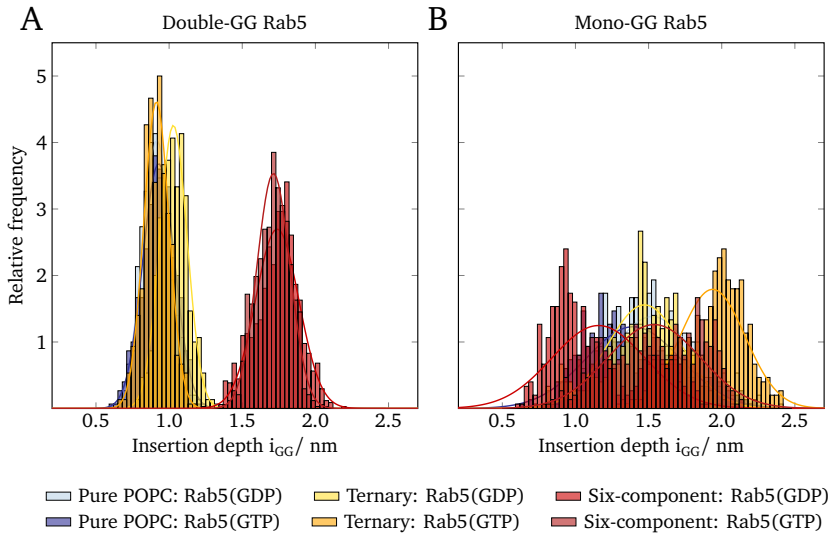


Figure B18: Distribution of the GG anchor insertion depths of A) double-GG Rab5 and B) mono-GG Rab5 in different model membranes.

B.6 Additional results for dimeric Rab5(GDP) and Rab5(GTP)

