

Molecular mechanism of membrane sensing

Dissertation zur Erlangung des Doktorgrades
der Naturwissenschaften

vorgelegt beim
Fachbereich Biochemie, Chemie und Pharmazie
der Johann Wolfgang Goethe-Universität
in Frankfurt am Main

von
Michael Gecht
aus Mogiljow in Belarus

Frankfurt am Main 2021
(D 30)

Diese Arbeit wurde vom Fachbereich Biochemie, Chemie und Pharmazie der Johann Wolfgang Goethe-Universität als Dissertation angenommen.

Dekan: Prof. Dr. Clemens Glaubitz

1. Gutachter: Prof. Dr. Volker Dötsch

2. Gutachter: Prof. Dr. Gerhard Hummer

Datum der Disputation: 13. Dezember 2021

Die vorliegende Dissertation wurde in der Zeit von März 2017 bis Dezember 2020 am Max Planck Institut für Biophysik in der Abteilung Theoretische Biophysik angefertigt.

Contents

List of Abbreviations	viii
List of Tables	x
List of Figures	xiii
Zusammenfassung	xiv
Summary	xix
1 Introduction	1
1.1 Structure of lipids	1
1.2 Membrane properties	2
1.3 Membrane compositions in organelles	4
1.4 Proteins sensing membrane properties	5
1.5 Biosynthesis of glycerophospholipids	8
1.6 Different tastes of amphipathic helices	10
1.7 Viral replication in host organelles	12
1.8 Molecular dynamics	16
1.8.1 Force fields	16
1.8.2 Equations of motion	18
1.8.3 Periodic boundary conditions	19
1.8.4 Thermodynamic ensembles	19
1.9 Aim of this thesis	21
2 Lipid recognition by amphipathic helices	22
2.1 Introduction	22
2.2 Methods	23
2.2.1 Coarse-grained molecular dynamics simulations	23
2.2.2 Backmapping of CG systems to atomistic representations	24
2.2.3 Atomistic simulations of PIP5K AH	24
2.2.4 Analysis of lipid localization	25
2.2.5 Calculation of distances	26
2.2.6 Calculation of residence times	26
2.3 Results	26
2.3.1 Charged lipids preferentially localize along the Opi1 AH	27
2.3.2 Coordination of lipids is primarily driven by lysines	29
2.3.3 Phosphate moieties of PA lipids are the major interaction partners	30

2.3.4	PS headgroups allow for two possible modes of interaction	30
2.3.5	The 3K motif provides space only for PA	32
2.3.6	PA lipids can reside longer at the Opi1 AH	33
2.3.7	PIP5K AH interacts with a variety of lipids	34
2.3.8	PIP5K AH is lined with charged residues on all sides	36
2.3.9	PI4P interacts with PIP5K AH through electrostatic network	36
2.4	Discussion	38
3	Interaction of Opi1 AH with membranes	40
3.1	Introduction	40
3.2	Methods	41
3.2.1	Atomistic MD simulation	41
3.2.2	Secondary structure calculations	42
3.2.3	Distance calculations between peptide and membrane	42
3.2.4	Calculation of backbone dihedral angles	43
3.2.5	Sequence similarity analysis	43
3.2.6	Biased molecular dynamics simulations	43
3.3	Results	45
3.3.1	Generation of unfolded Opi1 AH conformations	45
3.3.2	Recruitment and tethering to membranes	46
3.3.3	Secondary structure formation at lipid-water interfaces	52
3.3.4	G120 mediates the AH folding and unfolding	55
3.3.5	Folding of the amphipathic helix using enhanced sampling	60
3.4	Discussion	65
4	Structural dynamics of the HIV-1 glycoprotein	68
4.1	Introduction	68
4.2	Methods	69
4.2.1	Membrane composition	69
4.2.2	Preparation of membrane based structures	70
4.2.3	Preparation of soluble HIV-1 gp140 structure	70
4.2.4	Determination of HIV-1 gp140 stability	71
4.2.5	Calculation of the hydrophobicity profile	71
4.2.6	Setup of membrane systems	72
4.2.7	Calculation of tilt angles	72
4.2.8	Distances of N- and C-termini	73
4.2.9	Density calculations	73
4.2.10	Number of hydrogen bonds	74
4.2.11	Alignment of antibody crystal structures	74
4.2.12	Preparation of antibody structures	74
4.2.13	Preparation of combined HIV-1 structures	75
4.2.14	Contacts between antibody and its epitope	76
4.3	Results	76

4.3.1	R696 adapts two stable conformations in monomeric gp41 TMD . . .	77
4.3.2	The gp41 TMD exhibits two conformations depending on R696 . . .	77
4.3.3	Interactions of MPER with the membrane depend on TMD orientation	79
4.3.4	Trimeric gp41 TMD rearranges in viral-like membranes	80
4.3.5	R696 snorkels to the cytofacial leaflet in the trimeric assembly . . .	82
4.3.6	Arginines restrict possible placements of TMD in lipid bilayers . . .	83
4.3.7	Rearrangement of trimeric structures in DMPC bilayers	85
4.3.8	Formation of MPER epitopes accessible by bnAbs	87
4.3.9	HIV-1 gp140 simulations in solution	89
4.3.10	Construction of an almost full-length HIV-1 model	90
4.3.11	Accessibility considerations of bnAbs to MPER	92
4.4	Discussion	94
5	Optimization of molecular dynamics simulation performance	97
5.1	Introduction	97
5.2	Methods	98
5.2.1	Additional background on simulations	98
5.2.2	General procedure	99
5.2.3	Systems description	99
5.2.4	Fitting Amdahl's law to performance data	100
5.3	Results	100
5.3.1	Implementation details of MDBenchmark	101
5.3.2	Biomolecular systems are mostly composed of solvent	102
5.3.3	Defining a minimal run time of benchmarks	102
5.3.4	Optimizing performance of two systems	104
5.3.5	Performance estimates from Amdahl's law	106
5.3.6	Optimal parameters depend on system size	107
5.3.7	Running multiple simulations on a single node	110
5.4	Discussion	111
5.5	Appendix	113
6	Conclusions	117
	Bibliography	122
	Declaration of contributions	149

List of Abbreviations

Amino acids are abbreviated using their single letter code.

ABMD	adiabatic bias molecular dynamics
ADP	adenosine diphosphate
AH	amphipathic helix
AIDS	acquired immune deficiency syndrome
ALPS	amphipathic lipid packing sensor
ATP	adenosine triphosphate
bnAb	broadly neutralizing HIV-1 antibody
CDP	cytidine diphosphate
CDP-DAG	cytidine diphosphate-diacylglycerol
CDR	complementary determining region
CG	coarse-grained
CLI	command-line interface
COM	center of mass
CPU	central processing unit
CRAC	cholesterol recognition/interaction amino acid consensus
CTD	C-terminal domain
CV	collective variable
DAG	diacylglycerol
DHPC	1,2-dihexanoyl-sn-glycero-3-phosphocholine
DMPC	1,2-dimyristoyl-sn-glycero-3-phosphocholine
DOPA	1,2-dioleoyl-sn-glycero-3-phosphate
DOPC	1,2-dioleoyl-sn-glycero-3-phosphocholine
DOPS	1,2-dioleoyl-sn-glycero-3-phosphoserine
Env	envelope
ER	endoplasmic reticulum
ERAD	endoplasmic reticulum-associated degradation
ESCRT	endosomal sorting complexes required for transport
Fab	antigen-binding fragment
FP	fusion peptide
GPI	glycosylphosphatidylinositol
GPL	glycerophospholipid
GPU	graphics processing unit
HIV	human immunodeficiency virus
HR1	heptad repeat 1 region
MAG	monoacylglycerol
MBP	maltose-binding protein

MCS	membrane contact site
MD	molecular dynamics
MPER	membrane-proximal extended region
MPI	message passing interface
MSA	multiple sequence alignment
NMR	nuclear magnetic resonance
OpenMP	open multi-processing
Opi1	over-production of inositol
PA	phosphatidic acid
PBC	periodic boundary condition
PC	phosphatidylcholine
PE	phosphatidylethanolamine
PG	phosphatidylglycerol
PI	phosphatidylinositol
PI(4,5)P ₂	phosphatidylinositol (4,5)-biphosphate
PI4P	phosphatidylinositol 4-phosphate
PLPC	1-palmitoyl-2-linoleoyl-sn-glycero-3-phosphocholine
PLPE	1-palmitoyl-2-linoleoyl-sn-glycero-3-phosphoethanolamine
PM	plasma membrane
POPA	1-palmitoyl-2-oleoyl-sn-glycero-3-phosphate
POPC	1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine
POPE	1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoethanolamine
PS	phosphatidylserine
RAM	random-access memory
REMD	replica-exchange molecular dynamics
RMSD	root-mean-square deviation
RMSF	root-mean-square fluctuation
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
SFA	saturated fatty acid
SREBP	sterol regulatory element-binding protein
SSM	sphingomyelin
TAG	triacylglycerol
TMD	transmembrane domain
TTS	time-to-solution
UPR	unfolded protein response

List of Tables

2.1	Lipid compositions used in Opi1 AH simulations.	24
2.2	Lipid compositions used for atomistic simulations of PIP5K with a model PM-like bilayer.	25
3.1	Sequences of Opi1 amphipathic helix (AH) used for atomistic molecular dynamics simulations.	41
3.2	Systems of Opi1 AH simulated with molecular dynamics without biases. . .	42
3.3	STRIDE secondary structure codes and mappings used for analysis.	42
3.4	Membrane compositions used for binding simulations with unfolded Opi1 peptides.	47
3.5	Mean, minimal and maximal distances between C_{α} atoms of unfolded Opi1 AH and membrane center.	49
4.1	Lipid composition of viral-like membranes.	70
4.2	Side chain dihedral angles used for the three R696 rotamers.	70
4.3	Missing residues in HIV-1 gp140 model that were constructed with MOD-ELLER.	70
4.4	Simulated systems of HIV-1 glycoprotein structures.	73
4.5	Atom groups used to calculate tilt angle of helical segments in gp41 TMD structures.	73
4.6	Epitopes of bnAbs superimposed onto gp41 MPER-TMD structures.	74
4.7	Putative hydrogen bonds and salt bridge contacts between 4E10 antibody and its MPER epitope.	76
5.1	Atomistic simulation systems used for benchmarks.	113

List of Figures

1.1	Diversity of different lipid species.	2
1.2	Physicochemical properties of membranes induced by lipids.	3
1.3	The endoplasmic reticulum and plasma membrane form two lipid bilayer extremes.	4
1.4	Regulation of membrane saturation by DesK/DesR in <i>Bacillus subtilis</i>	6
1.5	Ire1 senses the membrane elasticity as a function of its lipid saturation.	6
1.6	CCT α regulates the synthesis of PC.	7
1.7	PA is at the center of phospholipid biosynthesis.	8
1.8	Opi1 senses PA at the ER membrane and downregulates GPL biosynthesis in the nucleus.	10
1.9	Different types of amphipathic helices.	11
1.10	Late phase of HIV-1 infection.	14
1.11	HIV-1 Env incorporation into budding viral particles.	15
2.1	Snapshot of a fully folded Opi1 AH embedded in a lipid bilayer.	27
2.2	Time-averaged positions of phosphate headgroups from DOPA lipids.	28
2.3	Time-averaged positions of phosphate headgroups from DOPS lipids.	28
2.4	Time-averaged positions of phosphate headgroups from DOPA and DOPS lipids.	29
2.5	Structure of fully folded Opi1 AH when embedded in a lipid bilayer.	30
2.6	PA interacts with the Opi1 AH at different interaction sites.	31
2.7	PS interacts with the “3K motif” in two different modes.	31
2.8	Probability density of the distance between the “3K motif” and lipid phosphates.	32
2.9	Probability density of the distance between the “KRK motif” and lipid phosphates.	33
2.10	Residence times of different lipid species at the “3K motif” and “KRK motif”.	34
2.11	Time-averaged positions of different lipids in a PM-like model bilayer.	35
2.12	Structure of fully folded PIP5K AH embedded in a model PM-like lipid bilayer.	36
2.13	Charged lipids interact with PIP5K AH.	37
3.1	Simulation snapshots of Opi1 30-mer at 400 K.	46
3.2	Structural composition of Opi1 30-mer at 400 K.	46
3.3	Initial structure of Opi1 30-mer binding simulations.	47
3.4	Distance distributions of backbone C α atoms to membrane center for four different membrane compositions.	48

3.5	Distances of backbone C_α atoms to membrane center for four different membrane compositions as a function of time.	50
3.6	Distance density of L113/I117 C_α atoms to membrane center.	51
3.7	Flotation assay of Opi1 AH WT and I117 mutants.	51
3.8	Secondary structure ratio of Opi1 30-mer with different membrane compositions.	53
3.9	Snapshots of Opi1 30-mer at lipid-water interfaces of membranes with different compositions.	54
3.10	Partially folded Opi1 22-mer AH from unbiased simulations.	54
3.11	Backbone dihedral angles ϕ and ψ follow the folding of an α -helix.	55
3.12	Sequence similarity of Opi1 AH and similar sequences.	56
3.13	<i>S. cerevisiae</i> strains grown on inositol lacking media secrete inositol.	57
3.14	Structural composition of 22-mer peptide with the G120A point mutation.	58
3.15	Dihedral angle time series for residues A120-N123 of G120A point mutant.	59
3.16	Structural composition of 22-mer peptide with the G120A/K121A double mutant.	60
3.17	Dihedral angle time series of a G120A/K121A double mutation simulation.	61
3.18	Observables from ABMD simulations for different values of κ	62
3.19	Snapshots from ABMD folding simulation of AKGK in 22-mer peptide.	63
3.20	Dihedral angles of 22-mer Opi1 AH from steered MD simulation.	64
3.21	Snapshots from steered MD folding simulation of 22-mer Opi1 AH.	65
4.1	Orientation of R696 in monomeric gp41 TMD.	77
4.2	Stable configurations from two independent atomistic simulations with a different starting R696 rotamer.	78
4.3	gp41 monomers adopt different tilt angles depending on the orientation of R696.	78
4.4	The orientation of R696 determines different conformations of gp41 MPER.	79
4.5	gp41 MPER-TMD constructs adopt the same tilt angle.	80
4.6	Trimeric gp41 TMD rearranges during atomistic simulations.	81
4.7	R696 orients towards the cytofacial region in trimeric gp41 TMD.	81
4.8	N-termini of trimeric gp41 TMD stay close, while the C-termini dissociate.	82
4.9	Trimeric gp41 TMD protomers obtain high tilt angles.	82
4.10	gp41 TMD trimer rearranges in a lipid bilayer.	83
4.11	R696 snorkels to the cytofacial leaflet with its polar side chain.	83
4.12	Arginine residues form hydrogen bonds with water molecules.	84
4.13	Analysis of the gp41 TMD hydrophobicity in 18-residue windows.	84
4.14	gp41 TMD cannot span the membrane without at least one arginine inside the bilayer.	85
4.15	Trimeric gp41 MPER-TMD construct dissociates in DMPC bilayers.	86
4.16	gp41 MPER-TMD rearranges into conical trimers.	86
4.17	Pairwise distances of gp41 MPER-TMD trimers in DMPC membranes.	87

4.18	gp41 MPER-TMD fragments superimposed onto corresponding regions from MD simulation.	88
4.19	gp41 MPER-TMD with superimposed bnAb crystal structures.	88
4.20	Crystal structure of HIV-1 Env as monomer and trimer.	89
4.21	HIV-1 gp140 RMSF and RMSD observables in solution.	90
4.22	Terminal distances in HIV-1 gp140 and gp41 MPER-TMD constructs. . . .	91
4.23	Model of HIV-1 Env with a membrane anchor embedded into a DMPC bilayer.	92
4.24	Model of HIV-1 Env with bound 4E10 and a membrane anchor embedded into a DMPC bilayer.	93
4.25	Distances of putative hydrogen bonds and crystal contacts between bnAb and its MPER epitope.	93
5.1	Workflow and implementation of MDBenchmark.	101
5.2	Size and composition of MD simulation systems.	103
5.3	Performance estimated for different run times.	103
5.4	Scaling of two systems on CPU-only nodes.	104
5.5	Scaling of two systems on mixed CPU-GPU nodes.	105
5.6	Amdahl's law captures the scaling behavior.	106
5.7	Optimal number of MPI ranks depends on system size.	107
5.8	Performance comparison between different architectures.	109
5.9	Multiple simulations per node increase performance.	110
5.10	Scaling of all 23 MD systems on CPU-only and mixed CPU-GPU nodes. . .	114

Zusammenfassung

Das zelluläre Leben ist auf Biomembranen angewiesen. Diese Membranen trennen das Zellinnere von ihrer Umgebung ab und teilen die Zelle in etliche kleinere Kompartimente auf, sogenannte Organellen. Membranen nehmen eine aktive Rolle in der Aufrechterhaltung vieler Zellfunktionen ein und stellen dabei nicht nur passive Barrieren dar. Lipide und Proteine bilden die Hauptbestandteile von Membranen und ihre Zusammensetzung unterscheidet sich von Organelle zu Organelle. Proteine können in mehrere Kategorien eingeteilt werden. Einige von ihnen werden bei der Translation in die Membran inseriert und bleiben mithilfe Ihrer Transmembrandomänen verankert. Andere lösliche Proteine werden zwar bei der Translation durch die Membran transportiert, bleiben aber nicht mit dieser assoziiert. Diese können posttranslationalen Modifikationen unterzogen werden, bevor sie ihre Zielmembran erreichen. Proteine, die vollständig löslich sind, werden ohne jegliche Membranassoziation im Zytosol translatiert. Mithilfe von speziellen Membranbindungsdomänen können diese Proteine aber zu spezifischen Organellen rekrutiert werden. So sind diese drei Proteinkategorien auf die korrekte Funktion von Organellen angewiesen. Da die Lipidzusammensetzung einer Organelle einen Einfluss auf die Funktion von Proteinen hat, muss die ständige Produktion und Wiederverwertung von Lipiden aufrechterhalten werden. Dadurch schafft es die Zelle, sich auf Veränderungen ihrer Umgebung anzupassen. Im umgekehrten Sinne muss auch die Membran die richtige Lipidzusammensetzung aufweisen, um die korrekte Proteinfunktion zu ermöglichen. Die Membranen von Organellen haben sich also auf ihre Proteinbestandteile angepasst und umgekehrt. So sind zum Beispiel einige eukaryotische Proteine darauf spezialisiert, die Lipidzusammensetzung von Membranen zu erkennen. Während das Protein Mga2 die Membranfluidität registriert, dient Ire1 als mechanischer Sensor und passt sich an die Steifigkeit der näheren Umgebung an. Diese Sensoren regulieren die Lipidhomöostase in der Zelle, indem sie auf die Veränderungen der Membraneigenschaften reagieren und Signale an nachgeschaltete Prozesse senden.

In dieser Thesis wurden Molekulardynamik (MD) Simulationen verwendet, um die Interaktion verschiedener Proteine mit Biomembranen zu untersuchen. MD-Simulationen ermöglichen es, die Wechselwirkungen von Proteinen mit Lipiden im atomistischen Detail zu studieren. Dabei dienen MD-Simulationen als ein sogenanntes “Computermikroskop”, welches eine höhere zeitliche und räumliche Auflösung erlaubt als gängige experimentelle Methoden. Diese Thesis ist in mehrere Kapitel aufgeteilt, welche sich mit unterschiedlichen Aspekten der Protein-Lipid-Interaktion auseinandersetzen. So wird zuerst die Sensorfähigkeit von amphipathischen Helices (AHs) von Opi1 und PIP5K untersucht, welche es ihnen ermöglicht, die Lipidzusammensetzung von Biomembranen zu detektieren. Darüber hinaus beschäftigt sich ein Kapitel mit der Faltung der Opi1 AH an Lipid-Wasser-Grenzflächen. Das darauf folgende Kapitel zeigt die Konformationsdynamik der HIV-1 Env Glykoproteine in Biomembranen. Im letzten Kapitel wird die Leistung von MD-Simulationen getunt,

wodurch eine effizientere Nutzung der verfügbaren Rechenkapazitäten ermöglicht wird.

Die Lipidhomöostase wird durch unterschiedliche zelluläre Signalwege reguliert. Dabei steht in *S. cerevisiae* die Phosphatidsäure (PA) im Mittelpunkt. Mit ihrer Hilfe können komplexere Phosphoglyceride produziert werden. Zu diesen zählen Phosphatidylcholin (PC), Phosphatidylethanolamin (PE), Phosphatidylinositol (PI), Phosphatidylserin (PS) sowie Phosphatidylglycerol (PG). Wenn ausreichend PA in der Membran des endoplasmatischen Retikulums (ER) vorhanden ist, kann diese als Vorprodukt zur Produktion weiterer Phosphoglyceride genutzt werden. Die dafür benötigten Gene werden hierbei aktiv im Zellkern von *S. cerevisiae* transkribiert. Die Phosphoglycerid-Synthese wird herunterreguliert, wenn PA nicht im ausreichenden Ausmaß verfügbar ist. Das Protein Opi1 ist in der Lage das Vorhandensein von PA zu ertasten und ist somit in die Regulation der Phosphoglycerid-Synthese in *S. cerevisiae* involviert. Dabei vermittelt die Membranbindedomäne des Proteins in der Anwesenheit von PA die Interaktion mit dem ER. Wenn der Vorrat des Lipids verbraucht wurde, begibt sich Opi1 in den Zellkern der Zelle und reguliert dort die Transkription von denjenigen Genen herunter, die für die Phosphoglycerid-Synthese verantwortlich sind. Die PA-Bindedomäne von Opi1 wurde als AH charakterisiert, da sie in der Präsenz von Liposomen eine α -helikale Sekundärstruktur aufweist. In dieser Thesis wurde die Lipiderkennung der Opi1 AH mithilfe von atomistischen MD-Simulationen untersucht. Hier konnte gezeigt werden, dass die AH ihre α -helikale Sekundärstruktur nach Einbettung in Biomembranen beibehält. Gleichzeitig wurde die Interaktion von den negativ geladenen Lipiden PA und PS an zwei gegenüberliegenden Regionen der AH beobachtet. Diese beiden Regionen bestehen aus jeweils drei positiv geladenen Resten, welche in der Thesis als “KRK-Motiv” und “3K-Motiv” bezeichnet werden. Die Anordnung dieser Aminosäuren auf der α -helikalen Windung steuert zur Lipidspezifität der AH bei. Das “3K-Motiv” bildet eine schmale Bindungstasche, welche nur Platz für die kleinen Kopfgruppen von PA Lipiden bietet. Obwohl PS Lipide binden können, werden diese durch elektrostatische Interaktionen ihrer größeren Kopfgruppe in einem größeren Abstand gehalten. Unter Vorhandensein von PA und PS in der gleichen Biomembran wurde beobachtet, dass PA die Fähigkeit besitzt, die Bindetaschen für sich zu beanspruchen und so PS zu vertreiben. Die kleine Kopfgruppe von PA ermöglicht eine bessere Bindung an das “3K-Motiv”. Mithilfe dieser simplen Motive ist es der AH von Opi1 möglich, zwischen den Kopfgruppen von verschiedenen Lipidspezies zu unterscheiden. Diese Ergebnisse tragen dazu bei, den molekularen Mechanismus dieses Membransensors besser zu verstehen.