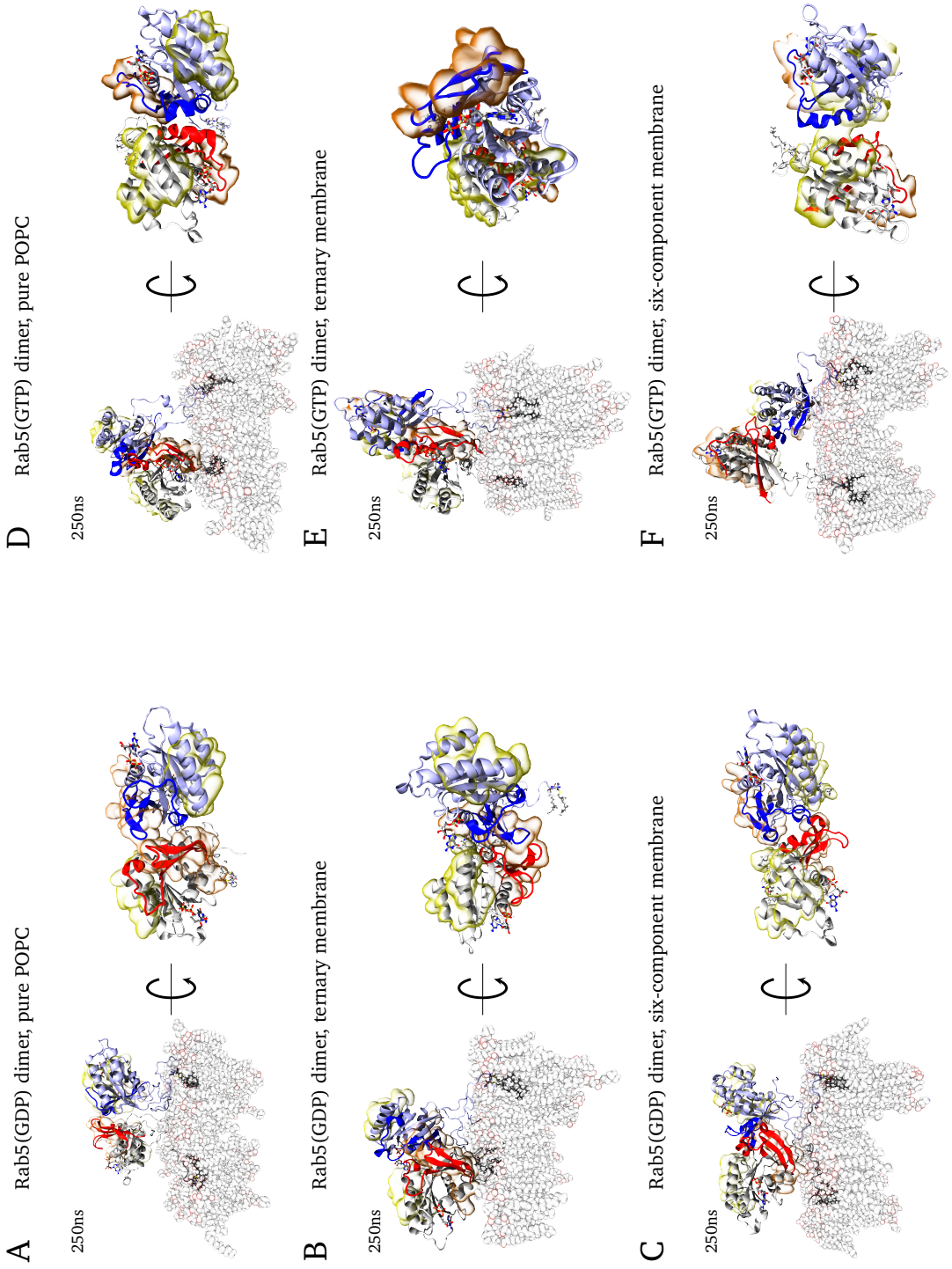


Figure B19: Final conformations of two associated Rab5 monomers from side and top views after 250 ns of MD simulation. A-C) Homotypic Rab5(GDP) and (D-E) homotypic Rab5(GTP) dimers in different membrane models. The helical and β-sheet dimer interfaces predicted from K-Ras are highlighted in yellow and orange, respectively.

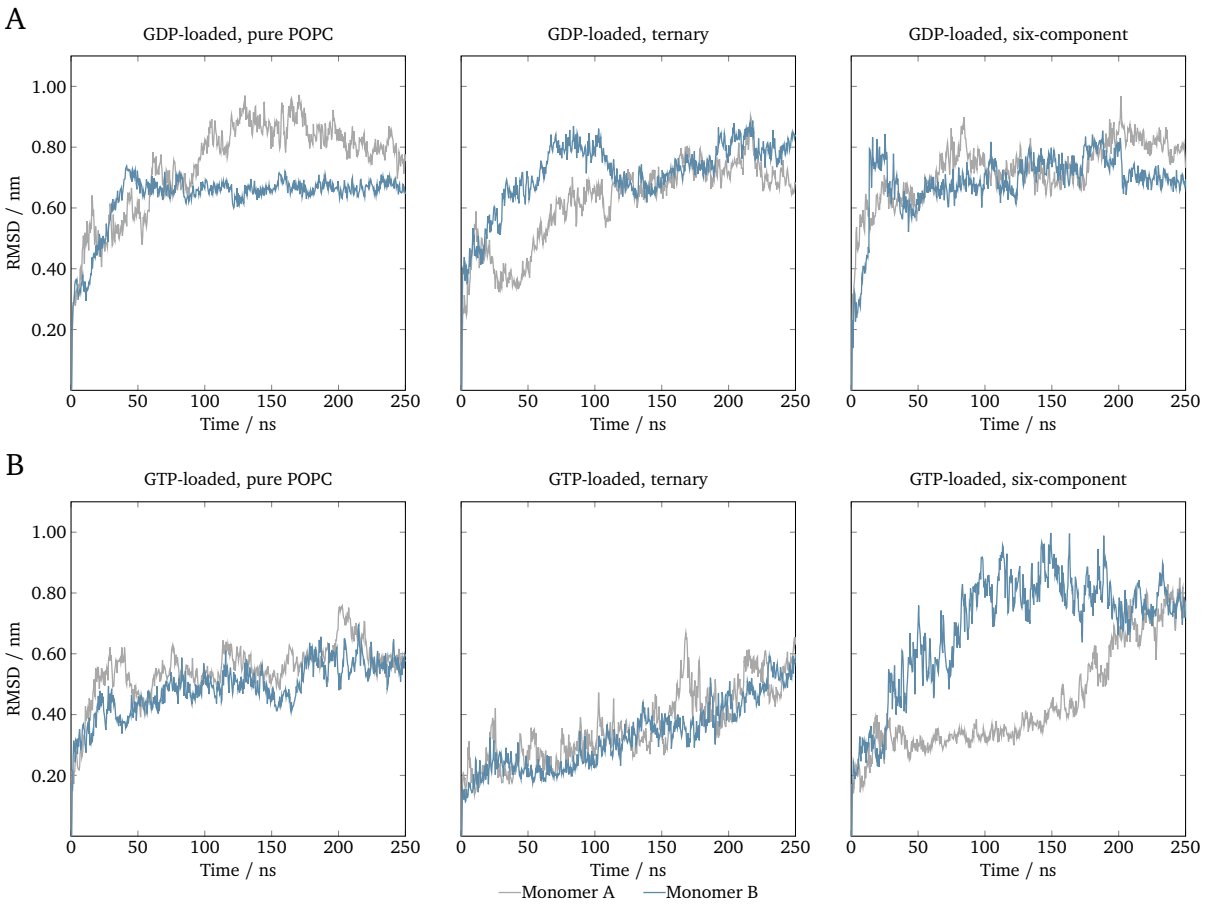


Figure B20: The RMSD of Rab5 monomers in A) the GDP-loaded and B) the GTP-loaded states over 250 ns MD simulation in different membrane systems (pure POPC: left, ternary mixture: middle, six-component membrane: right).

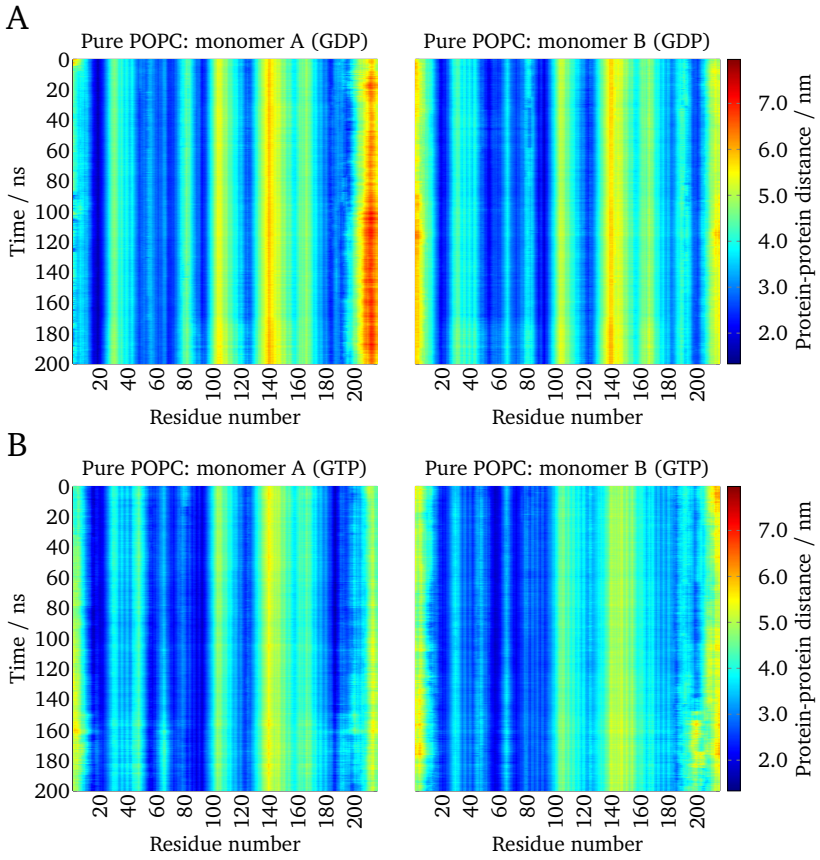


Figure B21: Protein-protein distances in the dimeric A) Rab5(GDP) and B) Rab5(GTP) complexes in pure POPC revealed the minimum intermolecular distance for each amino acid of the respective monomer to the other monomer center.