Das Protein PIP5K katalysiert die Reaktion von Phosphatidylinositol-4-phosphat (PI4P) zu Phosphatidylinositol-4,5-bisphosphat (PI(4,5)P₂). Phosphatidylinositole sind essenzielle Botenstoffe in der Lipid-Signalübertragung und dem Membrantransport. Die Erkennung von PI4P durch PIP5K erfolgt durch eine sogenannte “Spezifitätsschleife” (engl. “specificity loop”). Obwohl die Sekundärstruktur in den vorhandenen Kristallstrukturen von PIP5K nicht aufgelöst ist, konnte gezeigt werden, dass diese Spezifitätsschleife bei Membranbindung eine AH bildet. In dieser Thesis wurden atomistische MD-Simulationen genutzt, um die Lipidbindung von PI4P an die PIP5K AH zu charakterisieren. In mehreren Simulationen konnte gezeigt werden, dass PI4P in der Lage ist, an beide Seiten der AH

von PIP5K zu binden. Die Aminosäuren sind im Gegensatz zu denen auf der AH von Opi1 nicht in spezifischen Bindetaschen lokalisiert. Stattdessen finden sich die basischen Aminosäuren auf der gesamten AH verteilt. Der großen, geladenen Kopfgruppe von PI4P ist es möglich, sich großflächig auf die AH zu legen und damit mit mehreren positiv geladenen Aminosäuren gleichzeitig zu interagieren. Ebenfalls wurde in den Simulationen gezeigt, dass PI4P in der Lage ist, andere geladene Lipide wie z. B. jene mit PS Kopfgruppe, von der AH zu verdrängen. Hierzu reicht es mit seiner großen Kopfgruppe über das andere Lipid hinweg und beansprucht die elektrostatischen Interaktionen mit der AH für sich. Die hochgeladene Oberfläche dieser AH hat sich auf die Bindung von PI4P Lipiden spezialisiert, wodurch diese im nahe gelegenen katalytischen Zentrum zu PI(4,5)P₂ umgewandelt werden können.

Im ungebundenen Zustand weisen AHs keine Sekundärstruktur auf und bilden diese erst nach oder während der Membranbindung aus. AHs werden zuerst durch elektrostatische Wechselwirkungen an Biomembranen rekrutiert, woraufhin sie Sekundärstrukturen ausbilden und sich in die Membran inserieren. Mithilfe von atomistischen MD-Simulationen wurde in dieser Thesis die Rekrutierung, Assoziation, Insertion und Faltung der Opi1 AH an Biomembranen untersucht. Hierzu wurde eine vollständig entfaltete Opi1 AH oberhalb Biomembranen unterschiedlicher Lipidzusammensetzung platziert. Obwohl Membranen mit einer Zusammensetzung von 100% PC in der Lage waren, die AH zu rekrutieren, wurde keine dauerhafte Assoziation beobachtet. Hingegen konnten Membranen mit PE oder PA Lipidkopfgruppen die AH permanent rekrutieren und binden. Auf die Rekrutierung der Opi1 AH folgte in jeder Simulation die Insertion einer von zwei hydrophoben Aminosäuren, L113 oder I117. In einer Simulation wurde darüber hinaus die Faltung von zwei α -helikalen Windungen beobachtet. Ein konserviertes Glycin G120, welches beidseitig von je einem Lysin umgeben ist, stoppte die weitere Faltung. Darauf Folgende *in vivo* Experimente konnten zeigen, dass Glycin G120, für die Regulation der Opi1 AH Membranbindung verantwortlich ist. Eine G120W Punktmutation führte zum Kompletterlust der PA-Sensitivität, wodurch Opi1 auch nach dem Erschöpfen der PA Vorräte durchgehend an der ER Membran gebunden blieb. Um zu testen, ob sich die Faltung mithilfe von Punktmutationen auch in Simulationen zeigen lässt, wurde die AH mit einer G120A Einfach- und G120A/K121A Doppelmutation simuliert. In diesen Simulationen wurde die zeitweise Bildung einer α -helikalen Sekundärstruktur beobachtet, wobei die AH sich nie vollständig in die Biomembran inserierte. Um den Faltungs- und Insertionsprozess zu beschleunigen, wurde auf ein sogenanntes “enhanced sampling” Verfahren zurückgegriffen. Hierfür wurde die partiell gefaltete AH aus einer früheren Simulation genommen, und es wurde Kraft auf die ϕ und ψ Diederwinkel des Polypeptid-Rückgrates ausgeübt. Dadurch konnte die vollständige α -helikale Sekundärstruktur der Opi1 AH generiert werden, welche sich anschließend in die Biomembranen inserierte. Diese Simulationen tragen zum besseren Verständnis der Interaktion von ungefalteten AHs an Biomembranen sowie ihrer Insertion und Faltung an der Lipid-Wasser-Grenzfläche bei.

Künstliche Membranen werden häufig zur Strukturaufklärung eingesetzt, um die native Membranumgebung von Proteinen zu imitieren. Während solche künstlichen Membranen

leichter zu handhaben sind und zur Stabilisierung des Proteins beitragen können, unterscheiden sie sich in ihren biophysikalischen Eigenschaften von den Membranen in lebendigen Zellen. Beispielsweise werden für Kernspinresonanz Experimente häufig Bizellen genutzt. Diese setzen sich in vielen Fällen aus den kurzkettigen Lipiden 1,2-Dimyristoyl-sn-Glycero-3-Phosphatidylcholin (DMPC) und 1,2-Dihexanoyl-sn-Glycero-3-Phosphatidylcholin (DHPC) zusammen. Mithilfe solcher Bizellen und Kernspinresonanz wurde kürzlich die Tertiärstruktur der Transmembrandomäne (TMD) des HIV-1 Virushüllen Glykoproteins Env gelöst. Trotz gleicher experimentaler Bedingungen wird die Struktur von unterschiedlichen Forschungsgruppen sowohl als Trimer als auch als Monomer beschrieben. Die Trimerstruktur zeigt eine C_3 -Symmetrie, während die Monomerstruktur eine überaus lange TMD beschreibt. Die Unterschiede in diesen Strukturen wurden zum Anlass genommen, um deren Stabilität und Dynamik mithilfe von MD-Simulationen zu untersuchen. Als Erstes wurde die Monomerstruktur der TMD in Biomembranen simuliert, wessen Lipidzusammensetzung die der Virushülle imitieren soll. Die Simulationen zeigten, dass die konservierte Aminosäure R696, welche sich in der Mitte der TMD befindet, sich nach nur wenigen Nanosekunden in Richtung der Lipid-Wasser-Grenzfläche umorientiert. Diese Umorientierung, in der Thesis als "schnorcheln" bezeichnet, beeinflusst die Gesamtkonformation der TMD. Darüber hinaus wird durch den geladenen Rest Wasser in den hydrophoben Bereich der Membran gezogen. In den Simulationen der C_3 -symmetrischen Trimerstruktur wurde ebenfalls Wasser durch die drei R696 Aminosäuren in die Membran gezogen, wodurch die symmetrische Struktur am C-terminalen Ende auseinanderfiel. Um zu testen, ob das Fehlen der N-terminalen Membran-Proximalen Erweiterten Region (MPER) oder die unterschiedliche Lipidzusammensetzung der Membranen zum Auseinanderfallen der Strukturen führte, wurden neuen Simulationen aufgesetzt. Hierfür wurde die ebenfalls veröffentlichte C_3 -symmetrische MPER-TMD Trimerstruktur in DMPC Membranen simuliert. Erneut wurde die Dissoziation der C-terminalen Region der TMD in sechs unabhängigen Simulationen beobachtet. Die zwei α -helices der MPER Region sind durch eine kurze, unstrukturierte Region miteinander und der TMD verbunden. Die unstrukturierte Region zwischen der MPER Region und der TMD formte in allen Simulationen nach kurzer Simulationszeit eine α -helikale Sekundärstruktur und wurde somit zur Erweiterung der TMD. Überraschenderweise formten sich durch diese Umstrukturierung die Epitope von breitneutralisierenden Anti-HIV-Antikörpern (bnAbs). Das co-kristallisierte Epitop in den Kristallstrukturen der bnAbs wurde mit dem Simulationsepitop überdeckt, wodurch die Übereinstimmung der Strukturen gezeigt werden konnte. Zum Abschluss wurde ein Volllänge Modell des HIV-1 Env-Glykoproteins, ohne die C-terminale Domäne modelliert. Anhand dieses Modells wurde die Zugänglichkeit des Epitops dargestellt. Obwohl das Modell keine Glykosylierung aufweist, dient es zur Veranschaulichung wie Antikörper an die Virushüllenproteine von HIV-1 binden können.

In dieser Thesis wurden MD-Simulationen verwendet, um die Dynamik von Proteinen zu untersuchen, welche von Doppellipidschichten umgeben sind. Mithilfe von MD-Simulationen kann die zeitliche Entwicklung von Partikelsystemen berechnet werden. Die Bewegung der Teilchen folgen dabei den Regeln eines "Kraftfeldes", welches die Potenziale von intra-

und intermolekulare Wechselwirkungsenergien definiert. Dabei werden die Dehnung und Stauchung der Bindungslängen zwischen Atomen, die Oszillation ihrer Bindungswinkel und Torsionsschwingungen sowie die Wechselwirkungen durch hydrophobe und elektrostatische Kräfte berücksichtigt. Die Zeitentwicklung wird in kleinen Schritten in der Größenordnung von 10^{-15} s berechnet, während biologische Prozesse auf der Größenordnung von 10^{-9} s bis 10^3 s ablaufen. Da zur Beobachtung dieser Prozesse viele Zeitschritte berechnet werden müssen, werden Supercomputer wegen ihrer enormen Rechenkapazität eingesetzt. Diese Supercomputer bestehen aus Hunderten bis Tausenden einzelner Rechenknoten, welche jeweils mehrere CPUs und GPUs enthalten können. Die Zeitentwicklung von MD-Simulationen wird mit spezialisierter Software berechnet, die in der Lage sind, die heterogene Supercomputer Hardware zu nutzen. Gleichzeitig ist es möglich, verschiedene Einstellungen vorzunehmen, um die Berechnungsgeschwindigkeit anzupassen. In dieser Thesis wurde das Softwarepaket MDBenchmark entwickelt, welches es den Nutzer*Innen erlaubt, die bestmöglichen Einstellungen für Ihre MD-Simulationen herauszufinden. MDBenchmark wurde verwendet, um Skalierungsstudien an 23 repräsentativen MD-Systemen durchzuführen. Hierbei wurde die Anzahl der “message passing interface” (MPI) Ranks und “open multiple process” (OpenMP) Threads variiert, während jedes System gleichzeitig auch von 1–10 Rechenknoten skaliert wurde. Auf reinen CPU-Knoten, d. h. Knoten ohne GPUs wird die höchste Berechnungsgeschwindigkeit erreicht, wenn die Anzahl der MPI Ranks der Anzahl der physischen Kerne entspricht. Auf gemischten CPU-GPU-Knoten hängt die Anzahl der MPI Ranks allerdings von der Systemgröße ab. Ein System mit 300 t Atomen erreicht die höchste Berechnungsgeschwindigkeit mit $n_{\text{ranks}} = 10$. Im Gegensatz dazu erreichen Systeme mit mehr als 1 M Atomen die beste Berechnungsgeschwindigkeit mit denselben Einstellungen wie reine CPU-Knoten. Diese Skalierungsstudien haben des Weiteren gezeigt, dass der Übergang von einem zu zwei gemischte CPU-GPU-Knoten nur zu einer geringen Leistungssteigerung führt, da nicht alle Ressourcen voll ausgenutzt werden. Um das Potenzial der übrigen Ressourcen zu testen, wurden statt einer mehrere Simulationen auf einem einzigen gemischten CPU-GPU-Knoten berechnet. Wenn bis zu vier unabhängige Simulationen auf demselben Knoten ausgeführt werden, kann die kumulierte Gesamtleistung der Simulationen um einen Faktor von 1.5 bis 4 gesteigert werden. Diese Skalierungsstudie zeigt, dass MDBenchmark dabei helfen kann, die Berechnungsgeschwindigkeit von Simulationen in nur wenigen Schritten zu steigern, ohne detaillierte Kenntnisse über die Funktionsweise von Supercomputern zu haben.

Diese Thesis trägt zum besseren Verständnis von Protein-Lipid-Interaktionen aus der Perspektive verschiedener Proteine bei. Während AHS die Lipidzusammensetzung von Biomembranen spüren und unterscheiden können, müssen sie sich zunächst Falten und in die Membran inserieren. Die Stabilität der Tertiärstruktur der HIV-1 Virushüllen Env-Glykoproteine ist von der Lipidzusammensetzung abhängig und die Dynamik ihrer Struktur kann mit MD-Simulationen untersucht werden. Mit Hilfe des MDBenchmark-Softwarepakets, welches hier vorgestellt wurde, lässt sich die Berechnungsgeschwindigkeit dieser MD-Simulationen verbessern.

Summary

Lipid membranes are a crucial part of biological life. Membranes separate the cellular inside from the outside environment, but also divide the cell into further compartments, so-called organelles. Membranes are not only passive barriers but also play a crucial role in all aspects of the cellular function. The constituents of membranes, lipids, and proteins, interact in synergy and orchestrate what we know as life. Proteins can be divided into three categories. Upon translation of proteins, some are inserted into the membrane and remain anchored with the help of transmembrane domains, while others remain soluble but are transported through the membrane. All of these proteins undergo further processing in specialized organelles, before arriving at their final destination. Other soluble proteins are translated in the cytosol without any membrane anchorage, but some have special membrane-binding domains that recruit them towards lipid bilayers. All of these proteins are dependent on the proper function of organelles. And because the lipid composition of an organelle directly affects its function, the production and recycling of lipids have to be kept up constantly to allow the cell to adapt to changes in the internal or external environment. Vice versa the membrane must also have the correct lipid composition to allow for proper protein function. Both organellar membranes, as well as proteins, have adapted to one another. For example, some eukaryotic proteins are specialized to sense the lipid composition of membranes, i.e., the protein Mga2 senses the fluidity of the membrane, whereas Ire1 is a mechanical sensor that adapts to its stiffness. These sensors can react to changes in the membrane properties and signal to downstream processes that ultimately adapt the lipid homeostasis of the cell.

In this thesis, molecular dynamics (MD) simulations are used to study the interaction of different proteins with lipid bilayers. MD simulations can be used as a “computational microscope” to gain atomistic insights into the interactions between proteins and lipids that can barely be accessed in such detail by experimental methods. The different chapters of this thesis address the lipid sensing functionality of amphipathic helices (AHs) when bound to membranes, the folding of AHs at lipid-water interfaces, as well as the conformational dynamics of the HIV-1 Env glycoproteins in viral-like and experimental bilayers. In the last chapter, the possibilities to enhance the performance of MD simulations are explored, leading to the more efficient usage of computational resources.

The regulation of lipid homeostasis is orchestrated through a variety of cellular pathways. In the presence of phosphatidic acid (PA), the cell is able to produce the more complex glycerophospholipids (GPLs) phosphatidylcholine (PC), phosphatidylethanolamine (PE), phosphatidylserine (PS), phosphatidylinositol (PI) and phosphatidylglycerol (PG). The genes responsible for this *de novo* synthesis are actively transcribed in the cell nucleus of *S. cerevisiae* when PA is abundant. In contrast, when PA is depleted from the endoplasmic reticulum (ER) membrane, the biosynthesis of GPLs has to be downregulated. In *S. cere-*

visiae the protein Opi1 is at the center of the GPL *de novo* synthesis. Its membrane-binding domain retains it at the ER membrane when PA is present and Opi1 shuttles into the nucleus to downregulate the biosynthesis of GPLs in its absence. Opi1 was characterized to contain a PA-binding domain, as it lacks any specificity to PA in its absence. This domain was also shown to have an amphipathic character and to fold into an AH in the presence of PA containing liposomes. In this thesis, the lipid recognition of the Opi1 AH was studied with extensive atomistic MD simulations. After being embedded into lipid bilayers, the AH stayed associated and a fully folded α -helix in all simulations. At the same time, negatively charged lipids such as PA and PS were recruited to two distinct regions of opposite sides of the AH. These two regions are composed of three positively charged residues each and were named “KRK motif” and “3K motif”. The residues in both regions are located on different heights of separate turns along the α -helix and provide the lipid specificity. The “3K motif” forms a narrow binding pocket that only allows PA lipids to come into proximity. PS lipids are also able to bind but remain at a bigger distance to the AH. In mixed bilayers with both PA and PS present, PA outcompetes PS for interactions at both motifs. More specifically, the small headgroup of PA allows for a more intimate binding to the “3K motif”, whereas PS lipids are held at a bigger distance. These simple motifs of three basic residues allow the Opi1 AH to differentiate between the headgroups of lipids. These results help to better understand the molecular mechanism of membrane sensing.

Similar to Opi1, the protein PIP5K catalyzes the phosphorylation from phosphatidylinositol 4-phosphate (PI4P) to phosphatidylinositol (4,5)-biphosphate (PI(4,5)P₂). Phosphoinositides are required for lipid signaling and membrane trafficking. A “specificity loop” in PIP5K is responsible for PI4P recognition, but its secondary structure is missing in available structures. This specificity loop was later shown to be an AH, similar to that of Opi1. In this thesis, extensive atomistic MD simulations were used to characterize the lipid-binding of PI4P to the PIP5K AH. PI4P was shown to be able to bind to both sides of the PIP5K AH in simulations. In contrast to the Opi1 AH, the residues are not localized to two distinct binding pockets but are found all around the AH. The charged and bulky headgroup of PI4P wrapped around the AH when in proximity and stably interacted with multiple charged residues at once. If other lipids such as 1,2-dioleoyl-sn-glycero-3-phosphoserine were interacting with one of the positively charged residues on the AH, PI4P was able to outcompete them by reaching over their smaller headgroups in multiple instances. The highly charged surface of the AH was made to recruit PI4P lipids and allow for its proximal catalytic side to phosphorylate the lipid to PI(4,5)P₂.

The sensing of folded and embedded AH in lipid bilayers is predated by the insertion of the unstructured region into the membrane. AHs are recruited to membranes by charged interactions, followed by folding and insertion into the membranes. In this thesis, the recruitment, tethering, insertion, and folding of the Opi1 AH to membranes was studied using extensive atomistic MD simulations. A fully unfolded Opi1 AH was placed above lipid membranes of different lipid compositions containing PC, PE, and PA lipids to study the recruitment. Recruitment to 100% PC membranes was observed, but the association was

never permanent. In contrast, the AH is permanently recruited to membranes containing PE or PA. After recruitment, two hydrophobic residues, L113 and I117 tethered the unfolded Opi1 AH to the membrane. The partial folding of two α -helical turns were observed in one simulation instance. The folding stopped at a highly conserved glycine roughly at the center of the AH. This glycine, G120, is surrounded by one lysine on each side that disallowed further folding. *In vivo* experiments have confirmed that G120 acts as a regulator of Opi1 function. Replacement of this residue with a more hydrophobic tryptophan keeps the protein at the ER, independent of the PA abundance. This led to the assumption that G120 acts as a switch, both slowing down folding as well as speeding up the dissociation from the membrane. The ability of G120A single mutants and G120A/K121A double mutants to more quickly fold in lipid bilayers was assessed with additional atomistic simulations. Partial folding was observed in multiple simulations, but the AH never properly inserted into the lipid-water interface. Enhanced sampling methods were used with the partially folded AH embedded in the lipid bilayer, to better understand the folding process. The AH was successfully folded by applying a force on the ϕ and ψ dihedral angles of residues located on separate helical turns. In these simulations, all turns folded outside of the membrane, and the AH inserted in a single step. These simulations help to better understand the sensing of AHs at lipid bilayers and how they fold at the lipid-water interface.

Structures of membrane proteins are often solved in artificial membranes that mimic the native environment. These membranes are easier to handle and manage to properly stabilize the protein, but differ in their biophysical properties. For example, bicelles that are commonly used in NMR experiments are composed of the very short 1,2-dimyristoyl-sn-glycero-3-phosphocholine (DMPC) and 1,2-dihexanoyl-sn-glycero-3-phosphocholine (DHPC) lipids. Recent experimental studies have solved the NMR structure of the TMD of the HIV-1 Env glycoprotein gp41. Although the experimental procedures were identical, the solved structures were not. The first structure of HIV-1 gp41 TMD revealed a C_3 symmetric bundle of α -helices. In contrast, the second structure has revealed a monomeric TMD. Although it is clear that the HIV-1 Env glycoprotein is a homotrimer, these differences in the tertiary structure are important to better understand its function. In this thesis, these differences are addressed with MD simulations. First, the monomeric conformations of the gp41 TMD were simulated in viral-like bilayers. The simulations show that the highly conserved R696, which is found at the center of the TMD, can snorkel towards the lipid-water interface and thus affects the conformation of the TMD. The snorkeling of the charged R696 pulls water into the membrane and prevents the TMD to rotate around its principal axis. After simulating the monomeric structure, the published C_3 symmetric trimeric structure of gp41 TMD was simulated in the same viral-like bilayer, with the C_α atoms of R696 centered in the hydrophobic bilayer core. In both repetitions of the simulation, all three arginine side chains snorkeled towards the cytosolic leaflet of the membrane. This led to the introduction of water molecules into the hydrophobic core of the membrane and ultimately to the dissociation of the C-terminal regions of gp41 TMD. This opening of the structure was permanent and the initial C_3 symmetric structure was not assumed again. New simulations were set up to test if the lack of the membrane-proximal extended region

(MPER) domain or the difference in bilayer composition to the experiments destabilize the trimeric structure. The trimeric gp41 MPER-TMD structure was placed in a simple DMPC model bilayer to mimic bicelles. In all six independent simulations, the C-terminus of the TMDs dissociated after the introduction of water molecules due to R696 snorkeling. The trimers lose their C_3 symmetry after dozens of nanoseconds and equilibrate in either an almost symmetrical conical or unsymmetrical tent-shaped conformation. The MPER region initially resembles two α -helices that are connected by a short unstructured loop. The C-terminal helix of MPER also forms an angle with the TMD. In all simulations, this C-terminal MPER helix disappears and becomes part of an elongated TMD. Also, the conformational rearrangement of the MPER surprisingly recovers the epitopes of broadly neutralizing HIV-1 antibodies (bnAbs). Structures of bnAbs in complex with their epitope are available on the PDB. This epitope is part of the MPER-TMD junction and considers a consecutive α -helix in most cases. Superimposing the bnAbs onto the recovered epitope from simulations shows that the experimentally solved bnAbs can indeed bind to the recovered epitope. Finally, a model of the full-length HIV-1 Env glycoprotein was built to showcase how the bnAb can reach its epitope. This model lacks glycans that would complete the picture, but already establishes a first look at the bound state. These results show that MD simulations can analyze the stability of NMR structures and also recover epitopes of bnAbs to complete the full picture of the bound target.