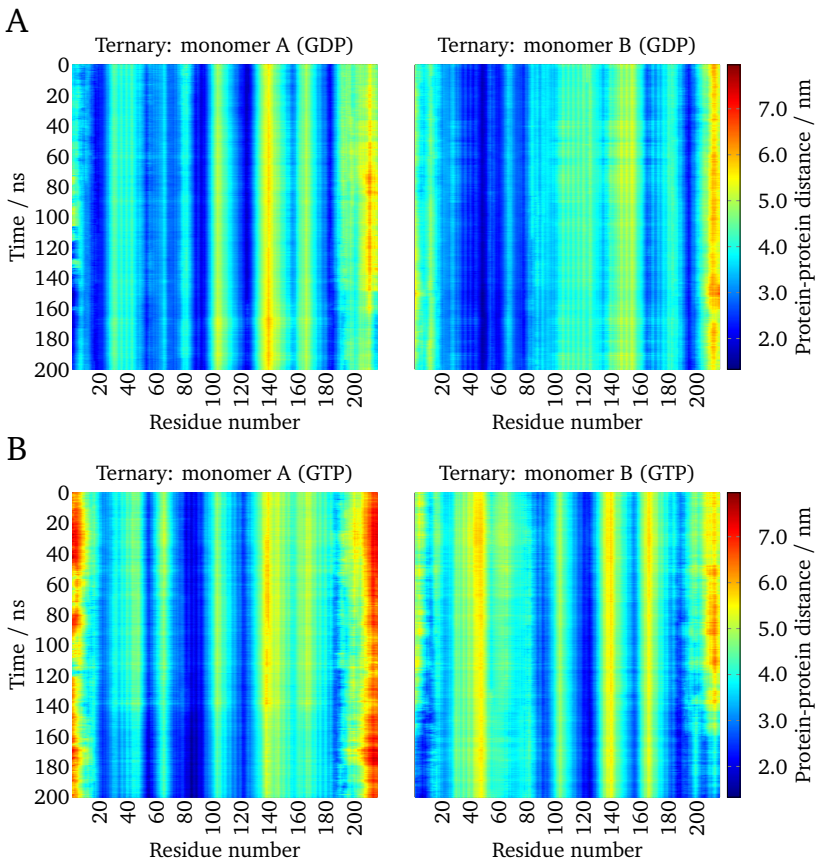


Figure B22: Amino acid-wise protein-protein distance maps of dimeric A) Rab5(GDP) and B) Rab5(GTP) in the ternary mixture. Data were calculated as the distance of each amino acid of one monomer to the other monomer center.

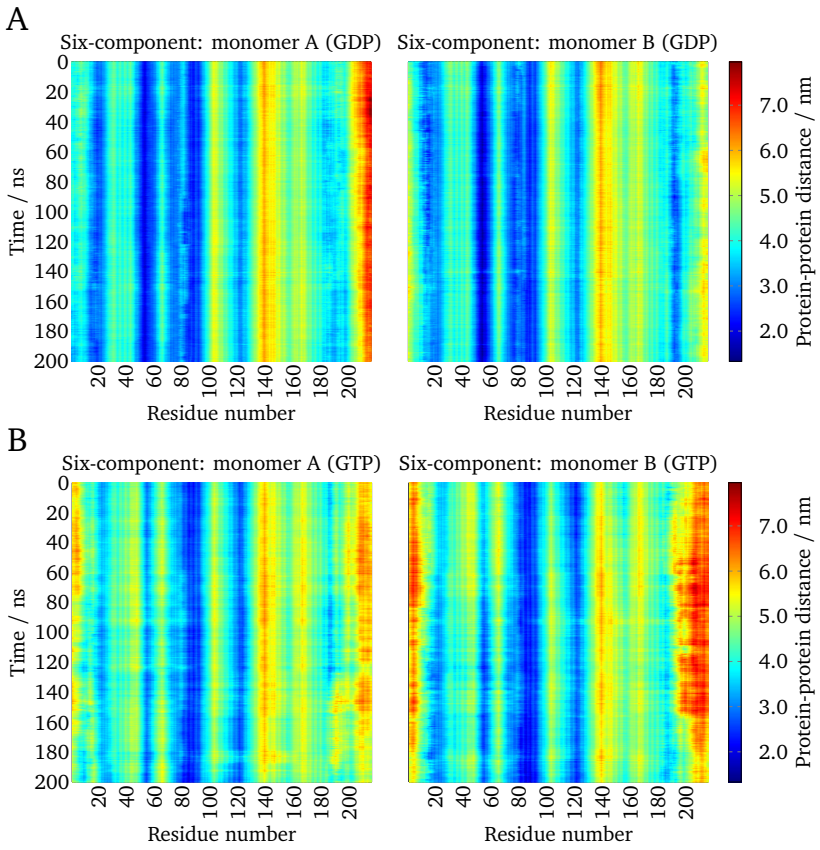


Figure B23: Protein-protein distances within the dimeric A) Rab5(GDP) and B) Rab5(GTP) complexes in the six-component membrane revealing the minimum intermolecular distance for each amino acid of the respective monomer to the other monomer center.

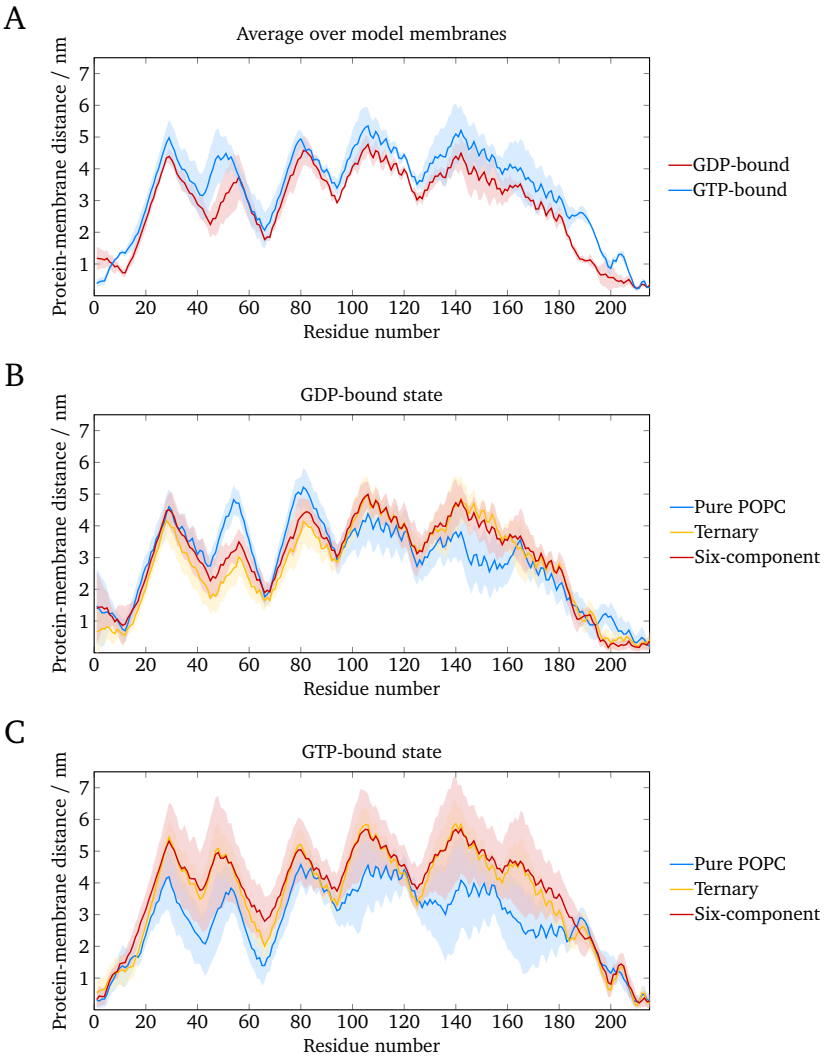


Figure B24: Dimeric Rab5 residue-wise protein-membrane distances averaged over A) the different model membranes as well as averaged over the individual monomers of a B) GDP-bound and C) GTP-bound dimer. The shadowed area corresponds to the standard deviation between the monomers of one dimer and in A) the different bilayer models.

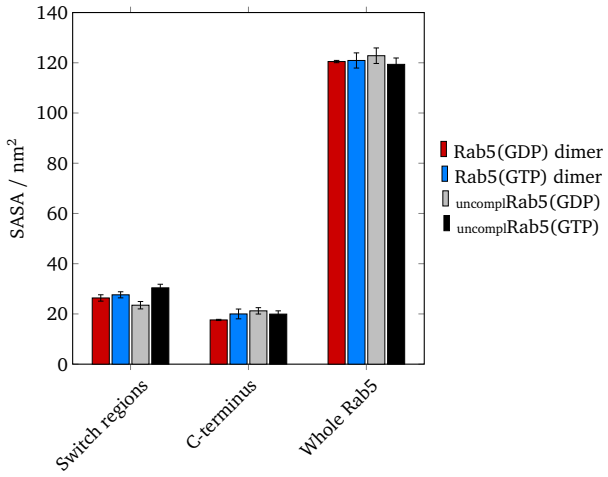


Figure B25: SASA of the Rab5 switch regions, the C-terminal HVR (residues 185-215) and the whole protein for dimeric Rab5(GDP) (red) and Rab5(GTP) (blue) compared to monomeric Rab5(GDP) (gray) and Rab5(GTP) (black) in the six-component membrane. Data was averaged over three independent simulations in case of monomeric Rab5 and over the two individual dimer monomers in case of dimeric Rab5.