In this thesis, MD simulations were used to study the dynamics of proteins surrounded by lipid bilayers. In MD simulations the evolution of particle systems is computed over time. The particles obey the rules of a so-called “force field” that defines the interaction energies between bonded and non-bonded atoms. Here, the stretching motions of atomistic bonds, their rotation around angles, and limitation by dihedrals, as well as interactions through hydrophobic and electrostatic forces are taken into account. The time evolution is computed in steps on the order of 10^{-15} s. Given that biological processes happen on the 10^{-9} s to 10^3 s timescale, a lot of MD steps have to be computed. Supercomputers, also called high-performance clusters, are used to calculate the trajectories of MD simulations. These supercomputers are composed of hundreds to thousands of single computing nodes, each containing CPUs and GPUs. MD simulations are calculated with highly optimized software that can make use of the heterogeneous hardware, but the environment also allows to adjust the settings and thus change the speed of the calculations. Inexperienced researchers might not be aware of the best settings to choose, and the defaults provided by the supercomputers are not necessarily the best for every simulation. In this thesis, the software package MDBenchmark was developed to help researchers benchmark their simulations and in this way try out different settings quickly. The software is written in Python and published on PyPI and conda-forge. Extensive documentation can be found online that describes the installation process and helps users to get started. As a proof of concept, MDBenchmark was used to perform scaling studies of 23 representative MD systems. The numbers of message passing interface (MPI) ranks and open multi-processing (OpenMP) threads were tuned, while each system was scaled from 1–10 nodes. On CPU-only nodes, those without any GPUs, the number of MPI ranks should equal to the number

of physical cores for the best performance. In the case of the tested MPCDF cluster cobra, $n_{\text{ranks}} = 40$ for all systems independent of their size. In the presence of GPUs, on mixed CPU-GPU nodes, the number of MPI ranks strongly depends on the system size. A system with 300k atoms achieves the highest performance on $n_{\text{ranks}} = 10$, whereas systems with more than 1 M atoms require the same settings as on CPU-only nodes, such that $n_{\text{ranks}} = 40$ delivers the best results. Hyperthreading does not make a difference on mixed CPU-GPU nodes but is beneficial on CPU-only nodes. These scaling studies revealed that going from one to two mixed CPU-GPU nodes leads to a small increase in performance and that not all resources are fully utilized. To test this, multiple simulations were run on a single mixed CPU-GPU node. The results highlight the opportunity to run up to four independent simulations on the same node and achieve higher overall performance. While each simulation instance produces a shorter trajectory, the total length of all trajectories combined is increased by 1.5 to 4-fold. This scaling study shows that usage of MDBenchmark can help researchers achieve better performance of their simulations without in-depth knowledge of the underlying supercomputer theory.

In summary, this thesis contains the investigations of protein-lipid interactions from the perspectives of different proteins. AHs can sense the lipid composition, but must first fold and embed into the membrane. The oligomeric conformation of viral proteins depends on the correct lipid composition and their structural dynamics can be studied with MD simulations. The calculation efficiency of these simulations can be enhanced with the MDBenchmark software package that was presented in this thesis.

Introduction

Cellular life in its current form could not exist without membranes. The oily film with a width of $\sim 30 \text{ \AA}$ engulfs all living cells and brings them into existence. Without a membrane, the cellular constituents would freely diffuse and mix with the outside environment. It is the actual membrane that holds all the molecular machines and their co-factors together, leading to the emergence of life. Like a modified version of a Russian “Matrjoschka”, more membrane compartments are found inside the main cellular body. Similar to how the outer cellular membrane separates the cellular interiority from the outside world, the cells’ membrane compartments separate their interior from other parts of the cell. This has allowed biological life to grow in complexity. Like a well-oiled factory with different departments in charge of various parts of a product, the cellular compartments all work in symphony to arrive at a selfish outcome: proliferation of their genes [1]. These cellular compartments are called organelles and are highly specialized in form and function. Each organelle can be identified by shape, contents, localization inside of the cell, or contacts formed with other organelles. Although all organelles can be differentiated from one another, they for the most part share the same building block: lipids.

1.1 Structure of lipids

Lipids serve multiple functions in the cellular context. They provide the structural components for membranes, store energy and heat, are used as signaling molecules, establish platforms for protein recruitment, and act as substrates for post-translational protein modifications [2–7]. Their amphipathic nature allows them to be used as building blocks of cellular membranes: they are built from two distinct regions, a polar headgroup and apolar acyl chain. Due to their hydrophobic properties the acyl chains tend to self-assemble, and depending on the type and amount, form micelles, liposomes or planar lipid bilayers [8, 9]. In the biological context liposomes and differently shaped lipid bilayers are of biggest importance. The secret of how lipids come together to form these distinct organelles of varying forms and functions lies in their structural variety.

Lipids share similar structural motifs, but are also very distinct from one another and can be divided into eight different categories depending on their backbone, acyl chain and general structure [10]. In the following, a short overview of the categories relevant for this thesis is given. Fatty acids are composed of a single hydrocarbon chain that is terminated by a carboxylic acid. They can be found in an esterified form in lipid droplets, which are used by eukaryotic cells for energy storage. This class of lipids is also used as a building block

for another important and well known lipid class: glycerophospholipids (GPLs). GPLs are the most abundant type of lipid found in eukaryotic cells as they are the building blocks of membranes. Their name indicates the three building blocks that compose each GPL (Figure 1.1A). The glycerol backbone is esterified to two fatty acids. These hydrophobic acyl chains vary in their length and degree of saturation. In addition, a phosphoric acid is also esterified to the glycerol. In turn, different headgroups are esterified to the phosphoric acid, changing the chemical variety of the GPLs. Instead of the phosphoric acid a third acid chain can be esterified to the glycerol backbone to produce a glycerolipid. Glycerolipids are primarily used for energy storage. Sphingolipids are more complex than GPLs. They share a common sphingoid base backbone, mostly sphingosine (Figure 1.1B). A single acyl chain is amide-linked to the sphingoid base. Without a headgroup this lipid is called ceramide. Its hydroxyl group can be esterified with a phosphocholine or phosphoethanolamine to produce sphingomyelin. The plasma membrane (PM) is abundant in sphingolipids, where they add to the bilayers mechanical stability. Another lipid class found predominantly in the PM are sterols, who belong to the group of steroids. They have a very distinct shape and structure, compared to GPLs and sphingolipids. They are made up of a four-core ring structure with a carbon chain on the hydrophobic side and a hydroxyl group defining the polar side (Figure 1.1C). Sterols such as cholesterol are used by the cell to control the fluidity of a membrane. When lipids of different classes and species come together, their chemical and physical properties affect the properties of the formed membranes, such as fluidity and shape.

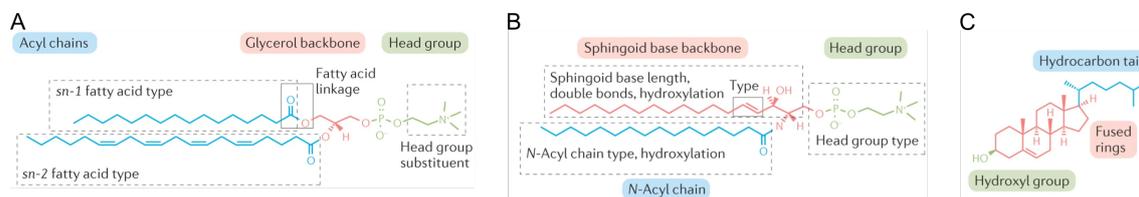


Figure 1.1: Diversity of different lipid species. (A) GPLs are made of three distinct parts: two acyl chains (blue), a glycerol backbone (red) and a head group (green). (B) Sphingolipids consist of an N-acyl chain (blue) connected to a sphingoid base (red) and a head group (green). (C) Cholesterol has a hydrocarbon tail (blue) connected to fused rings (red) with a hydroxyl group (green). Adapted from Takeshi Harayama *et al.* (2018) [11].

1.2 Membrane properties

The esterified headgroup of lipids, as well as their acyl chains influence the physico-chemical properties of membranes. A variety of different headgroups can be esterified to the glycerol backbone of GPLs: phosphatidic acid (PA), phosphatidylethanolamine (PE), phosphatidylserine (PS), phosphatidylcholine (PC), phosphatidylglycerol (PG) and phosphatidylinositol (PI). PA is the simplest GPL with a single hydrogen atom as a headgroup. PE is the second smallest headgroup with an ethanolamine group. The addition of a carboxylic acid produces the PS headgroup. The other two lipid headgroups have either an inositol ring or a glycerol esterified to the phosphate moiety. The mixture of lipids with different headgroups has specific effects on lipid bilayers. PC and PS headgroups tend

to adapt a cylindrical shape, because their headgroup has a similar breadth as the lipids acyl chains (Figure 1.2A). These cylindrical lipids favor flat lipid bilayers. In contrast, the smaller PE and PA headgroups invoke a negative curvature due to their conical shape, where the headgroup occupies less area than the acyl chains (Figure 1.2B). When the headgroup is larger than the acyl chains, as is the case of PI, a positive curvature is induced by the inverted-conical lipid shape (Figure 1.2C). For example, inclusion of PE into PC membranes leads to curvature stress. The curvature stress leads to membrane fission, fusion and budding [12].

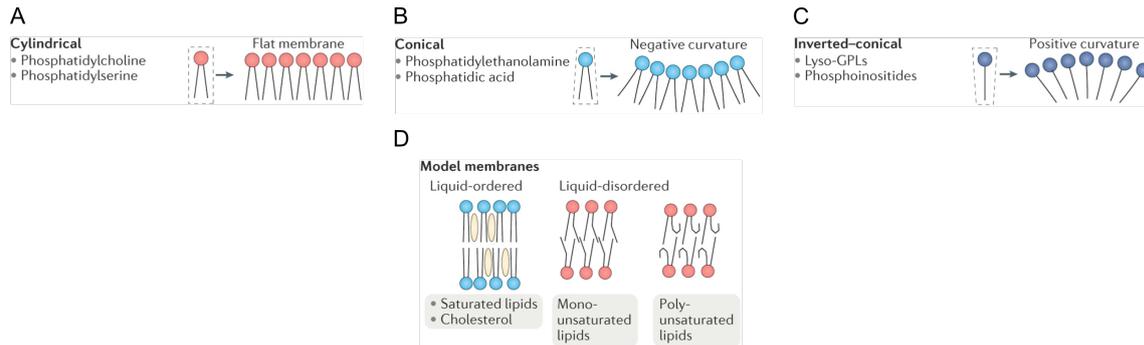


Figure 1.2: Physicochemical properties of membranes induced by lipids. (A) Headgroups of similar breadth as that of acyl chains lead to a cylindrical shape and flat membranes. (B) A headgroup smaller in breadth than the acyl chains produced a conical shaped lipid and negative curvature. (C) Broad headgroups, like those of PI lipids, lead to positive curvature. (D) The level of acyl chain saturation results in liquid-ordered and liquid-disordered domains in membranes. Adapted from Takeshi Harayama *et al.* (2018) [11].

The membrane properties of a lipid bilayer are not only affected by lipid headgroups. The length and saturation of each acyl chain influences the overall membrane properties as well. Long and unsaturated acyl chains, as found for example in sphingolipids, lead to thicker and less fluid membranes. This is owed to the tighter packing of hydrophobic acyl chains and in turn stronger lipid-lipid interactions [6, 13]. In complex membranes with different lipids species, lipids and cholesterol like to segment into different phases depending on their ratio [13]. In membranes with a high amount of lipid packing, liquid-ordered (L_o) phases start to segregate (Figure 1.2D). The PM is said to contain such L_o phases due to its high contents of sphingolipids and cholesterol [14]. These phases were named “lipid rafts”. In contrast, liquid-disordered (L_d) phases contain lipids with a low lipid packing, i.e., found in the endoplasmic reticulum (ER). These two phases can be differentiated in their fluidity. Due to its high abundance in sphingolipids, GPLs with high saturation and long acyl chains, as well as cholesterol, L_o phases have a low fluidity. This affects the diffusion of proteins that move slower through these phases. In contrast, L_d phases contain lipids with unsaturated acyl chains, which make it harder for lipids to pack together tightly, increasing the overall membrane fluidity. As a general rule of thumb, the *sn*-1 fatty acids of GPLs are often saturated and the *sn*-2 fatty acids contain one or multiple double bonds [3]. In the PM acyl chains of GPLs vary in length between 14 and 24 carbons [15]. The N-acyl chains of sphingolipids are saturated and on average longer than those of GPLs [2].

1.3 Membrane compositions in organelles

Cellular organelles are functionally specialized compartments that require certain lipid and protein compositions for their identity. Eukaryotic organelles are characterized by their specific charge distribution and lipid packing. For example, the ER, lipid droplets, nuclear envelope and *cis*-Golgi have a low abundance of anionic charges and a high abundance of unsaturated lipids [13,16]. In contrast, the membranes of the *trans*-Golgi, endosomes and PM have a high abundance of anionic and saturated lipids. The ER and PM form the two extremes of cellular organelles (Figure 1.3). They are connected through the secretory pathway, which is managed by the Golgi apparatus. The Golgi is the cellular hub for lipid and protein sorting, as well as post-translational modifications of different kinds. Lipids are understood to shuttle between the ER and PM through vesicular, non-vesicular and membrane contact site (MCS) mediated transport [9,17,18]. The differences that are seen with the extremes, the ER and PM, provide a platform for lipid-protein interactions tuned for certain proteins. For example, the high PS content of the PM allows targeting of the oncogene K-Ras through electrostatic interactions [19,20]. K-Ras is a relay for signals from the outside environment to the cell nucleus.

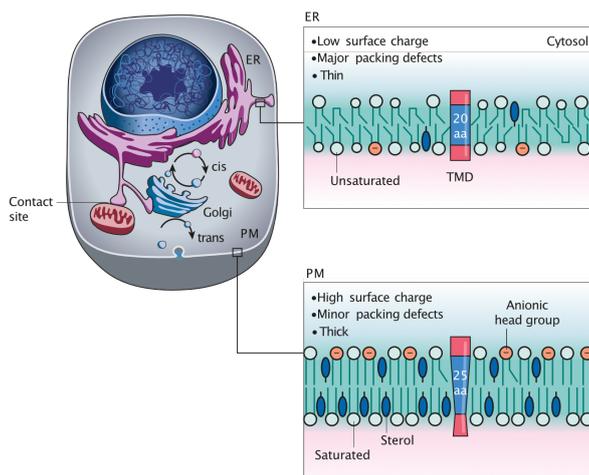


Figure 1.3: The ER and PM form two lipid bilayer extremes. Lipids are for the most part produced at the ER or the *trans*-Golgi. Organellar identity is kept up through lipid segregation. Two different processes, vesicle mediated through the *trans*-Golgi, and non-vesicle mediated through MCSs exchange lipids between organelles. The ER has a low surface charge with many packing defects and a relatively thin membrane, whereas the PM is thicker with many saturated lipids and a high abundance of anionic charged lipids. Adapted from Holthuis *et al.* (2014) [9].

The targeting of proteins to specific membranes is organized by the abundance of lipids in specific organellar membrane leaflets. Lipids are not equally distributed among all organelles and leaflets, but are localized in certain asymmetries [13]. Although PS is synthesized in the ER, it is only found in the luminal facing leaflet [21]. At the PM in contrast, PS is found facing the cytosol [9]. Monounsaturated acyl chains are mostly found at the ER, nuclear envelope and *cis*-Golgi, contributing to the conical shape of the lipid [13]. This physicochemical effect is similar to the contribution of small headgroups to curvature stress. The conical shape of lipids leads to so called “packing defects”. These can be thought of as holes in the lipid-water interface of a membrane, exposing the apolar innards with

hydrophobic patches [22, 23]. Proteins containing the amphipathic lipid packing sensor (ALPS) motif are sensitive to these lipid packing defects. Due to the nature of lipid distribution across organelles, lipid packing defects are predominantly present at the start of the secretory pathway, at the ER, nuclear envelope or *cis*-Golgi [24].

Lipids can be transferred between organelles in different ways. In the vesicular pathways lipids are transported between the ER and PM through the *trans*-Golgi network using COPI and COPII vesicles. COPI vesicles for example are less abundant in cholesterol and sphingolipids, than their parental Golgi membranes [25]. This sorting is caused by the curvature of the membranes [26, 27]. Another type of lipid sorting can be induced by protein-lipid interactions. BAR domains are excluded from budding vesicles and are found in the tubular neck region, connecting the vesicle and parent membrane [28]. The interaction of phosphatidylinositol (4,5)-biphosphate (PI(4,5)P₂) lipids with these BAR domains lowers their diffusion coefficient by ~ 2 -fold and excludes them from the budding vesicle [29]. Sterols are highly abundant in the late secretory pathway and kept at low abundance levels in the ER. Nevertheless, 5% of the ER are made up of sterols that act as regulatory elements for sterol synthesis. Proteins of the sterol regulatory element-binding protein (SREBP) family are activated when sterols levels in the ER deviate too much from the 5% optimum [30]. PC is a cylindrical lipid that forms flat bilayers. Its depletion causes the PE density to increase, leading to curvature stress and in turn the exposure of packing defects. These packing defects recruit the CCT α AH to the ER membrane, which will be touched upon later [31].

1.4 Proteins sensing membrane properties

Various membrane property sensors have evolved in different organisms, allowing the host to react to changes in temperature or nutritional availability. They have evolved to monitor the physical and chemical properties of organelle lipid bilayers [16]. DesK is a well understood membrane property sensor found in *Bacillus subtilis*. It reacts to temperature changes by sensing the bilayer thickness, as a function of its fluidity (Figure 1.4). A lower temperature reduces the bilayer fluidity and results in a thicker membrane. This change in membrane thickness triggers a conformational change in the DesK protein which in turn autophosphorylates and activates itself. This activated version of DesK then activates the transcriptional factor DesR, which upregulates the transcription of a fatty acid desaturase. The change in fluidity again deactivates DesK by autodephosphorylation [32, 33].

As the ER is the main site for protein folding, it also reacts to protein folding stress. To alleviate the aggregation of misfolded proteins at the ER, the organelle is expanded after the unfolded protein response (UPR) has been activated [34, 35]. The UPR consists of three branches: (i) IRE1 signaling through XBP1, (ii) ATF6 and (iii) PERK acting via eIF-2 α . All branches are activated when misfolded proteins accumulate in the ER, but only the first one will be explained in more detail. The activation of the UPR downregulates the translation, increase chaperone expression and/or induce cell death through apoptosis [36]. Ire1 contains a single spanning transmembrane domain and is monomeric when inactive (Figure 1.5). Its N-terminal unfolded protein sensor domain resides in the lumen of the

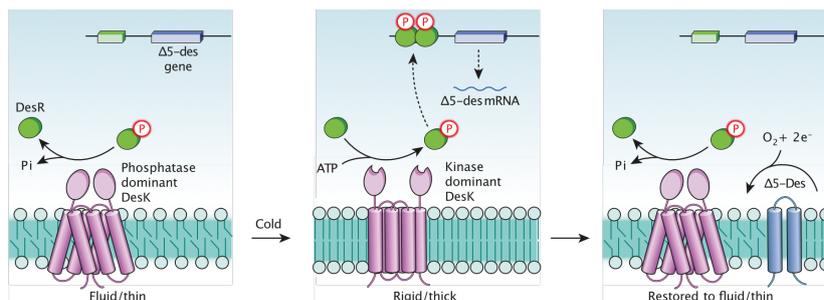


Figure 1.4: Regulation of membrane saturation by DesK/DesR in *Bacillus subtilis*. The membrane bound DesK senses the fluidity of membranes. Under physiological conditions its phosphatase function dominates and continuously dephosphorylates DesR. Upon activation through decreased membrane fluidity and increased rigidity, the activity of DesK shifts towards its kinase function, activating DesR. In turn, DesR acts as a transcription factor of $\Delta 5$ -Des, which restores membrane fluidity by oxidation of lipid acyl chains. DesK shifts back into its deactivated phosphatase dominant form and again dephosphorylates DesR. Adapted from Holthuis *et al.* (2014) [9].

ER, where misfolded proteins can accumulate after their translation. The protein forms higher oligomers upon its activation, when unfolded proteins are bound and the chaperones Kar2/BiP dissociate [37, 38]. Upon oligomerization, its cytosolic endoribonuclease domain is activated, leading to the splicing of the XBP1/Hac1 mRNA [39]. The translocated product of this unconventionally spliced mRNA then mediates different downstream events of the UPR [39]. Although initially described as a protein folding sensor, Ire1 was found to be sensitive to changes in the lipid environment as well. Indeed, when the unfolded protein-sensing domain is removed, the remaining protein maintains its ability to activate the UPR [40, 41]. A recent study has shown that activation of IRE1 by lipids requires a highly conserved AH [42]. This AH senses the elasticity of the membrane, which is a function of its lipid saturation.

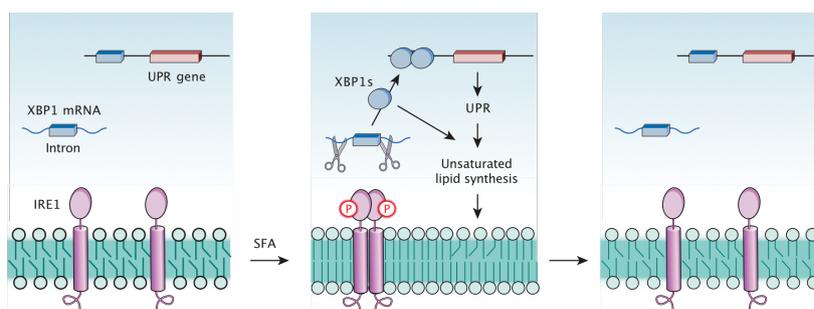


Figure 1.5: Ire1 senses the membrane elasticity as a function of its lipid saturation. Under physiological conditions Ire1 is found in a monomeric conformation, with its N-terminal unfolded protein sensor domain bound to the chaperones Kar2/BiP. Upon activation the protein oligomerizes, activating its endoribonuclease domain. Subsequent unconventional splicing of XBP1 mRNA leads to the translation of XBP1, which in turn upregulates the transcription of UPR. Ultimately unsaturated lipids are produced to ease the stress on the ER membrane. Ire1 can be activated by supplementing saturated fatty acids (SFAs). Adapted from Holthuis *et al.* (2014) [9].

CCT α is a membrane sensor that controls the synthesis of PC. As mentioned earlier, PC tends to form flat lipid bilayers due to its cylindrical shape. As it is one of the main building blocks of membranes, its biogenesis is required upon membrane expansion. CCT α is mostly soluble and shuttles between its soluble and membrane bound form (Figure 1.6).

When sufficient PC is present in the ER membrane, no “packing defects” between the lipids are present and CCT α is found in its completely soluble form, where it is auto-inhibited by its C-terminal M-domain [43]. Upon depletion of PC lipids, and a density increase of conical lipids, the abundance of packing defects increases on the lipid-water interface [22, 23]. This allows the M-domain to tether itself to the membrane and form an AH [44, 45]. In its active form CCT α then catalyzes the formation of cytidine diphosphate (CDP)-choline, a crucial step in the PC synthesis. Intriguingly, the activity of CCT α is regulated through a feedback loop by its downstream product. Here, the general shape of the lipids is crucial for sensing by CCT α .

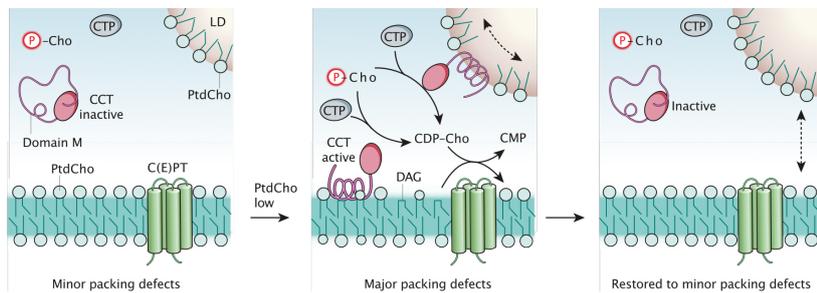


Figure 1.6: CCT α regulates the synthesis of PC. The cylindrical shape of PC does not allow for “packing defects” in the membrane, forcing CCT α to stay completely soluble and deactivated. When PC is depleted, the density of other lipids such as PE increases, leading to more packing defects on the cytosolic side of the ER. The M-domain of CCT α undergoes a conformational change, tethers to the membrane and forms an AH. This activates CCT α and allows the production of CDP-choline. Adapted from Holthuis *et al.* (2014) [9].

Another membrane sensor is Ole1, the sole fatty acid desaturase in yeast. It catalyzes the oxidation of the $\Delta 9$ bond of GPLs. Ole1 is activated by the membrane-bound transcription factors Mga2 and Spt23, whose cytosolic domains are able to shuttle into the nucleus after cleavage and regulate the expression of Ole1. This cleavage reaction is dependent on the degree of membrane saturation and thus fluidity [46]. A recent study combining both computational and molecular biological experiments has revealed that Mga2 is indeed activated by its rotational flexibility, which is ultimately affected by the degree of saturation of the lipids in a lipid bilayer [47].