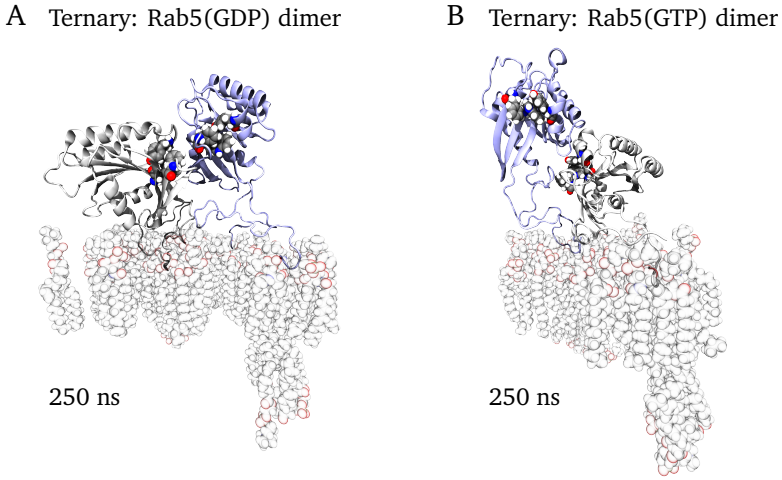


Figure B26: The residues of the invariant hydrophobic triad, Phe^{57}, Trp^{74}, and Tyr^{89}, were shielded from solvent in the Rab5(GDP) dimer, but effector-accessible in the Rab5(GTP) dimer. Structures shown were bound to the ternary mixture.

B.7 Additional results for the Rab5:GDI complex

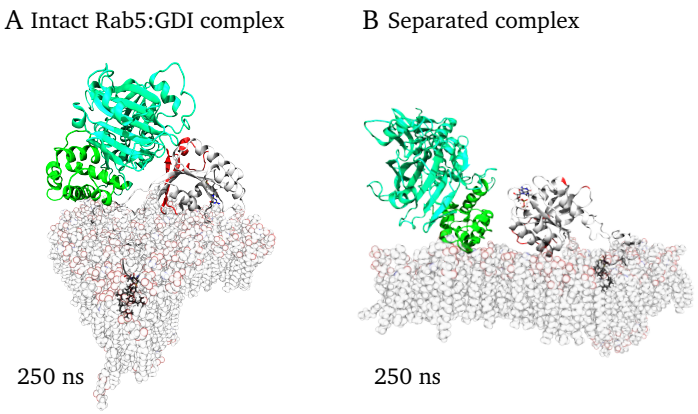


Figure B27: Final conformations of membrane-bound Rab5 A) in complex with GDI, $_{\text{insert}}$Run3 and B) separated from GDI due to distorted intermolecular interactions.

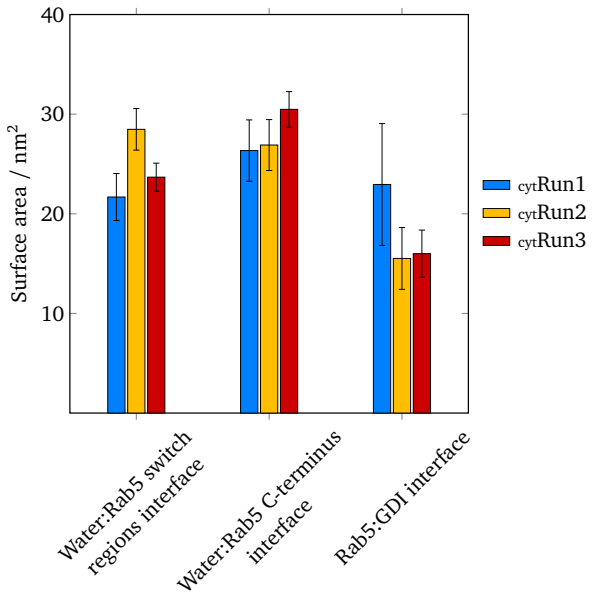


Figure B28: SASA of the Rab5 switch regions and HVR as well as the total Rab5:GDI interaction interface in the cytoplasmic complex. Data were averaged over 200 ns of MD simulation, error bars indicate deviations between the individual simulations.[279]