The sensors described above can be categorized into three distinct classes [48]. Sensors in class I recognize the surface properties of membranes. The packing density of surface molecules can be recognized by proteins containing the ALPS motif [49]. In addition, the surface charge is sensed by class I sensors. Class II sensors instead are able to sense the lipid packing density with their amino acid side chains, i.e., Mga2 uses its tryptophan residue to assess the degree of lipid saturation [47]. Similar to class II sensors, proteins of the class III family use a transmembrane domain (TMD) to sense the membrane properties. The TMD acts across the width of the bilayer by stretching, squeezing or bending. All three of these mechanic effects are affected by the bending rigidity and thickness of the membrane [50].

1.5 Biosynthesis of glycerophospholipids

The *de novo* synthesis of GPLs is tightly regulated. The expansion of the ER, upon protein folding stress, requires the synthesis of new lipids. These lipids are immediately built into the membrane and extend the volume of the organelle. As in any process, the product can only be produced when sufficient amounts of reagent are available. Producing and maintaining the machinery required to drive the reaction from reagent to product requires energy. It is not feasible to upkeep the machinery responsible for a downstream product, when one of the upstream reagents is not available. In this case, one is interested to regulate or entirely shutdown the downstream process and save energy, until sufficient amount of the upstream reagent are available. In a cellular context, PA is the main precursor for all GPLs. It is used up in various pathways to produce other lipids, among others PC, PE or PI. As the *de novo* synthesis of GPLs from PA is a multi-step process, cells require to know when PA is depleted to stop producing the machinery for downstream reactions. In *S. cerevisiae*, the abundance of PA is observed by the protein over-production of inositol (Opi1). Its name stems from the observation that Opi1 deletion mutants produce the sugar inositol to such extent that they start to secrete it into the outside environment.

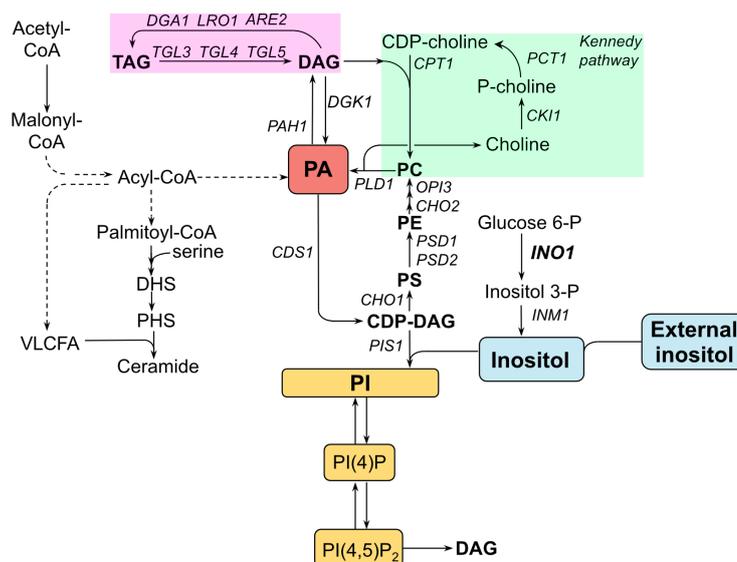


Figure 1.7: PA is at the center of phospholipid biosynthesis. It is produced in a multistep reaction from Acyl-CoA. The enzyme Cds1 converts PA to cytidine diphosphate-diacylglycerol (CDP-DAG). From here PI, PS, PE, PC or PG (not pictured) can be produced. PC and PE can also be generated without CDP-DAG through the Kennedy pathway and exogenous choline/ethanolamine (green box). The steps for PE are not pictured. Inositol can be produced by conversion of glucose-6-phosphate or through exogenous uptake (blue boxes). PA can be transformed to diacylglycerol (DAG) and subsequently to triacylglycerol (TAG) for lipid storage in lipid droplets (magenta box). Solid arrows show a single reaction, whereas dashed arrows indicate a multistep process. Adapted from Henry *et al.* (2014) [51]. This figure is licensed under a CC BY-NC-SA license.

All GPLs are synthesized from the single chained fatty acid lyso-PA, which is metabolized into PA. Now a decision is made whether to metabolize PA into CDP-DAG and produce new membranes or rather into DAG for lipid storage. In the CDP-DAG pathway the enzyme Cds1 converts PA into CDP-DAG [52]. CDP-DAG is then converted to PC in five steps. First, PS is generated by converting CDP-DAG through the Cho1 enzyme [53].

Next, PS has to be shuttled into the inner mitochondrial membrane to be converted to PE by Psd1, or into the Golgi where Psd2 resides [54, 55]. After this conversion, PE has to be shuttled back into the ER for the next three methylation reactions that convert it to PC. The first is catalyzed by Cho2, whereas the Opi3 protein is responsible for the last two methylations [56]. If PA is not present in sufficient amounts, then *S. cerevisiae* is unable to generate GPLs through the CDP-DAG pathway. Instead, it can use the Kennedy pathway to synthesize both PC and PE from exogenously supplied choline and ethanolamine, respectively [57]. The kinases Eki1 and Cki1 produce ethanolamine-phosphate and choline-phosphate [58, 59], respectively, which in turn are used up by Ect1 and Pct1 to generate CDP-ethanolamine and CDP-choline [60, 61]. At the final step DAG, in combination with CDP-ethanolamine or CDP-choline is used by Ept1 or Cpt1 to produce either PE or PC [62, 63]. Interestingly, both pathways are used even in the absence of exogenous ethanolamine and choline [64]. CDP-DAG can also be used to produce PI with the Pis1 enzyme [65]. PI is shuttled from the ER into the PM to generate phosphatidylinositol 4-phosphate (PI4P) and PI(4,5)P₂. The sugar inositol plays a major role in the biogenesis of phospholipids in yeast. Inositol can be produced from glucose-6-phosphate that is obtained in the glycolysis pathway [66]. The sugar is essential for the synthesis of PI, sphingolipids and also glycosylphosphatidylinositol (GPI) anchors [67–69]. As mentioned above, PI competes with PS for the pool of available CDP-DAG molecules. When exogenous inositol is taken up, the production of PI is increased [70]. This in turn leads to the depletion of PA and TAG levels, as the latter are transformed back into DAG and ultimately PA. PA can also be converted into DAG by the Pah1 enzyme if membrane expansion is not required [71]. DAG is then converted into TAG and transferred to lipid droplets. Here, it can be converted back to DAG, monoacylglycerol (MAG) or into free fatty acids [72, 73]. The lipid droplets provide storage of GPL building blocks in the form of TAG and steryl esters [74].

Different genes are tightly regulated by the presence and absence of inositol. In the absence of inositol, most *de novo* GPL synthesis is active. All of these genes contain the 5'-C ATGTGAAT-3' consensus [75]. This consensus is known as the inositol sensitive upstream activating element (UAS_{INO}) [76]. The two transcription factors Ino2 and Ino4 form a heterodimer and bind the UAS_{INO} to induce the transcription of genes under this promoter control [77]. *S. cerevisiae* is auxotrophic for inositol and can produce it autonomously, if it is not present in the environment. In this case, the transcription of the *INO1* gene is upregulated. Upon addition of external inositol yeast cells were shown to inhibit their inositol production, pointing to a downregulation of gene transcription by inositol [78]. In opi1⁻ cells, when the Opi1 protein is absent, the cell maintains its production of inositol independent of its abundance. This has led to the name of Opi1 [79, 80]. As mentioned previously, PA levels are depleted by production of PI when inositol is supplemented. Opi1 was shown to contain a “PA-binding motif” that is responsible for the recognition of PA lipids [70]. Opi1 is bound to the ER when PA is present, but dissociates from the membrane on its depletion (Figure 1.8). It then shuttles into the nucleus to interact with Ino2 [81]. This interaction leads to the downregulation of almost all enzymes involved in the *de novo*

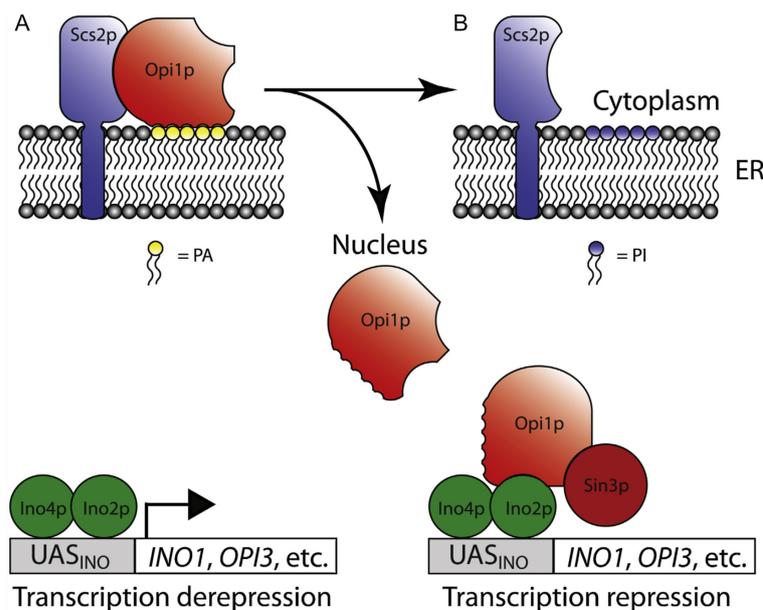


Figure 1.8: Opi1 senses PA at the ER membrane and downregulates GPL biosynthesis in the nucleus. (A) In the presence of PA at the ER membrane, Opi1 recognizes the charged headgroups and is retained at the cytosolic leaflet, interacting with Scs2. PA acts as a precursor of GPLs and the transcription of enzymes involved in the *de novo* synthesis of lipids commences in the nucleus. (B) Upon PA depletion, by addition of inositol and production of PI, Opi1 shuttles into the nucleus and interacts with Ino2 which causes a transcription repression, downregulating the *de novo* synthesis of GPLs. Adapted from Henry *et al.* (2014) [51]. This figure is licensed under a CC BY-NC-SA license.

synthesis of GPLs, as all of them are under control of the UAS_{INO} promoter.

1.6 Different tastes of amphipathic helices

AHs are a versatile structural motif that is used for various functional roles. All AHs are defined by three key parameters: (i) size of hydrophobic residues and their density per turn, (ii) the nature, charge and distribution of polar residues and (iii) the length of the AH. Their functions range from bilayer deformation, sensing membrane curvature [82], recognition of specific lipids, coating lipid droplets [83], and protecting membranes from stress [84]. Hydrophobic residues are distributed at every third or fourth ($N + 3 / N + 4$) position along an AH. Their hydrophobic face allows them to form even in stably folded proteins and cover the hydrophobic inside [85]. Most recently and prominently, they are found to negotiate the interaction between soluble proteins and membranes, and also act as sensors for membrane embedded proteins, as described earlier [42, 86].

Pex11 is a highly conserved protein that is found in the luminal leaflet of peroxisomes. Its structure is predicted to contain two TMDs, with an N-terminal AH (Figure 1.9A). Its apolar face is rich in bulky hydrophobic residues, whereas its hydrophilic face is lined with polar amino acids. Upon membrane binding the AH of Pex11 induces tubulation, followed by fission that is catalyzed by Dnm1 [87]. The expression level of Pex11 modulates the abundance of peroxisomes in cells.

Arf1 is a G protein involved in the vesicular trafficking of cellular contents. Its hydrophobic face is highly tuned, as each of the three turns contains either a combination of

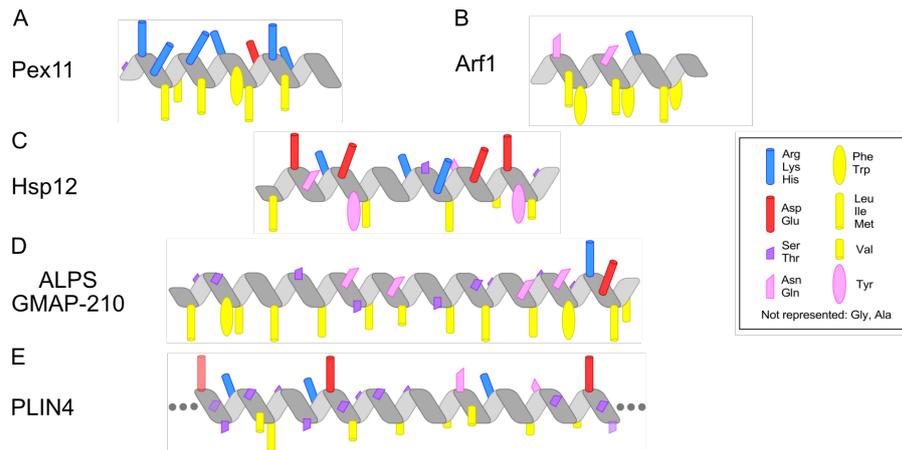


Figure 1.9: Different types of AHs. (A) Pex11 is a peroxisomal protein found inside of peroxisomes. It initiates the tubulation of peroxisomal membranes. (B) The Arf1 AH is folded both in solution and when bound to membranes. Proteins of its family are responsible for various functions, i.e., recruitment of proteins for the formation of COPI vesicles. (C) The heat shock protein Hsp12 stabilizes the PM of yeast cells upon sudden increases of temperature. (D) Proteins with an ALPS motif are insensitive for charged membranes, but are very specific for highly curved membranes. (E) Proteins of the perilipin family coat the outside of lipid droplets and are able to replace their monolayer of lipids. They are characterized by their extraordinary length. Adapted from Giménez-Andrés *et al.* (2018) [49].

Leu-Phe or Ile-Phe amino acids (Figure 1.9B). The AH is folded even in solution, but not necessarily embedded in the membrane. In its inactive form, Arf1 is bound to a guanosine diphosphate and the AH buries its hydrophobic face inside of the soluble protein. When a nucleotide exchange factor replaces the guanosine diphosphate with a guanosine triphosphate, the AH does not maintain the contact with the soluble protein and instead embeds into a lipid bilayer [88]. The hydrophobic face of Arf1 is tuned to such extent that it is able to bind to membranes of varying degrees of saturation and charges [89,90]. This membrane insensitivity allows Arf1 and proteins of the ARF family to be involved in different cellular functions, such as COPI vesicle formation, adaptation of membrane lipid composition and also the formation of viral membranes [91]. It has also been shown to regulate the activity of PIP5K at Golgi membranes [92,93].

Heat shock proteins are expressed in response to a sudden change in temperature. In yeast, Hsp12 is used to stabilize the fluidity of the plasma membrane when temperatures rise above a critical level [84] (Figure 1.9C). Hsp12 is a small protein of only 109 amino acids, but contains a total of four AHs that are unfolded in solution. They fold upon heat stress and independently bind to the plasma membrane [84,94].

The ALPS motif combines a special family of AHs. They are characterized by their polar, but almost neutral hydrophilic face (Figure 1.9D). This makes proteins with an ALPS motif insensitive to charged membranes. Instead, their extended hydrophobic face provides a sensitivity to highly curved membranes [95]. In contrast to the AH of Arf1, the ALPS protein GMAP-210 has only one hydrophobic residue per turn on its hydrophobic face. This feature tunes it to sense the curvature and bulk membrane properties of membranes between the ER and Golgi [96].

Perilipin coats lipid droplets with its unusually long amphipathic helix. Indeed, the

AH of perilipin-4 contains 33 amino acid repeats and spans almost 1000 amino acids in total [97] (Figure 1.9E). Its hydrophobic face is perfectly optimized for lipid droplets, which are composed of neutral lipids at their core, coated by a monolayer of lipids and proteins [90]. The hydrophobic face has only few bulky residues and thus allows the protein to interact directly with the neutral core of lipid droplets, replacing the phospholipid monolayer [97].

These different modes of interaction show how AHs can sense, among others, bulk membrane properties and recognize specific lipid types. A detailed view of the mechanism of sensing is currently missing.

1.7 Viral replication in host organelles

Viruses often use the ER membrane as a platform to set up their replication machinery. To do this, they alter the fatty acid metabolism in the host [98]. Lipid rafts are used by various pathogens including HIV-1 in different stages of their lifecycle. For example, during the infection, immune evasion and the exit of assembled HIV-1 virions [99–101]. The entry of HIV-1 can be inhibited by alteration of GPL synthesis and the abundance of cholesterol [102, 103]. Viruses are internalized after interacting with the receptors in the hosts plasma membrane. They use the hosts cellular machinery to produce all building blocks required for another round of productive infections [104]. The successful infection of a virus requires the presence of cellular organelles [105, 106]. If the virus is unsuccessful and is trapped in an unsupported organelle or is degraded inside a cellular compartment, the infection is considered to be “nonproductive”. Only if the viral machinery is able to generate more viral particles the infection event is called “productive”. Viruses have evolved to manipulate the host immune response and hide in plain sight. They use the endoplasmic reticulum-associated degradation (ERAD)-dependent proteasomal degradation to degrade certain proteins that report on cellular infections. For example, the herpes simplex virus blocks the transporter associated with antigen processing-mediated peptide transport [107]. Ultimately, major histocompatibility complex class I is unable to present viral peptides on the cellular membrane, failing to report a viral infection. Instead, HIV-1 uses its Vpu protein to induce degradation of CD4 receptors through ERAD [108, 109]. The downregulation of the CD4 receptor, which HIV-1 interacts with before entering the host, plays an important role in the stable infection of HIV-1. This downregulation helps HIV-1 evade the cellular immune response in different ways. By removing CD4 from the cellular surface, further infections of the same cell, so called “superinfections” can be mitigated [110]. In addition, the longer retention of CD4 in the ER before its degradation allows for interactions with the synthesized Env precursor [108].

HIV-1 belongs to the *Retroviridae* family [111]. Viruses of this family carry their genome as single stranded RNA and first reverse-transcribe it to produce DNA, out of which the cellular polymerases produce mRNA. The HIV-1 genome contains a total of nine open reading frames that encode 15 proteins, out of which the precursor polyproteins Gag, GagPol and Env are the major components [112]. All three of them are proteolytically cleaved during a certain point in the lifecycle of the virus. Gag is composed of the four proteins matrix, capsid, nucleocapsid and p6, and is cleaved after budding to initiate the finalization

of viral maturation. Env, also known as gp160, is cleaved into gp120 and gp41, producing two separate parts that interact through non-covalent bonds. These two glycoproteins, gp120 and gp41, are essential for binding to host cells and viral fusion. GagPol contains the three enzymes protease, reverse transcriptase and integrase. These GagPol proteins are essential for viral function and localize inside the virion [113]. Six additional proteins are encoded by the HIV-1 genome: two regulatory proteins, Tat and Rev, as well as four accessory proteins Vif, Vpr, Vpu and Nef. Vpu is involved in the assembly of virions, whereas the other three localize inside the finished viral particles [114].

To enter a host cell, HIV-1 must first bind a receptor on the host cell membrane. CD4 is the surface protein that HIV-1 interacts with for cell entry. It belongs to the Immunoglobulin superfamily and is found on the cellular membranes of macrophages and T cells [115]. In addition to CD4, HIV-1 must interact with either of two co-receptors, CCR5 or CXCR4, to enter its host cell [116–118]. These two chemokine receptors belong to the G-protein coupled receptor superfamily, identified by their characteristic seven transmembrane domains. These receptors are commonly used in signal transduction pathways and HIV-1 uses them to prepare cells for virion production [119]. HIV-1 has evolved to hide its co-receptor interacting surfaces until gp120 interacts with CD4. By hiding its functional surfaces from antibody recognition the virus is able to evade the immune system of its host [120].

The replication cycle of HIV-1 can be divided into the early and late phase. All events leading up to the integration of viral DNA into the host genome are part of the early phase. The late phase includes the production of all viral machinery, trafficking through the host cell, assembly of new virions, as well as the budding and maturation of viral particles (Figure 1.10). After entering the host cell, the virus particle must translate its genome, replicate its proteins, assemble a particle and finally exit the host cell [104]. The single-stranded RNA genome of HIV-1 is first reverse-transcribed to DNA by the reverse transcriptase enzyme that HIV-1 carries inside its virion. The newly synthesized viral DNA is translocated into the nucleus through the nuclear pore using a pre-integration complex. This pre-integration complex is composed of the viral proteins Vpr, matrix and integrase. These three proteins all contain the nuclear localization sequence [121–123]. This allows them to enter the nucleus through the nuclear pore complex after interacting with nuclear transport receptors. Nuclear localization sequences are also found in the nucleoproteins of the influenza virus [124]. Inside the nucleus the viral DNA is integrated into the hosts genome using the viral integrase enzyme [125]. This step finalizes the early phase of HIV-1 infection.

HIV-1 uses the cellular machinery and organelles as a platform to produce new virions. The late phase of the infection starts with the transcription of the viral genome into mRNA by cellular polymerases. The mRNAs are subsequently translated by cellular ribosomes into their respective precursor proteins. The Gag polyprotein is responsible for the assembly, budding and release of immature viral particles, and is synthesized in the cytosol by ribosomes. Gag is a 55 kDa precursor protein, whereas GagPol is 160 kDa in size. Gag is recruited to the inner leaflet of the PM by basic residues found on an α -helix of the matrix

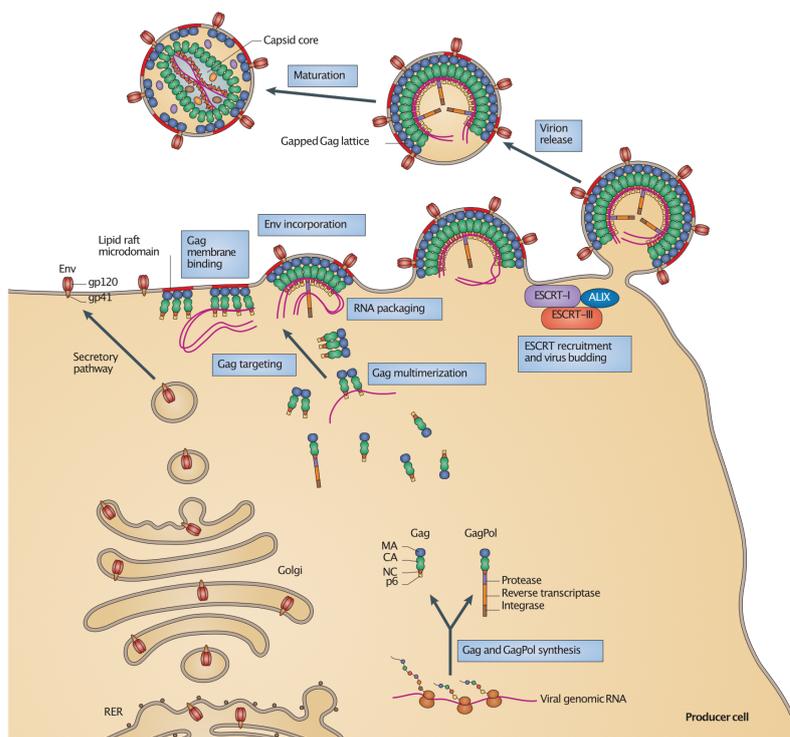


Figure 1.10: Late phase of HIV-1 infection. Env glycoprotein is translated at the rough ER and moves through the Golgi and secretory pathway towards the plasma membrane. It is heavily glycosylated while passing from the ER to the PM. The precursor proteins Gag and GagPol are translated in the cytosol and migrate to the PM. Gag is recruited to the membrane through the basic residues of its matrix (MA) domain. Env is incorporated into the Gag lattice. Two copies of single stranded viral RNA are packed into the budding virion by interactions with the nucleocapsid (NC). The endosomal sorting complexes required for transport (ESCRT) machinery is recruited through the p6 domain of Gag, leading to complete viral budding and release. To finalize the maturation, the protease of GagPol cleaves Gag into its constituents, leading to a rearrangement in the virion core, ultimately forming the final viral capsid by the capsid (CA) protein. Adapted from Freed (2015) [126].

domain [127]. More specifically, the basic residues were shown to sense the abundance of PI(4,5)P₂ on the inner leaflet of the PM [128]. Binding of PI(4,5)P₂ to the matrix domain is thought to trigger a conformational change exposing its N-terminal myristic acid moiety which subsequently tethers the protein at the PM by inserting into the lipid bilayer. The binding of Gag to the PM leads to the formation of lipid rafts [101, 129]. The Env glycoprotein is synthesized as gp160 at the rough ER by ER-resident ribosomes [130]. Its TMD acts as a hydrophobic stop-transfer signal and prohibits gp160 to be released into the ER lumen [131]. Instead, the extracellular domain of Env is found inside the ER lumen and its C-terminal domain (CTD) extends into the cytoplasm. Immediately after translation, the Env glycoprotein is heavily glycosylated at the ER [132, 133]. While shuttling through the *trans*-Golgi it also acquires high-mannose oligosaccharides. In fact, half of Env's mass is due to N-linked glycans [134]. Env is thought to oligomerize at the ER and move through the ER-Golgi network as a trimer, but other oligomeric forms such as dimers and tetramers were also observed [135, 136]. While shuttling through the Golgi, Env is proteolytically cleaved by furin and furin-like proteases, producing the two glycoproteins gp120 and gp41 [137]. They remain stably associated through non-covalent interactions

after cleavage. These glycoproteins move through the secretory pathway toward the plasma membrane. More specifically, HIV-1 gp41 has an unusually long TMD that is sorted towards the thicker PM [138]. The precursor Gag polyprotein is also targeted to the PM through its matrix domain. Four distinct models try to explain how Gag and Env co-localize during the assembly of nascent virions (Figure 1.11).

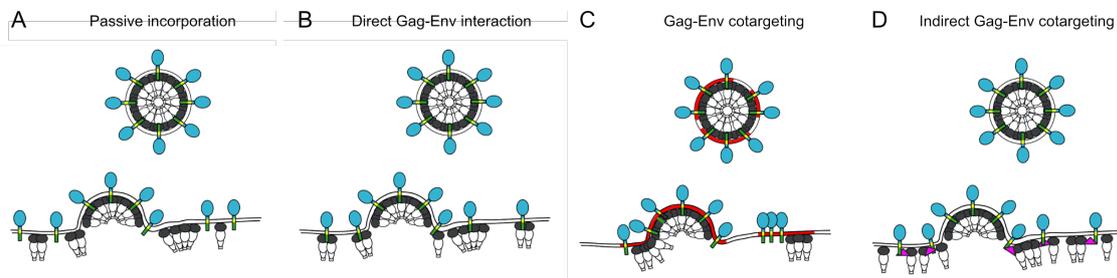


Figure 1.11: HIV-1 Env incorporation into budding viral particles. Four models describe the incorporation of the Env glycoprotein into budding viral particles. (A) In the “passive incorporation model” Gag and Env accumulate at the budding viral particle by chance. (B) Direct interactions lead to enrichment of Gag around Env in the “direct Gag-Env interaction model”. (C) Lipid rafts provide a platform for both Gag and Env, called the “Gag-Env cotargeting model”. (D) Auxiliary proteins in the host cell mediate the interaction between Gag and Env in the “indirect Gag-Env cotargeting model”. Gag is depicted in oval shapes with a black top. Env is shown as blue oval with a green and yellow membrane anchor. Lipid rafts are shown as red regions in the membrane. Auxiliary host proteins are shown as magenta triangles. Adapted from Checkley *et al.* (2011) [130].

In the “passive incorporation model” both Env and Gag come together by pure chance (Figure 1.11A). This model is supported by the observation that retroviruses are promiscuous and tend to incorporate host membrane proteins into virions [139]. In the “direct Gag-Env interaction model” the CTD of gp41 is thought to interact directly with Gag promoting a close proximity of both proteins at the PM (Figure 1.11B) [140, 141]. Lipid rafts might form a platform that helps to associate Gag and Env in an indirect manner, as proposed in the “Gag-Env cotargeting model” (Figure 1.11C) [142]. The fourth model of Env and Gag co-localization is the “indirect Gag-Env interaction model” (Figure 1.11D). Similar to the previous lipid raft model, a cellular protein could negotiate the interaction between the two proteins. This was shown for AP-1 and AP-2, two proteins involved in the trafficking and sorting of proteins through the secretory pathways [143]. After Env incorporation into the lattice formed by the precursor Gag polyprotein, RNA is packed into the nascent budding virion through interactions with the nucleocapsid domain (Figure 1.10). Two copies of the viral RNA are packed into every HIV-1 virion [144]. GagPol is also incorporated into the Gag lattice with its matrix and capsid domains. After the viral particle is assembled it must undergo a fission event to be released from the hosts PM. Again, HIV-1 uses the cellular machinery, more specifically endosomal sorting complexes required for transport (ESCRT), to release its virion [145]. After the budding of the viral particle has completed, the HIV-1 protease, an aspartyl protease, processes both Gag and GagPol. This ultimately leads to the rearrangement and morphological change of the capsid core inside the virion (Figure 1.10) [146]. This rearrangement finalizes the virus maturation and prepares the particle for a new round of infections.

Conformational dynamics of the HIV-1 glycoproteins could provide more information

of how the virus uses the membrane-anchored domain to infect its host and how the virus evades the immune system.

1.8 Molecular dynamics

A laboratory experiment allows the researcher to measure observable from a sample of interest. These experiments are subject to several limitations. First, in a classical biochemical experiment the properties of an ensemble of molecules are measured. These data are useful for statistical processes, such as the turnover of an enzymatic reaction. Second, the sample preparation is a labor intense process with a low yield. Failure of proper experimental handling can result in loss of the precious sample, forcing the experimentalist to redo the preparation steps. Third, more often than not it is helpful to know the structure of proteins of interest. Various imaging techniques such as X-ray crystallography, cryogenic electron microscopy and nuclear magnetic resonance (NMR) have been developed over the last century. Classically they were used to solve static structures of various proteins, often locked in a certain conformation to discourage the structural flexibility. More recent advances have allowed electron microscopy specialists to reassemble protein dynamics from their measurements [147]. Dubbed as the “computational microscope”, molecular dynamics (MD) can be used to study a multitude of processes both in atomistic detail and in more coarse grained representations [148, 149]. MD simulations bridge the gap between observations from biochemical assays and static structures solved using imaging techniques. The concepts of MD are explained in the following sections.

1.8.1 Force fields

The interaction between molecules is mandated by their interaction energy. The forces that act on each atom in a MD simulation are defined by a so-called “force field”. Such a force field is defined in terms of a functional form that is used to calculate the potential energy U for every atom in a system. Different factors contribute to the potential energy of a force field and are summed up by the bonded and non-bonded forces. The parameters of a force field are defined to reproduce the observations that are obtained from experiments. All atoms of a molecule are connected through covalent bonds. These bonds can undergo compression and stretch motions, atoms can rotate around the bonds by some angle, and molecules of at least four atoms can be described by a dihedral angle that is formed by the two planes defined by three atoms each. These three contributions account for the bonded forces of the potential energy. The bond between two atoms can be thought of being a spring. It can be approximated by Hooke’s law with a harmonic spring potential as a function of the bond length r :

$$U_{\text{bonded}} = U(r) = k_r(r - r_0)^2 \quad (1.1)$$

where k_r is the force constant and r_0 is the equilibrium distance of the two atoms. Similarly, the bond angle ϕ of three atoms can also be described by a harmonic potential around an

equilibrium angle ϕ_0 :

$$U_{\text{angle}} = U(\phi) = k_\phi(\phi - \phi_0)^2 \quad (1.2)$$

As mentioned above, four atoms connected in sequence can form two planes. These are defined by two sets of three atoms each and intersect at an angle ϕ . This angle describes the rotation of the middle bond between atoms 2 and 3, numbering all atoms 1–4. The overlap of the two outer atoms (1 and 4) leads to a potential energy that is defined by the molecules conformation. The energy can be described with a periodic potential:

$$U_{\text{dihedral}} = U(\phi) = k_\phi(1 + \cos(n\phi - \delta)) \quad (1.3)$$

with k_ϕ being the energy required to rotate the central bond, the periodicity n and the offset δ .

Non-bonded interactions can emerge between atoms of the same molecule (intramolecular) or between atoms of two separate molecules (intermolecular). The intramolecular interactions are modeled with the Lennard-Jones potential. Its functional form combines the Pauli repulsion, the force disallowing two atoms of being too close, and the London dispersion attraction, two non-polar atoms interacting through weak attractions. The Lennard-Jones potential U_{LJ} is generally defined as

$$U_{\text{LJ}} = U(r) = 4\varepsilon \left[\left(\frac{\sigma}{r} \right)^{12} - \left(\frac{\sigma}{r} \right)^6 \right] \quad (1.4)$$

where ε is the depth of the potential well, σ is the radius of the atom and r is the distance between two interacting atoms. For very close distances the two atoms repel each other and they do not interact at big distances. Only at a moderate distance do the two atoms attract each other. The minimum of a Lennard-Jones potential is found at $r = 2^{\frac{1}{6}}\sigma$ with a potential energy of $U_{\text{LJ}} = -\varepsilon$.

The intermolecular interactions of two distinct molecules can be modeled using Coulomb's law:

$$U_{\text{Coulomb}} = U(r) = \frac{q_A q_B}{4\pi\varepsilon r} \quad (1.5)$$

with q_A and q_B being the charges of atoms A and B, r their distance and the dielectric constant ε .

The kinetic energy U is finally defined as a sum of all the parts described above:

$$U = U_{\text{bonded}} + U_{\text{angle}} + U_{\text{dihedral}} + U_{\text{LJ}} + U_{\text{Coulomb}} \quad (1.6)$$

1.8.2 Equations of motion

Newton's laws of motion have established a framework to calculate the velocity and acceleration of a particle as a function of time. MD simulations make use of these equations of motion to describe the behavior of physical systems. The evolution of particles positions over time considering the forces that act upon each particle, as described above, are calculated in small time increments usually on the order of femtoseconds. Newton's equations of motions can be summarized with the following definitions. The change of the instantaneous position $\mathbf{r} = \mathbf{r}(t)$ over some small period of time dt can be described as

$$\mathbf{v} = \frac{d\mathbf{r}}{dt} \quad (1.7)$$

The velocity \mathbf{v} of a particle can also change with time, described by the acceleration \mathbf{a} :

$$\mathbf{a} = \frac{d\mathbf{v}}{dt} = \frac{d^2\mathbf{r}}{dt^2} \quad (1.8)$$

To compute the time evolution of all particles, one must also take into the account the force that acts on each particle at an instantaneous time. The force \mathbf{F} that acts on a particle with mass m and acceleration \mathbf{a} is defined by Newton's second law as $\mathbf{F} = m\mathbf{a}$. A Taylor series approximation can be used to derive the Verlet algorithm, which is commonly used to compute the trajectories of particles in a MD simulation. The position of a particle at time $t + \Delta t$ is approximated as follows:

$$\mathbf{r}(t + \Delta t) \approx \mathbf{r}(t) + \frac{d\mathbf{r}(t)}{dt}\Delta t + \frac{1}{2}\frac{d^2\mathbf{r}(t)}{dt^2}\Delta t^2 + \frac{1}{6}\frac{d^3\mathbf{r}(t)}{dt^3}\Delta t^3 + \mathcal{O}(4) \quad (1.9)$$

Similarly, a Taylor series can be evolved at time $t - \Delta t$:

$$\mathbf{r}(t - \Delta t) \approx \mathbf{r}(t) - \frac{d\mathbf{r}(t)}{dt}\Delta t + \frac{1}{2}\frac{d^2\mathbf{r}(t)}{dt^2}\Delta t^2 - \frac{1}{6}\frac{d^3\mathbf{r}(t)}{dt^3}\Delta t^3 + \mathcal{O}(4) \quad (1.10)$$

The sum of Equation (1.9) and Equation (1.10) eliminates all terms higher than the second order, while retaining accuracy up until the third order terms:

$$\mathbf{r}(t + \Delta t) + \mathbf{r}(t - \Delta t) = 2\mathbf{r}(t) + \frac{d^2\mathbf{r}(t)}{dt^2}\Delta t^2 \quad (1.11)$$

Solving for $\mathbf{r}(t + \Delta t)$ one arrives at Equation (1.12):

$$\mathbf{r}(t + \Delta t) = 2\mathbf{r}(t) - \mathbf{r}(t - \Delta t) + \frac{d^2\mathbf{r}(t)}{dt^2}\Delta t^2 \quad (1.12)$$

Substitution of the third term by consideration of Equation (1.8) and Newton's second law yields:

$$\mathbf{r}(t + \Delta t) = 2\mathbf{r}(t) - \mathbf{r}(t - \Delta t) + \frac{\mathbf{F}(t)}{m} \Delta t^2 \quad (1.13)$$

It is possible to substitute t for $t + \Delta t$ in Equation (1.10) and again form the sum with Equation (1.9). This allows the computation of the velocity \mathbf{v} and position \mathbf{r} at the same time t , as shown in Equation (1.14) and Equation (1.15), respectively.

$$\mathbf{v}(t + \Delta t) = \mathbf{v}(t) + \frac{\mathbf{F}(t) + \mathbf{F}(t + \Delta t)}{2m} \Delta t \quad (1.14)$$

$$\mathbf{r}(t + \Delta t) = \mathbf{r}(t) + \mathbf{v}(t) \Delta t + \frac{\mathbf{F}(t)}{2m} \Delta t^2 \quad (1.15)$$

1.8.3 Periodic boundary conditions

Simulations require the evaluation of a lot of interactions at tiny timesteps. Even moderately sized systems, such as a gram of water, requires 3×10^{22} molecules. At the time of writing, systems with a size of $10^6 - 10^7$ atoms are considered to be extraordinary big. To overcome current computational limitations, periodic boundary conditions (PBC) are utilized to mimic a seemingly infinite system. In a system with periodic boundary conditions all atoms are contained within the same box, the *unit cell*, as in a system without PBCs. Similarly, all atoms are allowed to move freely and interact with each other. With PBCs enabled, an atom leaving one side of the box will reenter it at the exact other side.

As described earlier, non-bonded interactions of molecules can be described with the Lennard-Jones potential (Equation (1.4)) and Coulomb's law (Equation (1.5)). These long-ranged interactions need special consideration when imposing PBCs on a system. As each molecule interacts with its neighbors, those at the borders of the unit cell will interact with the mirror images of atoms at the opposite side of this very unit cell. As the London dispersion attraction of the Lennard-Jones potential only converges slowly to zero, it will generate infinitely many interactions that need to be calculated in a system with PBCs. To overcome this issue, the interactions are truncated at a certain distance, while maintaining the continuity of the potential through a switching function. The electrostatic potential defined by Coulomb's law is split into a short-range and long-range term that are calculated in real space and reciprocal space, respectively. The *Ewald summation* method or its *Particle mesh Ewald* variant are often used to compute these long-range interactions in periodic systems [150].

1.8.4 Thermodynamic ensembles

The laws of thermodynamics govern the interactions of molecules with their environment. Closed systems are often considered in laboratory experiments, i.e., isothermal titration calorimetry, and enable researches to observe the heat exchange in biomolecular reactions to ultimately calculate the thermodynamic properties of a system. In a statistical ensemble, all possible states of molecules in a given system are considered all at once. This ensemble

represents all possible states that a system can be in. A statistical ensemble at equilibrium is called thermodynamic ensemble. Three thermodynamic ensembles are used in MD simulations.

A system that is not able to exchange heat with the outside environment, where the volume V and the number of particles N stay the same is called a microcanonical ensemble. The total energy E stays constant, because there is no heat exchange with the outside. This ensemble is typically abbreviated NVE.

When the system is coupled to a thermal bath, that is sufficiently larger than the system itself, it can freely exchange heat and have a fixed temperature T . This canonical ensemble is abbreviated NVT. As the system can freely exchange heat with the external bath, the total energy of the system is not constant. Due to the PBCs the system is its own neighbor and cannot be surrounded by an external bath. Instead, in MD simulations a thermostat is used. The classic Berendsen thermostat rescales the velocities of all particles at specific time intervals, keeping the kinetic energy constant [151]. Although the Berendsen thermostat is widely used to equilibrate systems and relax them towards a target temperature, it is unable to generate trajectories that are consistent with the definition of a canonical ensemble. Instead, more sophisticated algorithms such as the Nosé–Hoover thermostat are able to generate actual NVT ensembles [152]. Even though both the temperature and average kinetic energy $\langle E_{\text{kin}} \rangle$ are fixed in the NVT ensemble, the latter can fluctuate at an instantaneous time. The two quantities are proportional to each other:

$$\langle E_{\text{kin}} \rangle = \frac{3}{2} N k_{\text{B}} T \quad (1.16)$$

where N is the number of particles, k_{B} is the Boltzmann constant and T the temperature.

The third thermodynamic ensemble that is often used in MD simulations is the isothermal-isobaric ensemble. Here, the number of particles N , the pressure p and the temperature T are kept constant. In this ensemble, typically abbreviated N_pT or NPT, the volume of the system is allowed to change such as to keep the pressure constant. The temperature is kept constant by a thermostat, as described for the NVT ensemble in the previous paragraph. In addition to a thermostat, an external barostat is used to pressurize the system. The sides of the simulation box are used as pistons to exert an external force onto the system. Similar to the NVT ensemble, in the NPT ensemble the volume is scaled at specific time intervals. The Berendsen barostat can establish the desired pressure [151], but again other barostats such as the Parrinello-Rahman barostat [153] can better reproduce the NPT ensemble.

1.9 Aim of this thesis

The aim of this thesis is to advance the understanding of protein-lipid interactions using molecular dynamics simulations in conjunction with experiments performed by collaborators. Additionally, to help advance and lower the resource usage of molecular dynamics simulations a software package was developed to help scientists with the set up of simulations. This thesis is broken down into four separate results chapters:

1. The lipid sensing abilities of the AHs of Opi1 and PIP5K are studied in detail. The AHs were embedded into lipid bilayers of varying compositions and the lipid enrichment around each AH was quantified. The specific lipid interactions were used to identify key residues and corresponding mutants were tested experimentally.
2. The binding of an unfolded Opi1 AH to lipid bilayers was studied in atomistic detail. The binding to membranes of different lipid compositions revealed preferences to certain lipid species. A rate-limiting step of α -helix folding was identified in a conserved glycine and the observation was confirmed experimentally.
3. In the third part, the NMR structures of the HIV-1Env glycoproteins are studied, to determine their stability and conformational dynamics. The accessibility of broadly neutralizing antibodies to its epitopes on gp41 is tested both for the isolated membrane anchor, as well as for the full length Env glycoprotein.
4. The last part focuses on the optimization of resource usage by molecular dynamics simulations. Different parameters can be set on supercomputers to improve the simulation performance and increase efficiency. A Python software was developed to ease the search for the best parameters.

The thesis ends with a conclusion that brings the results of the different chapters into perspective.

Lipid recognition by amphipathic helices

All investigations in this chapter were performed at the Max Planck Institute of Biophysics, Frankfurt am Main, Germany. This chapter contains work from two publications. The investigations of the Opi1 AH lipid interaction was published in the article: H.F. Hofbauer, M. Gecht, S.C. Fischer, A. Seybert, A.S. Frangakis, E.H.K. Stelzer, R. Covino, G. Hummer, R. Ernst. The molecular recognition of phosphatidic acid by an amphipathic helix in Opi1. *J Cell Biol*, 217(9):3109–3126, 2018. The investigations of the PIP5K AH lipid interactions were published in the article: T. Nishimura, M. Gecht, R. Covino, G. Hummer, M.A. Surma, C. Klose, H. Arai, N. Kono, C.J. Stefan. Osh Proteins Control Nanoscale Lipid Organization Necessary for PI(4,5)P2 Synthesis. *Mol Cell*, 75(5):1043-1057.e8, 2019. All simulations, data analysis and interpretations were carried out by myself. This work contains discussions with Dr. Harald Hofbauer (Institute of Molecular Biosciences, University of Graz, Graz) and Prof. Dr. Robert Ernst (Institute of Medical Biochemistry and Molecular Biology, Saarland University, Homburg (Saar)) for the Opi1 AH part, as well as discussions with Dr. Taki Nishimura (Molecular Cell Biology of Autophagy, The Francis Crick Institute, London) and Dr. Christopher J. Stefan (MRC Laboratory for Molecular Cell Biology, University College London, London) for the PIP5K AH part. The work for both publications was supervised by Dr. Roberto Covino (Frankfurt Institute for Advanced Studies, Frankfurt) and Prof. Dr. Gerhard Hummer (Department of Theoretical Biophysics, Max Planck Institute of Biophysics, Frankfurt).

2.1 Introduction

Organelles establish specialized compartments in cells and are identified by their specific lipid bilayer composition [9, 16]. This composition provides the platform for molecular processes at organellar surfaces, for example the control of lipid metabolism [154]. Organelles are required to maintain their surface properties and lipid composition in a process called lipid homeostasis [155]. The ER membrane is the main site of *de novo* lipid synthesis and as such contains the precursors for structural phospholipids and cholesterol, and is also responsible for the production of ceramides [155, 156]. At the same time, the ER contains only low amounts of sterols and spingolipids, as these are constantly transported through the secretory pathway to the Golgi apparatus or ultimately to the PM. The ER and PM can be differentiated by their lipid saturation, where the former has a high content of unsaturated phospholipids and the latter is mostly composed of saturated lipids [5]. Also, the two organelles have different compositions of lipid headgroups. While both membranes

contain PS lipids, only a few percent are found in the ER, whereas high levels of PS are found in the PM [154]. The major difference is the localization of the lipids in different lipids. PS headgroups are found on the luminal side of the ER, whereas they are found on the cytosolic side in the PM [9,21].

To maintain the distinct properties of organelles through lipid homeostasis, cells have to resort to various sensors that report on the chemical and physical properties of membranes [50]. A common type of sensor found in eukaryotic cells are AHs. As stated in the introduction of this work, AHs are α -helical peptides with two distinct sides, one hydrophilic and the other hydrophobic. They not only deform bilayers, but also sense membrane curvature, recognize specific lipids, coat lipid droplets and protect membranes from stress [49]. In this chapter, two different AHs will be studied. As described in the introduction of this work, the protein Opi1 serves as a key regulator of phospholipid homeostasis in *S. cerevisiae*. In the presence of PA, it is bound to the ER, while it shuttles into the nucleus after depletion of PA. A “PA-binding motif” was identified to be responsible for the specificity of Opi1 to PA [70]. The removal of this binding motif deactivates the specificity of Opi1 for PA containing membranes. In a recent publication Hofbauer *et al.* [157] has shown that this “PA-binding motif” contains an AH. The second AH is found in the protein PIP5K. It is responsible for the phosphorylation of PI4P to produce PI(4,5)P₂ at the PM. Its “specificity loop” has been implicated to be responsible for membrane binding and PI4P recognition [158–161]. The mechanism with which both of these AHs sense their respective lipid of interest remains elusive and are subject of this chapter.

Aim. Here, extensive all atom MD simulations are used to elucidate the dynamics of both the Opi1 AH and PIP5K AH in model membranes representing the ER and PM, respectively. The α -helical peptides were embedded into the lipid bilayers and specific lipid interactions were studied over the course of extensive equilibrium MD simulations. The results show how both of these AHs are tuned to sense a specific lipid headgroup, while ignoring others. Their primary amino acid sequence maps onto the α -helical secondary structure to produce specific binding motifs. These provide each AH with their own specificity for certain lipid headgroups and help to explain their molecular sensing mechanism.