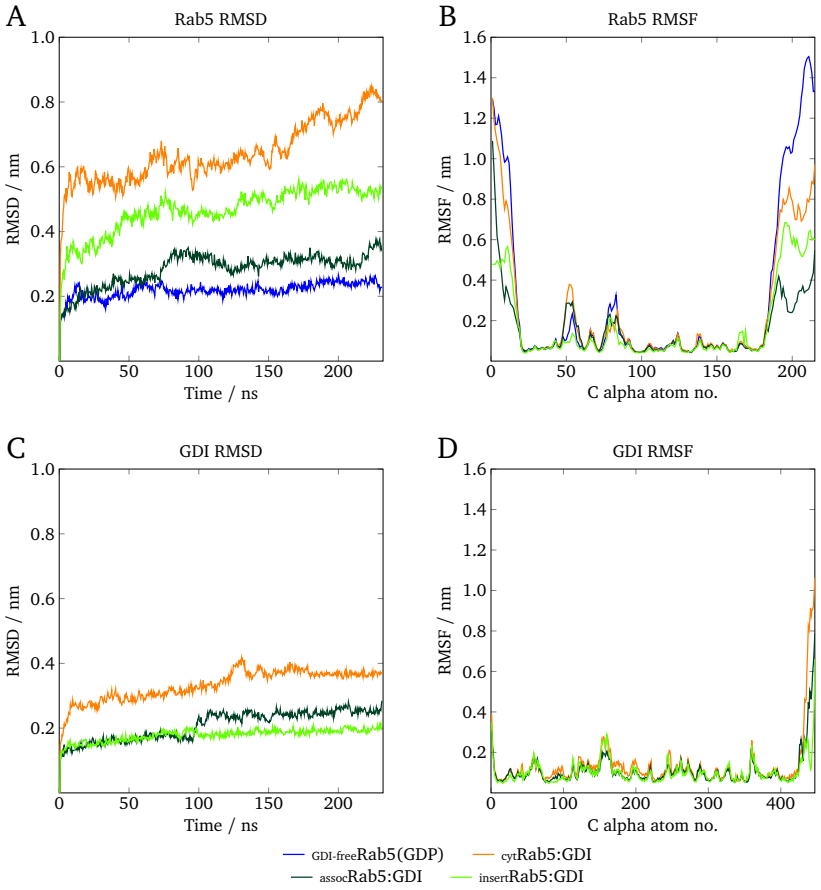


Figure B29: Comparison of the RMSD and RMSF of A) Rab5 and B) GDI in different simulations. Data are averaged over three independent simulations of GDI-free Rab5(GDP) anchored to the six-component membrane (blue) as well as the Rab5:GDI complex in cytoplasm (orange), membrane-associated (dark green) and membrane-bound (light green).

C List of Figures

D List of Tables

E List of Publications

Publications within the framework of this thesis

1. Edler,E.; Stein, M. Simulation of membrane-bound Rab5: structure and dynamics at different lipid compositions. *European Biophysics Journal with Biophysics Letters* **2015**, *44*, 165
2. Edler, E. et al. Membrane Localization and Dynamics of Geranylgeranylated Rab5 Hypervariable Region. *Biochimica Et Biophysica Acta-Biomembranes* **2017** *1859*, 1335-1349
3. Edler, E.; Stein, M. Probing the Druggability of Membrane-Bound Rab5 Investigated by Molecular Dynamics Simulations. *Journal of Enzyme Inhibition and Medicinal Chemistry* **2017** *32*, 434-443
4. Edler E.; Stein M. Recognition and stabilization of geranylgeranylated human Rab5 by the GDP Dissociation Inhibitor (GDI). *Small GTPases* **2017** 1-16

Earlier publications

1. Edler, E.; Stein, M. Spin-State-Dependent Properties of an Iron(III) Hydrogenase Mimic. *European Journal of Inorganic Chemistry* **2014** *22*, 3587-3599

Conference contributions

1. Biophysics Joint Meeting, **2015**, Hünfeld
2. 10th European Biophysics Congress, **2015**, Dresden
3. 30th Molecular Modeling Workshop, **2016**, Erlangen
4. 21st European Symposium on Quantitative Structure-Activity Relationship, **2016**, Verona