2.2 Methods

2.2.1 Coarse-grained molecular dynamics simulations

A 22-mer Opi1 AH peptide (QKLSRAIAKGKDNLKEYKLNMS) was modeled into an ideal α -helical conformation using UCSF CHIMERA [162] and converted into a coarse-grained (CG) representation for the MARTINI force field using the *martinize.py*-tool [163]. A cubic box of 10 nm³ containing bulk water, a lipid bilayer and 150 M NaCl was assembled using the *insane*-tool [164], putting the helical peptide at 4 nm from the bilayers center of mass. Different bilayer systems with varying ratios of phospholipids were set up. Acyl chains were kept constant with a ratio of 6:4 between 1-palmitoyl-2-oleoyl (16:0–18:1) and 1,2-dioleoyl (18:1), while the headgroup species and ratios were varied. The membranes were composed of different combinations of four lipid species: 1,2-dioleoyl-sn-glycero-3-phosphate (DOPA),

1,2-dioleoyl-sn-glycero-3-phosphoserine (DOPS), 1,2-dioleoyl-sn-glycero-3-phosphocholine (DOPC) and 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC). Their ratios were varied and the exact membrane compositions are listed in Table 2.1.

Table 2.1: Lipid compositions used in Opi1 AH simulations. The acyl chain ratio was kept constant, while the headgroups were varied. Each system was simulated in a single repetition for 3 μ s each.

System	mol%	Lipid
20% PA	60%	POPC
	20%	DOPC
	20%	DOPA
20% PS	60%	POPC
	20%	DOPC
	20%	DOPS
20% PA/ 20% PS	60%	POPC
	20%	DOPA
	20%	DOPS

CG simulations were performed using the GROMACS software package version 5.1.3 [165] with the MARTINI force field version 2.2 [163, 166–169]. The integration time step was set to 20 fs with a neighbor list update every 20 steps. Non-bonded interactions were cut off at 1.1 nm using the Verlet neighbor list. Temperature was kept constant at 310 K with the Velocity Rescale [170] thermostat with a characteristic time of 1 ps. Semiisotropic pressure coupling at 1 bar using the Parrinello-Rahman barostat was used with a characteristic time of 12 ps [171]. Periodic boundary conditions were imposed in all directions. Each system was run for a total of 3 μ s.

2.2.2 Backmapping of CG systems to atomistic representations

CHARMM-GUI [172, 173] was used to lift three Opi1 AH CG systems (20% PA, 20% PS and 20% PA/PS, see Table 2.1) to an atomistic description in the CHARMM36 force field [174]. The energy of the system was minimized for 5000 steps, keeping a restraint on all heavy atoms. The systems were equilibrated in a canonical (NVT) ensemble for 50 ns at a temperature of 310 K with a time step of 1 fs. Next, a constant pressure of 1 bar was established in the isothermal-isobaric (NPT) equilibration during 375 ns with a time step of 2 fs. Both the Berendsen barostat and thermostat were used during these two equilibration steps [151]. All non-bonded interactions were cut off at 1.2 nm. The Parrinello-Rahman barostat with semiisotropic pressure coupling was applied on the protein, membrane and solvent separately every 5 ps during production simulations [171]. Each systems was simulated for 10 μ s. In all simulations N- and C-terminus were acetylated and amidated, respectively.

2.2.3 Atomistic simulations of PIP5K AH

A 24-mer peptide (LQSYRLVKKLEHSWKALLHDGDTV) of PIP5K (phosphatidylinositol-4-phosphate 5 kinase, accession number Q50313) was modeled into an ideal α -helix with

UCSF CHIMERA [162]. CHARMM-GUI [172,175–178] was used to set up atomistic plasma membrane-like model systems with the CHARMM36m force field [179]. All simulations were run using the GROMACS 2018 software package [165]. Two systems with varying amounts of POPC, DOPS and PI4P were set up, with one also containing cholesterol. The membrane compositions are listed in Table 2.2. N- and C-terminus were neutralized, NNEU and CNEU respectively, whereas H391 and H398 were protonated. The peptide was embedded into the lipid headgroup region during system construction. The AH was translated by 16 Å along the z-axis. First, counter ions were added to neutralize the system and subsequent ions were added to a total concentration of 150 M NaCl. All systems were energy minimized with steepest descent and equilibrated in multiple subsequent steps. First the system was equilibrated in a canonical (NVT) ensemble with a time step of 1 fs for 25 ps, maintaining a constant temperature of 310 K with the Berendsen thermostat [151]. During the equilibration position restraints of $1000 \text{ kJ mol}^{-1} \text{ nm}^{-2}$ were applied to all lipid heavy atoms, whereas restraints on protein backbone heavy atoms were lowered from 4000 to $2000 \text{ kJ mol}^{-1} \text{ nm}^{-2}$. The Berendsen barostat was used to keep a constant pressure of 1 bar in the isothermal-isobaric (NPT) equilibration for the first 25 ps with a time step of 1 fs [151]. All further equilibration steps were performed for 300 ps with a time step of 2 fs. Here, lipid position restraints were decreased from 400 to $0 \text{ kJ mol}^{-1} \text{ nm}^{-2}$ and protein position restraints were decreased from 1000 to $200 \text{ kJ mol}^{-1} \text{ nm}^{-2}$. During production simulations no restraint were applied on the systems. The temperature was kept at 310 K using the Velocity Rescale thermostat [170] with a characteristic time of 1 ps. The Parrinello-Rahman barostat [171] was used with semiisotropic pressure coupling to keep a pressure of 1 bar. The pressure coupling was applied each on the protein, solvent and membrane with a characteristic time of 5 ps. All non-bonded interactions were cutoff at 1.2 nm in all simulations.

Table 2.2: Lipid compositions used for atomistic simulations of PIP5K with a model plasma membrane-like bilayer. The time column shows the cumulative simulation time of all repetitions.

System	mol%	Lipid	Repetitions	Time (μs)
With Cholesterol	69%	POPC	3	23.8
	20%	Cholesterol		
	10%	DOPS		
	1%	PI4P		
Without Cholesterol	89%	POPC	1	10
	10%	DOPS		
	1%	PI4P		

2.2.4 Analysis of lipid localization

For every frame in the trajectory phosphate atoms belonging to a specific lipid type were binned according to their x and y positions on a two-dimensional grid, using a bin area of 1 \AA^2 and a sampling time of 500 ps. The rotation angle for every frame was obtained by using an in-place least-square fit of the backbone C_α atoms to a reference structure

of the AH aligned along the x-axis. In every frame, the two-dimensional grid was rotated by the calculated angle orthogonal to the plane of the membrane, using third order spline interpolation. By considering periodic boundary conditions, the grid in every frame was extended and cropped after the rotation to keep the original dimensions. Finally the lipid density was calculated by averaging all two-dimensional grids along the full 10 μs simulations. The code to determine the in-place least-square fit was written and provided by Dr. Roberto Covino.

2.2.5 Calculation of distances

The distance between atom groups of interest was calculated with the *distances* module in the MDAnalysis package [180,181]. For lipids the phosphate atom was used as a reference. In the “3K motif” the center of mass (COM) of C_α atoms of K121, K125 and K128 was used, whereas in the “KRK motif” the COM of C_α atoms of K112, R115 and K119 were used. The distance distributions were evaluated by dividing the distances smaller than 13 Å into 85 spatial bins and in each frame counted the number of distances in every bin.

2.2.6 Calculation of residence times

For each frame of the trajectory the phosphate atom with the smallest distance to the reference group was determined. The reference group was defined as the COM of C_α atoms of all residues in either the “3K motif” or “KRK motif”. The number of consecutive frames that this specific phosphate atom stayed below the cutoff of 5.5 Å were counted. These consecutive stretches were defined as single “binding events”. The sum of all binding event durations was calculated and divided by the number of binding events to determine the average.

2.3 Results

In the following section, the results of atomistic MD simulations of two AHs are described. The section starts with simulations of the Opi1 AH in model membranes containing different amounts of negatively charged lipid headgroups, either PA or PS. It reports on the enrichment of these lipids at specific locations around the AH. Also it shows how in mixed lipid bilayers, when both PA and PS are included in sufficient amounts, PA stays enriched at the AH, but PS is depleted from it. The molecular details of these interactions are studied in detail. Differences between the interactions of the AH with either PA and PS headgroups are explained. At the end, the residence times of different lipid types are estimated for various membrane compositions. The section then continues with the atomistic MD simulations of the PIP5K AH in model PM-like lipid bilayers. Again, the enrichment of negatively charged lipids is observed at distinct spots around the AH. The underlying molecular interactions between each lipid (or sterol) and the AH are considered in detail and explained. The implications on the sensing functionality of these two AHs are discussed.

2.3.1 Charged lipids preferentially localize along the Opi1 AH

To study the preference of the AH to interact with different lipid species, the AH was first modeled as ideal α -helix using UCSF CHIMERA and embedded in a lipid bilayer. More specifically, after modeling of the α -helix in an atomistic representation, the 22-mer peptide was coarse-grained for usage with the MARTINI force field. The procedure and results were previously described [157, 182], but will be summarized again in the following paragraph. Three distinct lipid bilayers were generated, containing either 20% PA, 20% PS or 20% PA and 20% PS headgroups with the remaining percentage being PC. The exact lipid compositions and acyl chain contents are listed in Table 2.1. The AH was placed at a distance of 4 Å from the lipid bilayer’s COM. During the CG simulation it embedded into the membrane. The AHs hydrophobic residues inserted into the membrane, whereas the hydrophilic residues remained either at the lipid-water interface or pointed towards bulk solvent. The AH never dissociated from the membrane even after a simulation time of 1 μ s. The system was then lifted back into an atomistic representation. The three atomistic systems were run for 10 μ s each, accumulating a total of 30 μ s simulation time (Figure 2.1).

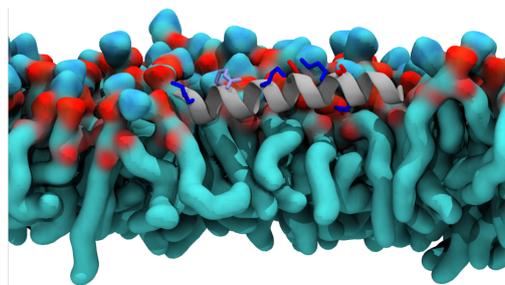


Figure 2.1: Snapshot of a fully folded Opi1 AH embedded in a lipid bilayer. The Opi1 AH embeds into lipid membranes such that it is found at the lipid-water interface, roughly at the height of the glycerol backbone. The AH is shown as gray ribbon with charged side chains as sticks (light blue: arginine, blue: lysine, red: aspartate and glutamate). Lipids are shown as densities with carbons in cyan, phosphates in orange, oxygens in red and nitrogens in blue. Only the leaflet with the inserted Opi1 AH is shown. Water and ions are not shown for clarity.

In these extensive atomistic MD simulations the Opi1 AH remained α -helical and did not dissociate from the membrane. A single glycine (G120) found in the AH was observed to transiently lose its α -helical structure, but did not lead to further unfolding of the remaining α -helix. All lipids were mobile and frequently explored the entirety of the lipid bilayer. Closer inspection revealed a long residence of both DOPA and DOPS lipids around the AH, but not of DOPC and POPC. Given the negative charge of -1 of both PA and PS headgroups, the first assumption was to check for electrostatic interactions between the lipids and the AH. To quantify the accumulation of lipid species around the AH, their localization probability was computed as an average over the whole 10 μ s trajectory. In the presence of only PA headgroups, distinct spots indicate the accumulation of the lipid at specific locations around the AH (Figure 2.2A). No lipid accumulation is observed on the opposite leaflet, indicating that there is no cross-talk between the two leaflets (Figure 2.2B). The accumulation is most prominent on the right-hand side towards the C-terminal end of the AH, with some visible density also at the N-terminal top. No lipid accumulation is observed at the C-terminal left-hand side.

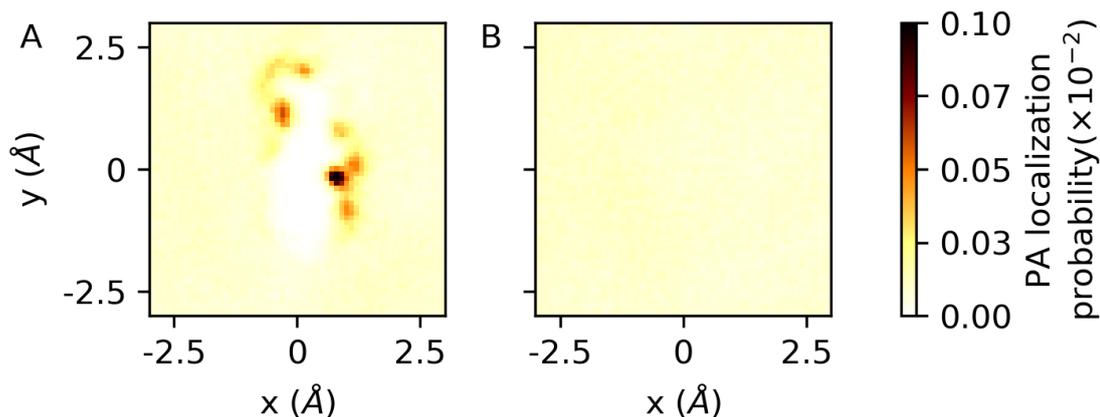


Figure 2.2: Time-averaged positions of phosphate headgroups in a 20% PA and 80% PC lipid bilayer. (A) Light yellow region at the origin (0, 0) represents the AH. Darker regions correspond to a higher density of lipids at that position over the course of the trajectory. DOPA enriches at distinct spots around the AH. (B) No enrichment of DOPA lipids is visible in the opposite leaflet.

In comparison, in the simulation with PS instead of PA headgroups, the lipid also accumulates around the AH, but in a different pattern (Figure 2.3A). Again, no lipid accumulation is observed on the opposite leaflet that does not contain the AH (Figure 2.3B). Here, the most prominent lipid accumulation is observed at the N-terminal left-hand side. Two distinct spots are visible, but one is more prominent than the other. Also, similar to the bilayer with PA headgroups, PS lipids enrich on the right-hand side towards the C-terminus of the AH. The spots visible in the PS densities are more smeared out (Figure 2.3A), when compared to those in the PA densities (Figure 2.2A). No accumulation of lipids with PC headgroups is observed in either of the two leaflets. Electrostatic interactions seem to drive the enrichment of negatively charged lipids at the positively charged sides of the Opi1 AH.

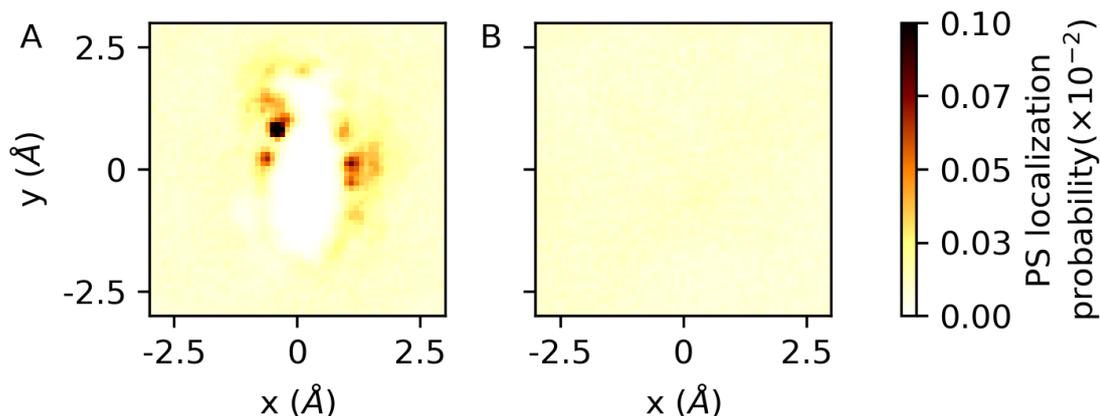


Figure 2.3: Time-averaged positions of phosphate headgroups in a 20% PS and 80% PC lipid bilayer. Light yellow region at the origin (0, 0) represents the AH. Darker regions correspond to a higher density of lipids at that position over the course of the trajectory. (A) DOPS enriches at distinct spots around the AH. (B) No enrichment of DOPS is visible in the opposite leaflet.

PA headgroups outcompete the bigger PS headgroup at the AH. Both headgroups have a net charge of -1 and thus are not differentiable by their overall charge alone. To test if

there is any specificity of the AH for PA over PS headgroups, a mixed bilayer containing 20% DOPA, 20% DOPS and 60% POPC was set up with the Opi1 AH embedded in one of the leaflets. After a 10 μ s equilibrium simulation the localization of the different lipid species was calculated (Figure 2.4). As described above both PA and PS headgroups are found in close proximity in membranes only containing one of the two negatively charged species. In contrast, in a mixed bilayer only PA is enriched at the AH (Figure 2.4A). PS is only found for short amounts of time at the AH (Figure 2.4B). These observations in a mixed bilayer point to the ability of the Opi1 AH to differentiate lipids of the same charge.

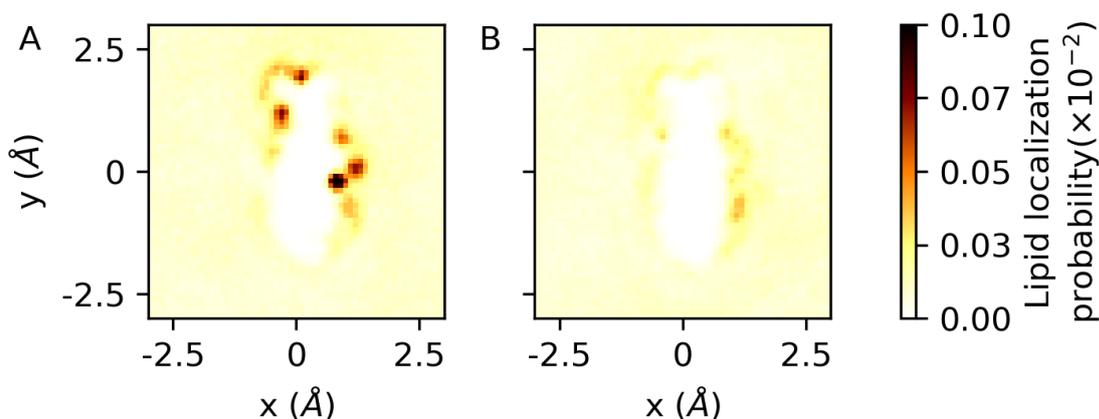


Figure 2.4: Time-averaged positions of phosphate headgroups in a 20% PA, 20% PS and 60% PC lipid bilayer. Light yellow region at the origin (0, 0) represents the AH. Darker regions correspond to a higher density of lipids at that position over the course of the trajectory. (A) DOPA enriches at distinct spots around the AH. (B) DOPS is depleted from the AH when PA is present in the same leaflet.

2.3.2 Coordination of lipids is primarily driven by lysines

To better understand the molecular interaction between the AH and charged lipids, the structure of the Opi1 AH was considered under more detail. The AH includes a total of eight charged residues. Five lysines are found stretching to the sides of the helix, while one arginine is placed on top of the N-terminal hydrophilic side of the AH. The two negatively charged residues D122 and E126 are found on one side of the C-terminus. Figure 2.5 visualizes the eight charged residues. It is possible to divide the positively charged residues into two distinct groups by their localization along the principal axis of the peptide. Two lysine residues and one arginine residue, K112, R115 and K119, form the first group at the N-terminus of the AH and are from here on called “KRK motif” (Figure 2.5). The second group consists of three lysine residues K121, K125 and K128 close at the C-terminus and will be called “3K motif”. It is noteworthy that the two lysines K121 and K128 lie on the same height, while K125 resides slightly shifted to the top of the helical turn (Figure 2.5B, bottom).

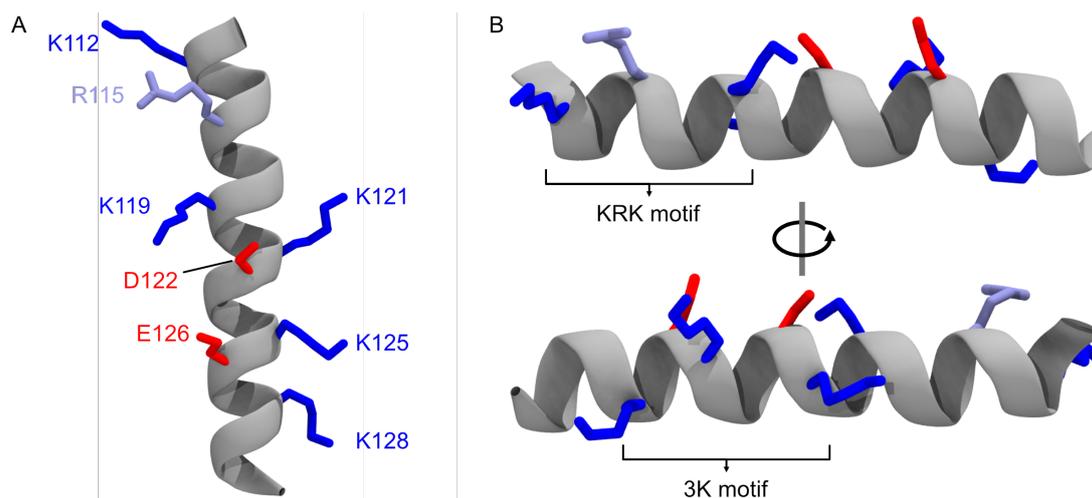


Figure 2.5: Structure of fully folded Opi1 AH when embedded in a lipid bilayer. Positively charged residues are shown as blue sticks and negatively charged residues as red sticks. (A) Top view of Opi1 AH from N- (top) to C-terminus (bottom). Two distinct lysine/arginine rich sides are visible at the top left and bottom right. (B) The “KRK motif” found at the top left consists of K112, R115 and K119. The “3K motif” consists of three lysines: K121, K125 and K128. Lipids, water and ions are not shown for clarity.

2.3.3 Phosphate moieties of PA lipids are the major interaction partners

The PA lipids interact at different positions with the AH. Specific electrostatic interactions of the phosphate moiety with different parts of the AH are observed throughout the trajectory (Figure 2.6). The PA lipids are predominantly found at four binding sites of the AH. Region 1, the “KRK motif”, corresponds to a DOPA lipid interacting with its PA headgroup with either K112, R115 or K119 (Figure 2.6A). The interactions with both lysines is mediated by their positively charged nitrogen atoms and the negatively charged oxygen atom of the phosphate moiety. R115 is observed to align its hydrogens in the general direction of the negatively charged oxygen atoms of the PA headgroup. Further, the hydrogen belonging to the hydroxyl-group of the phosphate moiety temporarily aligns itself to the carbonyl-group of arginines backbone. Electrostatic interactions with involvement of the backbone oxygen are present in all of the three regions. Region 2 originates from the interactions between the positively charged nitrogen in the side chain of K112 and the negatively charged oxygen atoms in the phosphate moiety (Figure 2.6B). Region 3, the “3K motif”, is formed by various interactions between the positively charged nitrogen atoms of the lysines K121, K125 and K128 with the phosphate moieties oxygen atoms (Figure 2.6C). Further the backbone carbonyl oxygen of K121 interacts with the phosphate moieties hydroxyl hydrogen. The phosphate headgroup seems to fit into the binding motif that the three lysines provide.

2.3.4 PS headgroups allow for two possible modes of interaction

The same interaction patterns between the AH and DOPS are observed, as already described for DOPA in the previous section. Intriguingly, two different modes of interaction are observed between the serine headgroup and lysines in the “3K motif” (Figure 2.7).

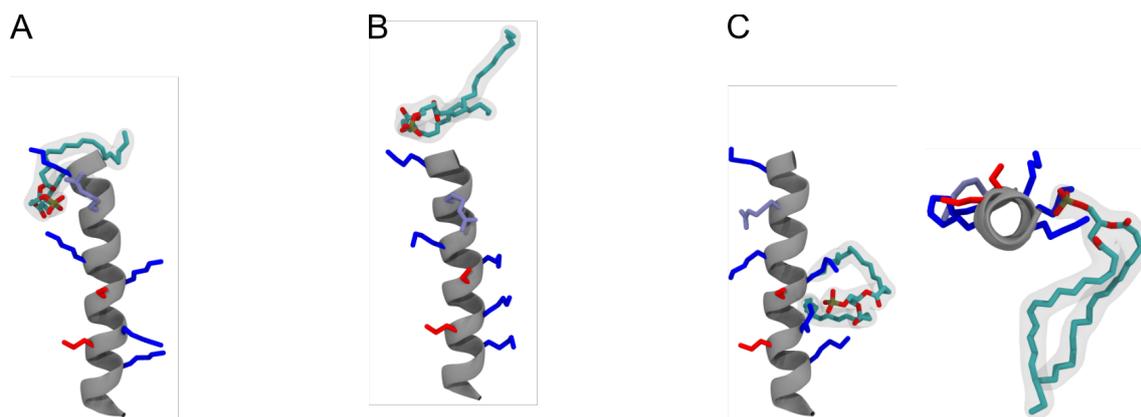


Figure 2.6: PA interacts with the Opi1 AH at different interaction sites. (A) PA is coordinated by K112, R115 and K119 at the N-terminal “KRK motif”. (B) PA interacts with K112 at N-terminal end. (C) PA is coordinated by the three lysines K121, K125 and K128 that form the “3K motif”. The side view (right panel) shows how the phosphate moiety of PA is wedged between the lysines. Lipids, solvent and ions are not shown for clarity.

In the first mode, the negatively charged oxygen atoms in the phosphate moiety interact with the positively charged nitrogens found in K121, K125 and K128 (Figure 2.7A). The second mode is defined by the interaction of the partially negatively charged oxygen atoms from the serine headgroup with the positively charged nitrogens from the same three lysines (Figure 2.7B). Depending on the interaction mode, DOPS is held at different distances to the AH. When the serine headgroup interacts with the lysines, it pushes back the phosphate moiety and increases its distance to the AH. This way the phosphate is either found closer to the AH (Figure 2.7A) or further away (Figure 2.7B). The visual inspection of the lifetimes of both modes indicates that they are short lived. This points to a possible explanation of why the DOPS localization probability appears to be broader and less localized (Figure 2.3), when compared to the localization probability of DOPA (Figure 2.2).

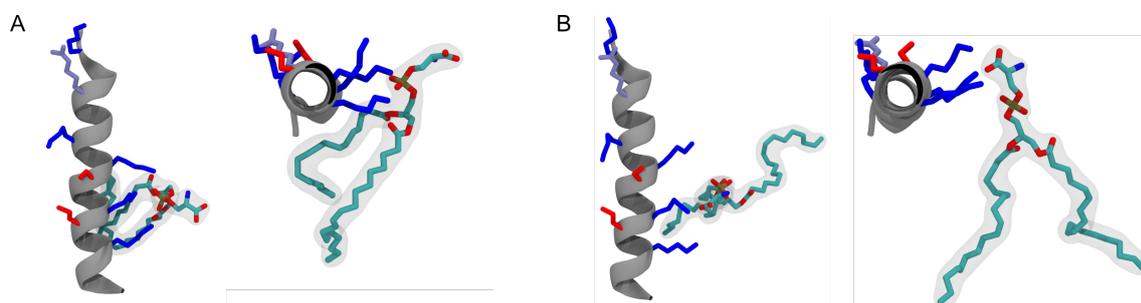


Figure 2.7: The PS headgroup interacts with the “3K motif” in two different modes. Two different snapshots from the 20% DOPS simulation show possible interaction modes of one DOPS headgroup with the lysines K121, K125 and K128 that form the “3K motif”. (A) Lysines interact with the oxygen phosphate of a PS headgroup. (B) Lysines interact with oxygen atoms of the serine side chain. Lipids, solvent and ions are not shown for clarity.

2.3.5 The 3K motif provides space only for PA

The distribution of the three lysine residues K121, K125 and K128 along the secondary structure of the peptide is that of a cage with a boundary to each side (K121 and K128) and one on top (K125) (Figure 2.6C). As mentioned in the previous sections, the two lipid headgroups PA and PS show differences in interaction with the “3K motif”. The phosphate headgroup of DOPA is able to move closer to the AH backbone, while being encapsulated inside the “3K motif”. Further, interactions with the backbone carbonyl oxygens are possible, as mentioned in Section 2.3.3. The phosphate head group of DOPS on the other hand is not able to move as close to the AH backbone. The additional serine provides steric hindrance for the moiety to get any closer. The two state interaction, as described in Section 2.3.4, may also keep the phosphate at a distance.

The distance distribution between the C_α atoms of the “3K motif” lysines and all phosphates in the leaflet was calculated to quantify the proximity between the AH and various lipid headgroups. The center of mass of all C_α atoms from the three lysines K121, K125 and K128 was assumed to be the reference point. The distance between the phosphates of any given lipid and this reference point was calculated for the whole trajectory (Figure 2.8). In membranes with DOPA content of 20%, the DOPA density peaks at roughly 4.5 \AA (Figure 2.8A). Lipids with PC headgroups are excluded up to a distance of 5 \AA and never enrich to such high levels as DOPA even at distances of more than 10 \AA . In membranes with 20% DOPS lipids, DOPS is almost never found at distances less than 5 \AA , similar as lipids with PC headgroups (Figure 2.8B). When both 20% DOPA and 20% DOPS lipids are present in the same leaflet, DOPA outcompetes DOPS as described in a previous section (Figure 2.8C). POPC is found at similar densities as for the other two compositions. As shown earlier, DOPA can interact with the “3K motif” at a closer distance, whereas lipids with PS headgroups are held at bigger distances.

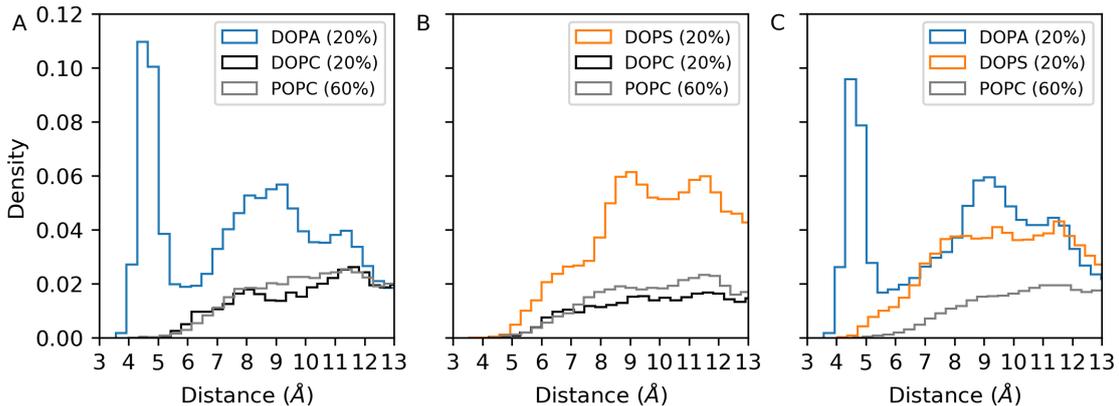


Figure 2.8: Probability density of the distance between the “3K motif” and phosphates in different membrane environments. Distance distributions between the combined center of mass of all C_α backbone atoms from the “3K motif” and every phosphate headgroup in DOPA (blue), DOPS (orange), DOPC (black) and POPC (gray) lipids. (A) DOPA comes into close proximity of $\sim 4.5 \text{ \AA}$, (B) whereas DOPS is excluded from these close distances. (C) In the presence of both PA and PS headgroups, PA outcompetes the latter and depletes it from the AH.

The distance distribution at the “KRR motif” shows a similar picture (Figure 2.9). In

membranes containing both PA and PC headgroups, PA lipids are found at a distance of ~ 5 Å to the C_α atoms of the three residues that form the “KRK motif” (Figure 2.9A). Intriguingly, DOPC and POPC are found at closer distances when compared to the “3K motif”. Also, all three lipids show a different fingerprint. In addition to the first peak at 5 Å, they are found at ~ 10 Å. In membranes containing PS headgroups, DOPS is predominantly found at a distance of 6.5 Å (Figure 2.9B). Intriguingly, it is also found at smaller distances like 4 Å. The different arrangement of the three residues K112, R115 and K119, each residue being positioned further up on the α -helical loop, seems to fit the PS headgroup better than at the “3K motif”. Again, the presence of PA headgroups almost completely depletes the DOPS lipids from the Opi1 AH (Figure 2.9C).

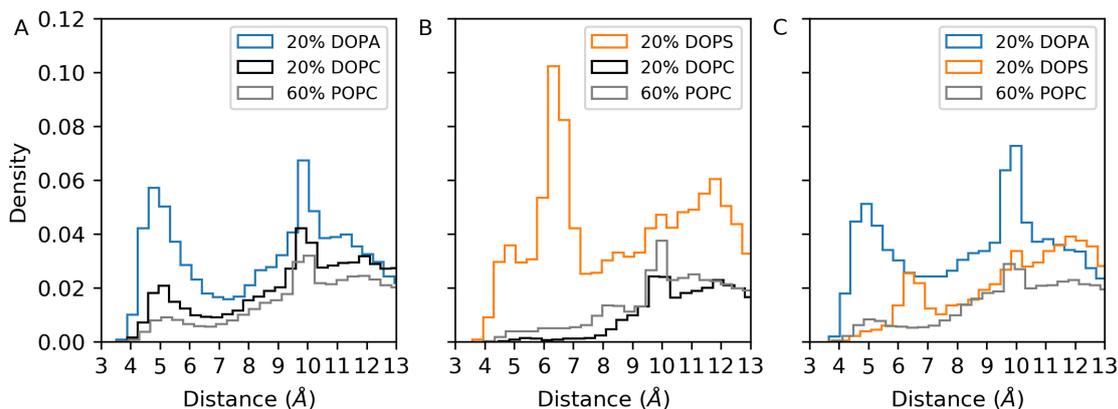


Figure 2.9: Probability density of the distance between the “KRK motif” and phosphates in different membrane environments. Distance distributions between the combined center of mass of all C_α backbone atoms from the “KRK motif” and every phosphate headgroup in DOPA (blue), DOPS (orange), DOPC (black) and POPC (gray) lipids. (A) DOPA is found at ~ 5 Å. (B) DOPS appears most prominent at ~ 8 Å. (C) In the presence of both PA and PS headgroups, PA outcompetes PS from the AH almost completely.

2.3.6 PA lipids can reside longer at the Opi1 AH

DOPS lipids stay at the “3K motif” for significantly shorter times than DOPA. The observations from the previous chapters have revealed that PS headgroups interact in two modes with the “3K motif”: (i) the lipids’ phosphate moiety interacts with the positively charged lysines, (ii) the oxygen atoms in the lipids’ serine headgroup interact with the charged lysines. Visual inspection of the trajectories revealed a difference in the times that either DOPA or DOPS reside at the “3K motif”. To quantify this observation the average residence time of all lipids at both the “3K motif” and the “KRK motif” was calculated over the whole trajectory in simulations containing either 20% PA or 20% PS. Figure 2.8 showed that only PA is able to come into contact of less than 5 Å. To make the comparison viable lipids were considered to be in the “bound state” if they were at most 5.5 Å away from either of the motifs. Again, the reference position inside the motifs was defined by the COM of the three residues C_α atoms. A residence event was defined by counting the number of consecutive frames in which the closest lipid to either of the two motifs is closer than 5.5 Å. The mean and standard error of the mean was calculated from all observed

residence events for each lipid species (Figure 2.10). ~ 60 events were observed for the three lipids at the “3K motif”, where DOPA resides the longest $86.6 \text{ ns} \pm 0.91 \text{ ns}$ (Figure 2.10A). The residence time of DOPS is lower by 2.5-fold with $34.0 \text{ ns} \pm 8.6 \text{ ns}$. POPC resides for an even shorter time of $15.7 \text{ ns} \pm 3.8 \text{ ns}$.

At the “KRK motif” the residence times are not differentiable. While DOPA has a residence time of $60.7 \text{ ns} \pm 12.6 \text{ ns}$, DOPS shows a similar residence time with a very high standard error of $50.4 \text{ ns} \pm 18.7 \text{ ns}$ (Figure 2.10B). The high standard error of DOPS also overlaps with the average residence times of POPC ($33.2 \text{ ns} \pm 6.6 \text{ ns}$). Only a low number of events was observed for DOPS ($n = 28$) leading to a high standard error. A longer simulation would be required to increase sampling and achieve a better estimate of the residence time at the “KRK motif” for DOPS.

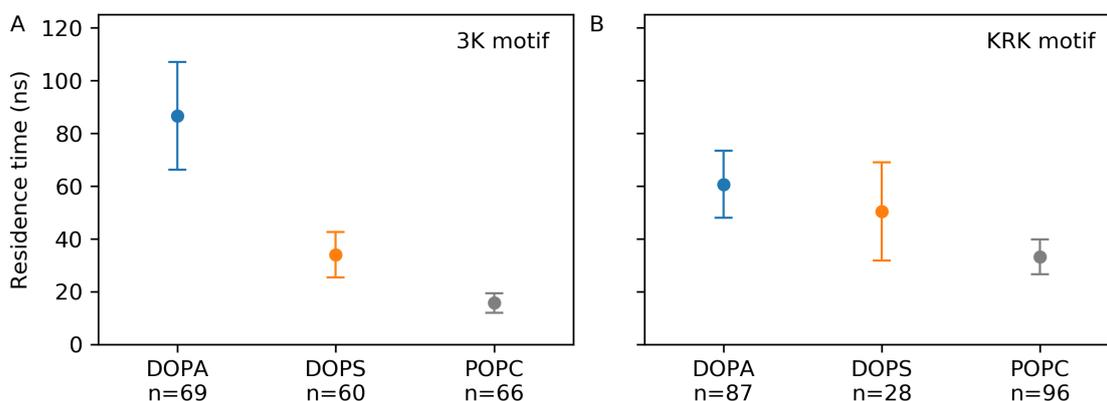


Figure 2.10: Residence times of different lipid species at the “3K motif” and “KRK motif”. The average residence times for the different lipid species found either in the (A) 20% PA or (B) 20% PS system. Only lipid headgroups with a maximal distance of 5.5 \AA to the average center of mass of backbone C_{α} atoms of the “3K motif” were considered.

2.3.7 PIP5K AH interacts with a variety of lipids

The previous sections showed how the Opi1 AH interacts with negatively charged lipid headgroups and specifically prefers PA over PS. This sensing functionality allows the AH to act as a sensor and retain Opi1 at the ER membrane in the presence of PA. As stated earlier, in the absence of PA the Opi1 protein dissociates from the membrane and shuttles into the nucleus to interact with Ino2/Ino4 and ultimately downregulate *de novo* synthesis of GPLs. Similarly, a “specificity loop” in PIP5K was shown to provide the overall membrane targeting and substrate specificity [158–160]. The protein catalyzes the phosphorylation of PI4P to PI(4,5)P₂. This loop was shown to be unstructured when PIP5K is in solution, but to fold into an AH upon membrane binding [161]. Intriguingly, the AH is in close proximity to the catalytic aspartate residue of PIP5K [161]. To test if the AH of PIP5K is able to enrich specific lipids similar to the AH of Opi1, multiple simulations were set up.

Experimental data by Taki Nishimura has shown that the presence of PS headgroups and sterols enhances the activity of PIP5K [183]. To get insights on how the AH of PIP5K interacts with different lipids, atomistic MD simulations were set up. Two different membrane compositions were considered and a fully folded PIP5K AH was embedded into

the bilayers at the lipid-water interface. The two membrane compositions are differentiated by the presence and absence of cholesterol and the mol% of POPC (Table 2.2). The low mol% of 1% PI4P in each bilayer represents the low abundance of the lipid in the PM. This resulted in only three PI4P lipids per leaflet. Three repetitions of the system with cholesterol were set up independently, to randomly distribute the PI4P lipid on the membrane. These repetitions were run for a total of 23.8 μs , one for 10 μs and the other two for almost 7 μs each. The system without cholesterol was run in a single repetition for 10 μs .

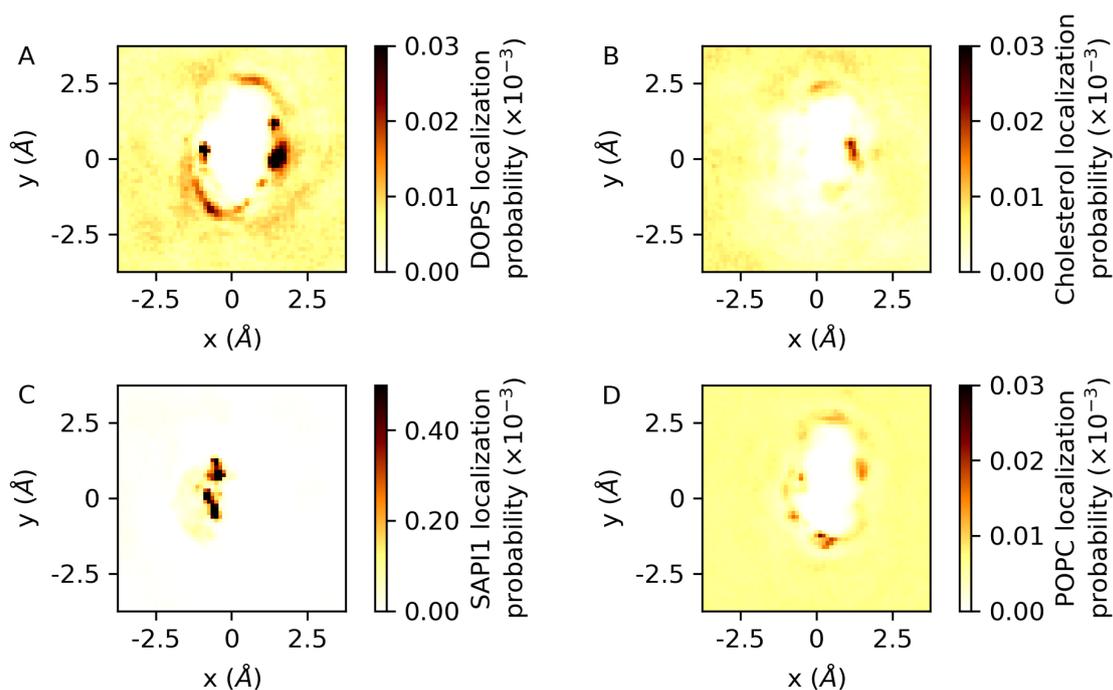


Figure 2.11: Time-averaged positions of different atoms in a plasma membrane-like model bilayer. (A) DOPS enriches at both sides of the AH, but most prominently on the right hand side. (B) Cholesterol enriches only on the right-hand side. (C) PI4P shows a density only on the left-hand side of the AH. The sensitivity of this panel was increased to make the lipid visible given its low abundance. (D) POPC does not significantly enrich at the AH.

The AH loses structure both at the N- and C-terminus, but remains fully α -helical at its core (R384–H398) and never dissociates from the membrane in the simulations. The lipid localization probability was calculated for all systems (Figure 2.11). DOPS is enriched at the PIP5K AH on the two sides, but most prominently on the right-hand side (Figure 2.11A). In contrast, cholesterol is only enriched at a narrow patch on the right-hand side (Figure 2.11B). PI4P is seen to permanently associate with the AH, never fully dissociating over the course of the 10 μs simulation. Two distinct spots can be seen for PI4P on the left-hand side of the AH (Figure 2.11C). In a different instance of the simulation PI4P is observed to also permanently interact at the opposite side of the AH. No clear enrichment can be observed for the highly abundant POPC at the PIP5K AH, indicating an affinity for charged interactions of the AH (Figure 2.11D). The single 10 μs simulation of the PIP5K AH embedded into a membrane without cholesterol shows the same behavior of the lipid enrichment around the AH. The absence of cholesterol did not

affect the binding affinity in the atomistic MD simulations.

2.3.8 PIP5K AH is lined with charged residues on all sides

Similar to the previous approach with the Opi1 AH, the PIP5K AH was studied under more detail to better understand the molecular interactions between its residues and charged lipids. The AH of PIP5K is lined with 6 positively charged residues, R384, K387, K388, H391, K394 and H398, as well as three negatively charged residues: E390, D399 and D401 (Figure 2.12). Both K388 and K394 are located at almost the same height, pointing directly into the lipid-water interface, whereas all other positively charged lipids are situated on top of their respective helical turns. R384, K387, H391 and H398 are able to interact with lipids on both sides of the PIP5K AH, because of their positions. The negatively charged residue E390 is found at the center of the AH pointing to the side, whereas D399 and D401 are located at the C-terminus positioned clearly above the lipid-water interface (Figure 2.12B).

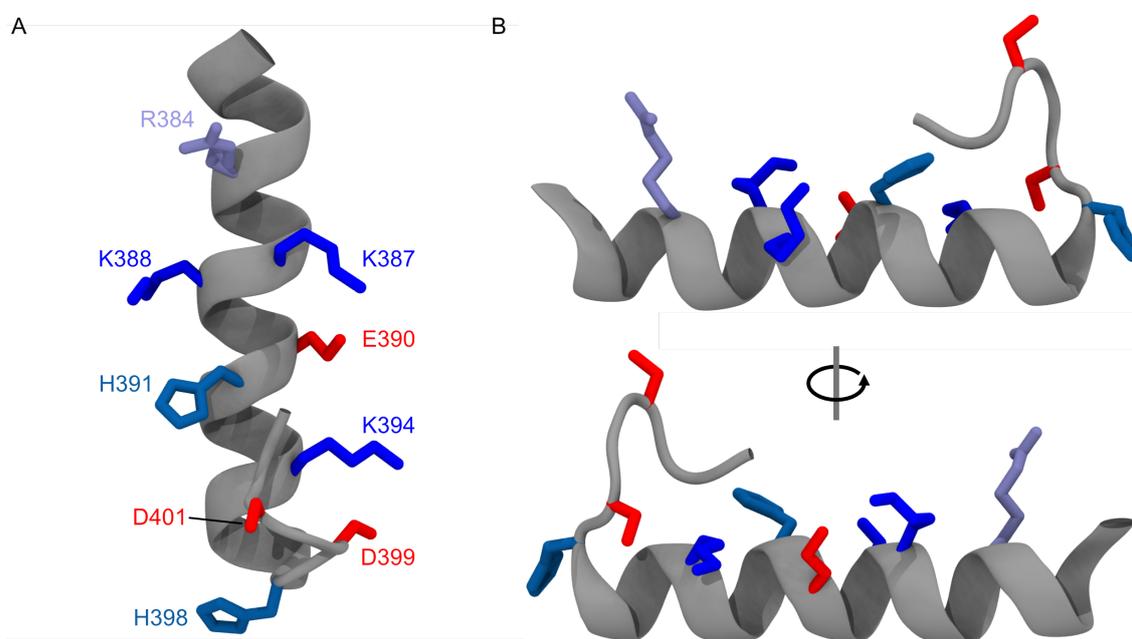


Figure 2.12: Structure of fully folded PIP5K AH embedded in a model PM-like lipid bilayer. Positively charged residues are shown as blue sticks and negatively charged residues as red sticks. (A) Top view of PIP5K AH from N- (top) to C-terminus (bottom). The left side contains only positively charged residues, whereas the right side also contains two negatively charged residues. (B and C) Side views of PIP5K AH highlighting its charged sides. Lipids, water and ions are not shown for clarity.

2.3.9 PI4P interacts with PIP5K AH through electrostatic network

The PIP5K AH interacts with different kinds of lipids in various ways, but most preferentially with PI4P. Visual inspection reveals different forms of interactions between the membrane embedded PIP5K AH and negatively charged lipids in the multi-microsecond atomistic MD simulations. Although only three PI4P molecules were placed into each leaflet they associate with the AH after several hundreds of nanoseconds. A closer look at the association process reveals an interesting event. Multiple DOPS molecules are found

close to the AH by chance, interacting with the positively charged residues. The negatively charged oxygen atoms of a single PI4P molecule, that is several lipid shells away from the AH, starts interacting with the positively charged nitrogen on the serine headgroup of these DOPS lipids. As these lipids are already in close proximity to the PIP5K AH, they ultimately lead the PI4P lipid towards the charged side chains of the AH they themselves are interacting with. The diffusion of this PI4P lipid appears to be directed towards the AH and it ultimately associates in close contact with it, displacing any DOPS molecules at the same binding area.

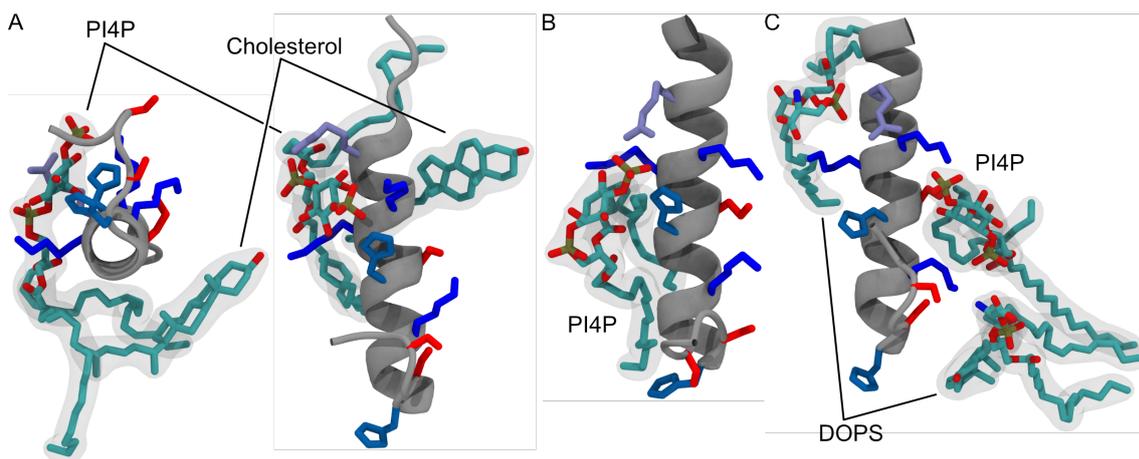


Figure 2.13: Charged lipids interact with PIP5K AH. Three snapshots of lipid interactions at different positions of the AH are shown. (A) PI4P is bound between R384 and K388. A single cholesterol is found at the opposite side. (B) PI4P is wedged between K388 and H391. (C) PI4P interacts at the opposite side of the AH, wedged in between K387 and K394. Two DOPS lipids interact with the AH as well. Colors are the same as in Figure 2.12. The Membrane, water and ions are not shown for clarity.

The interaction between PI4P and the PIP5K AH has three possible stable states (Figure 2.13). In the most prominent state, that was also highlighted in a previous section (Figure 2.11C), the lipid is found between R384 and K388, interacting also with K387 and H391 (Figure 2.13A). PI4P provides many oxygen atoms bound either to its two phosphate moieties or to the inositol ring. This allows the lipid to interact with four charged residues of the AH at the same time. In this way it also occupies the possible interaction partners for other charged lipids like DOPS. The polyunsaturated acyl chain of PI4P wraps underneath the AH and fills the void created by the peptide. In one simulation, the PI4P lipid was accompanied by a cholesterol molecule that wedged its hydrophobic carbon chain into the same void (Figure 2.13A). It also transiently interacted with K387 through its oxygen atom on its head, keeping it associated to the AH. In the second state of PI4P interaction with the PIP5K AH, the lipid is found between K388 and H391 (Figure 2.13B). Although its acyl chains and glycerol backbone are positioned at a slightly different spot along the AH, its large inositol headgroup with the additional phosphate group is able to reach over towards both R384 and K387. In this way, PI4P manages to interact in a similar fashion with the AH as in the first state. In the third state, which is observed only in one of the three independent trajectories, PI4P is found at the opposite side of the AH with two DOPS lipids also interacting with the peptide (Figure 2.13C). Here, the interactions

between PI4P and the PIP5K AH are limited to K387, H391 and K394. In this third state a DOPS lipid is interacting with R384 and K388, but might be replaced by a PI4P lipid approaching from the same side. Besides these three stable states, PI4P was also observed to approach the AH from the N-terminus interacting only with R384, before ultimately shifting into the state 1 interaction (Figure 2.13A). This observation highlights the very stable interaction between the many negatively charged oxygen atoms with the positively charged side chains of PIP5K AH.

2.4 Discussion

Homeostasis of the lipid environment is a crucial task for cells, as quick adaptation of lipid compositions to specific circumstances can decide between life and death. Membrane property sensors in the form of AHs are found throughout different biological domains [31, 86]. They act as curvature sensors, actively deform membranes or coat lipid droplets [49]. Specific AHs have evolved to sense lipid packing defects in membranes. These AHs contain ALPS motifs that report on the accessibility of the hydrophobic membrane core from the lipid-water interface [23]. These AHs sense bulk membrane properties, whereas others are able to determine the presence of specific lipid species. In this chapter, insights into the sensing mechanism mediated by the AH of Opi1 and PIP5K were presented. Opi1 is the main gate keeper in phospholipid homeostasis found in *S. cerevisiae* [184]. The AH is anchored to the membrane through its hydrophobic interface, while the hydrophilic residues are free to interact with the solvent and the polar lipid atoms. While bound to the ER membrane it interacts with PA lipids, which serve as precursors for the *de novo* synthesis of other GPLs. Upon depletion of PA, the Opi1 AH dissociates from the ER membrane and shuttles into the nucleus to downregulate the transcription of genes involved in the *de novo* synthesis of GPLs [70]. When embedded in the ER membrane the AH is able to discriminate different lipid species based on their headgroups. Although highly abundant, lipids with PC headgroups do not enrich at the AH in the simulations reported in this chapter. Instead, only negatively charged lipid species are able to accumulate at the AH and only in specific regions. In the simulations two charged lipid headgroup species were tested: PA and PS. Both have a net charge of -1 in the used force field and only differ by their headgroup. Both lipids accumulate at either of two distinct motifs: the “3K motif” and “KRK motif”. The two motifs are found on opposing sides of the AH. They are composed of three positively charged residues, either lysine or arginine. The arrangement of these residues forms some sort of cage that provides enough space for the headgroup of PA, but not PS. While the latter is able to interact with the “3K motif” in two modes, either with its phosphate moiety or the oxygen on the serine headgroup, its residence time is significantly lower than that of PA. These two modes compete for the interaction with the lysines in the “3K motif” and allow PA lipids to displace them. Short lived interactions between the two competing interaction partners in PS lead to a lower efficiency in binding, while PA only presents one group for interaction.

The second studied AH in this chapter was that of PIP5K. The proteins function is to phosphorylate PI4P and produce PI(4,5)P₂. To do this, it must bind a molecule of

adenosine triphosphate (ATP) and hydrolyze it to adenosine diphosphate (ADP). The phosphate is transferred to the inositol group of PI4P to produce PI(4,5)P₂. A conserved aspartate residue in the catalytic center of PIP5K is responsible for this phosphorylation reaction [161,185,185]. This aspartate residue is found both close to the “specificity loop” of PIP5K, which folds into its AH upon membrane binding and the ATP binding domain [161]. The simulations in this thesis have shown that PI4P is able to interact with the PIP5K AH in three major interaction modes. Two of these modes only differentiate by the position of the glycerol backbone of the lipid. It might be wedged between two specific residues along one side of the AH. The inositol headgroup of PI4P is big enough to interact with the positively charged residues on top of the PIP5K AH independent of its position. Even in the third mode the inositol group interacts with almost the same residues as in the other two modes. These observations, in combination with the position of the AH in the overall PIP5K structure, highlight the specific function of this AH. While the soluble domain of PIP5K binds ATP and contains the catalytic aspartate residue, the responsibility of the AH is to bind a PI4P molecule and position its inositol group such that it can be phosphorylated. Future work could try to include the AH in the overall structure, to better understand if PI4P must be bound on either of the two sides to reach the catalytic aspartate residue.

The localization of the positively charged residues at the side of an α -helix may be a conserved motif that nature uses for lipid sensing with AHs. The current scientific literature does not include any examples of further AHs that sense lipid species with their specific arrangement of charged residues. Other proteins, especially membrane spanning transmembrane domains, are known to organize certain lipids around their perimeter to help with proper function [186,187]. In general, AHs appear to be the perfect protein sensor for sensing at the lipid-water interface. Their hydrophobic side allows the α -helices to insert into the membranes and anchor the remaining protein at the lipid-water interface. The hydrophilic side then is responsible for the actual sensing. The differences in the sequences of the two AHs and the positions of these residues along the side of the α -helices show how simple motifs can have distinct and complex functions. The sequence of amino acids and their location on the α -helix may be the toolkit used to sense the variety of lipids on a cellular membrane. Mutations of these specific residues should produce mutants with a different affinity, either lower affinity to its preferred lipid or increased affinity to a different lipid species.

Interaction of Opi1 AH with membranes

The investigations of the conformational dynamics of the Opi1 AH comprise computational and experimental work and were performed at the Max Planck Institute of Biophysics, Frankfurt am Main, Germany. The flotation experiments were carried out by Dr. Harald Hofbauer (Institute of Molecular Biosciences, University of Graz, Graz) and myself. The growth tests of different *S. cerevisiae* strains were carried out by Dr. Harald Hofbauer. All experiments were supervised by Prof. Dr. Robert Ernst (Institute of Medical Biochemistry and Molecular Biology, Saarland University, Homburg (Saar)). All simulations, data analysis and interpretations were carried out by myself under the supervision of Dr. Roberto Covino (Frankfurt Institute for Advanced Studies, Frankfurt) and Prof. Dr. Gerhard Hummer (Department of Theoretical Biophysics, Max Planck Institute of Biophysics, Frankfurt).

3.1 Introduction

To maintain proper membrane functionality living cells resort to lipid homeostasis. Different proteins adopt the lipid composition to the current needs, depending on the available lipid precursors and the general state and lifecycle of the cell. The Opi1 AH is a key regulator of GPL *de novo* synthesis [70, 184]. As mentioned previously, the protein is bound at the ER membrane when PA lipids are abundant. PA is the precursor for the *de novo* synthesis of GPLs. *S. cerevisiae* will keep producing various GPLs when PA is abundant. After depletion of PA, Opi1 must translocate into the cell nucleus and acts as a transcriptional repressor. It interacts with the transcriptional factors Ino2/Ino4 which are responsible for the active transcription of genes involved in the *de novo* synthesis of GPLs [184]. The interaction with Opi1 deactivates Ino2/Ino4 and ultimately downregulates the transcription. A machinery of proteins has to be produced to upkeep the *de novo* synthesis of GPLs. This sole production of the machinery costs energy. Cells must therefore tightly regulate this process as to not waste energy for the production of proteins when the precursor PA is depleted. In the previous section it has been shown how the fully folded AH of Opi1 is able to differentiate between different lipid headgroups when it is embedded into a lipid bilayer. Steric interactions between arginine and lysine residues allow the AH to recognize PA lipids and differentiate these from other charged lipid species, i.e., PS lipids.

While the folded function of the AH was studied in the previous chapter, the remaining lifecycle of the Opi1 AH remains unresolved and multiple questions remain unanswered: how does the AH decide to bind membranes? Does it fold into an α -helical conformation before or during the insertion into the membrane? What is the sequence of these processes?

Does the Opi1 AH follow the normal lifecycle of generic AHs? AHs are described to first be recruited to lipid bilayers by charged interactions, followed by insertion of hydrophobic residues into the bilayer. This tethering keeps the AH close to the membrane, where it is able to fold and insert into the bilayer [86, 188]. The work in this chapter tries to answer the above questions using both computational and experimental methods and proposes a possible mechanism how α -helices might fold on lipid membranes.

Aim. Atomistic MD simulations are required to understand the dynamics of tethering and folding of AH to lipid bilayers. No computational studies of the Opi1 AH are available at the time of writing. The aim of this project was to understand the membrane recruitment, membrane tethering through hydrophobic residues and subsequent folding and insertion of the AH. The results reveal a functional role of a key glycine residue. It plays a role in the initial binding process to membranes, as well as when the AH decides to unbind again.

3.2 Methods

The following section describes the used methods. It includes the set up and procedure of the MD simulations. The procedures to analyze different observable of interest are described. Additionally, two enhanced sampling methods are described.

3.2.1 Atomistic MD simulation

All systems in this chapter were set up using CHARMM-GUI [172, 175–178], if not mentioned otherwise. Two sequences of different lengths were used for all simulations (Table 3.1).

Table 3.1: Sequences of Opi1 AH used for atomistic molecular dynamics simulations.

No.	Residues	Length	Sequence
1	111–132	22	QKLSRAIAKGGKDNLKEYKLNMS
2	107–136	30	TNKRQKLSRAIAKGGKDNLKEYKLNMSIESK

The TIP3P water model was used with the CHARMM36 force field [189] and the GROMACS 5.1 software package [165]. Ions were added for a total concentration of NaCl 0.15 M. Each system was energy minimized using steepest descent followed by a multi-step equilibration. The system was first equilibrated in the canonical (NVT) ensemble with an integration timestep of 1 fs for 50 ps, maintaining a temperature of 310 K using the Berendsen thermostat [151] during the first two steps. Position restraints of $1000 \text{ kJ mol}^{-1} \text{ nm}^{-2}$ were applied to lipid heavy atoms in this initial phase. Restraints on backbone heavy atoms of the protein were lowered from 4000 to $2000 \text{ kJ mol}^{-1} \text{ nm}^{-2}$, whereas restraints on protein side chain heavy atoms were lowered from 2000 to $1000 \text{ kJ mol}^{-1} \text{ nm}^{-2}$. The system was equilibrated for further 300 ps by using an integration time step of 2 fs in the remaining four equilibration steps. The Berendsen barostat was used with a timestep of 1 fs to keep a constant pressure of 1 bar isothermal-isobaric (NPT) [151]. Lipid position restraints were decreased from 1000 to $0 \text{ kJ mol}^{-1} \text{ nm}^{-2}$, and protein position restraints were released. During the production simulations no restraints were applied on the system.

The Velocity Rescale thermostat [170] was applied with characteristic times of 1 ps to keep a constant temperature of 310 K. The Parrinello-Rahman barostat [171] was applied with semiisotropic pressure coupling and characteristic times of 5 ps to keep a constant and pressure of 1 bar. The thermostats were applied independently on protein, membrane, and solvent. Throughout all simulations non-bonded interactions were cutoff at 1.2 nm. All molecular dynamics systems reported in this chapter are listed in Table 3.2.

Table 3.2: Molecular dynamics simulation systems of Opi1 AH used in this chapter. All simulations start in an unfolded conformation above the membrane. The cumulative simulation time is listed for all repetitions of a single membrane composition or construct.

#	Sequence	Mutation	Lipid composition	Reps	Time (μ s)
1	22-mer	–	70% PC / 20% PA	1	18.8
2	22-mer	–	70% PC / 20% PA	6	10.0
3	22-mer	–	100% PC	10	22.8
4	30-mer	–	100% PC	5	20.2
5	30-mer	–	70% PC / 30% PE	5	21.8
6	30-mer	–	50% PC / 30% PE / 20% PA	5	22.2
7	30-mer	–	70% PC / 20% PA	5	22.7
8	22-mer	G120A	70% PC / 20% PA	5	71.5
9	22-mer	G120A/K121A	70% PC / 20% PA	2	31.5

3.2.2 Secondary structure calculations

The secondary structure assignment was performed with STRIDE [190]. STRIDE assigns the structure by considering the hydrogen bond energy, as well as backbone torsional angle information. STRIDE assigns one of six possible secondary structure codes to each amino acid. These six codes were mapped onto a simpler representation of three possibilities: α -helix (H), β -sheet (B) and coil (C) (Table 3.3). In every frame t the fraction $f(s, t)$ of a secondary structure s was calculated by dividing the number of structures s by the number of residues N : $f(s, t) = s(t)/N$.

Table 3.3: STRIDE secondary structure codes and mappings used for analysis.

STRIDE Code	Description	Mapping
H	α -helix	H
G	3_{10} -helix	H
I	π -helix	H
E	Extended conformation	B
T	Turn	B
B/b	Isolated bridge	C
C	Coil	C

3.2.3 Distance calculations between peptide and membrane

For every frame the distance between C_α atom and the membrane COM was calculated in the z direction. Out of all distances only the minimal distance was saved, resulting in a single distance per frame. The distances of the whole trajectory were binned and the

resulting histogram was smoothed by Gaussian kernel density estimation as implemented in scikit-learn [191].

3.2.4 Calculation of backbone dihedral angles

The intersection angle between two planes is called dihedral angle. A minimum of three points are required to define a geometrical plane. The heavy atoms of a protein backbone can be used as such points. A total of three backbone dihedral angles are defined for all amino acids: ω between C^{i-1} and N^i , ϕ between N^i and C_α^i , as well as ψ between N^{i+1} and C_α^{i+1} , with i begin the index of the residue. The planarity of the peptide bond allows ω to only adapt one of two angles: 0 deg and 180 deg. In contrast, ϕ and ψ are able to rotate freely (except for proline).

With \hat{n}_A and \hat{n}_B being the normal vectors of the two planes, the dihedral angle φ can be expressed as the dot product of both vectors divided by their lengths:

$$\varphi = \frac{\hat{n}_A \cdot \hat{n}_B}{|\hat{n}_A||\hat{n}_B|} \quad (3.1)$$

The ϕ and ψ dihedral angles were calculated for every residue in every frame using the implementation available in MDAnalysis [180,181].

3.2.5 Sequence similarity analysis

The full sequence of Opi1 from the reference strain S288C (systematic name: YHL020C; SGD ID: S000001012) was used to perform a search for similar sequences on NCBI using the non-redundant protein sequences (nr) database (dated 2020/07/16) with the blastp algorithm [192]. The maximum number of aligned sequences was set to 20000, with an expected threshold of 10 and a word size of 3. Sequences were scored using the BLOSUM62 matrix with no compositional adjustments [193]. Regions of low compositional complexity were masked to exclude spurious results. The search returned a total of 716 sequences. These sequences were filtered by only considering organisms in the class of *Saccharomyces* and sequences with at least 40% identity. A multiple sequence alignment (MSA) was performed on the remaining 92 sequences using COBALT [194]. A visual representation of the sequence similarity in the MSA was created with WebLogo [195]. The MSA residues 429–455, corresponding to residues 109–136 in the original Opi1 sequence, were used for the visual representation.

3.2.6 Biased molecular dynamics simulations

Two different enhances sampling methods were used to accelerate the sampling and complete the folding of the Opi1 AH.

Adiabatic bias molecular dynamics

Adiabatic bias molecular dynamics (ABMD) simulations were run to fold the Opi1 AH, as implemented by PLUMED [196,197]. ABMD connects different conformational endpoints

by addition of a ratchet-and-pawl restraint to the system. All fluctuations that move the system closer to the desired value of a collective variable (CV) are left untouched. A harmonic potential is applied onto all fluctuations that would decrease the CV. The system is allowed to proceed towards the CV using thermal fluctuations, but kept from distancing itself with an external time-dependent biasing potential $V(t)$:

$$V(\rho(t)) = \begin{cases} \frac{\kappa}{2}(\rho(t) - \rho_m(t))^2, & \rho(t) > \rho_m(t), \\ 0, & \rho(t) \leq \rho_m(t), \end{cases} \quad (3.2)$$

where

$$\rho(t) = (S(t) - S_{\text{target}})^2. \quad (3.3)$$

is the distance S along the CV of the instantaneous configuration of the system with respect to the target value S_{target} , and

$$\rho_m(t) = \min_{0 \leq \tau \leq t} \rho(\tau) + \eta(t) \quad (3.4)$$

is the minimum distance along S that has been reached until time t . $\eta(t)$ is additional white noise acting on the bias position. The damping factor κ was varied in some of the simulations.

The correct choice of CV is important for biased simulations. The *alpharmsd* CV probes the α -helical content of a protein. Six consecutive amino acids are considered to be part of a ‘‘block’’. A block is superimposed on an ideal α -helix and the root-mean-square deviation (RMSD) is calculated. The RMSD is transformed to a value between 0 and 1 with a switching function and all possible six residue blocks are summed:

$$s(r) = \sum_i \frac{1 - \left(\frac{r_i}{r_0}\right)^6}{1 - \left(\frac{r_i}{r_0}\right)^{12}}, \quad (3.5)$$

with $r_0 = 0.08$ nm. The harmonic bias applied by ABMD acts on the CV, which is just the number of blocks that are in an α -helical conformation.

Steered molecular dynamics

Steered MD simulations add a time-dependent, harmonic potential $V(s, t)$ on a CV [198]. In contrast to ABMD, the harmonic potential always acts on the system, even if it is moving towards the CV through thermal fluctuations. The bias is defined for t_n and is linearly interpolated for all values between t_0 and t_n :

$$V(s, t) = \frac{\kappa(t)}{2}(s - s_0(t))^2, \quad (3.6)$$

with s being the instantaneous CV, $\kappa(t)$ is the interpolated spring-constant and $s_0(t)$ is the target value at time t . The bias was applied sequentially one residue at a time. t_n was set to 1 ns. The dihedral angles ϕ and ψ were used as CV for steered MD.

3.3 Results

In the following section, the results of atomistic MD simulations of Opi1 AH are summarized. It starts with the generation of unfolded conformations for two different lengths of the Opi1 AH. The structures of these unfolded conformations are subsequently used in membrane binding simulations, where tethering and partial folding of the unfolded peptides is observed. The observation of partial folding leads to an analysis of the sequence similarity in *Saccharomyces* and experimental work performed by Dr. Harald Hofbauer. After the functional role of G120 is established in experiments, additional membrane binding simulations containing single and double point mutations are evaluated. Finally, two different biasing schemes are used to complete the folding of the AH while tethered to lipid bilayers.

3.3.1 Generation of unfolded Opi1 AH conformations

Unfolded conformations of the Opi1 AH peptide were generated by running high temperature molecular dynamics simulations. To this end, two sequences of the peptide with different lengths were used: a 22 amino acid stretch that encompasses only the AH and another 30 amino acid stretch that pads the AH by four residues on each side (Table 3.1). The two sequences were modeled as an ideal α -helices with UCSF CHIMERA [162], inserted into their respective simulation box and solvated with TIP3P water. In the following these two constructs will be referred to as 22-mer and 30-mer indicating their sequence length. The set up and simulation procedure for the 22-mer AH was described previously [182]. In brief, the 22-mer AH was modeled as ideal α -helix and simulated using replica-exchange molecular dynamics (REMD) at 25 different temperatures (310 K to 351 K) for 121.5 ns each, accumulating a total of 3037.5 ns. These biased simulations at elevated temperatures explored different conformations of the 22-mer peptide, but ultimately unfolded the structure. Multiple fully and partially unfolded α -helical structures were assessed visually and extracted for further use.

The 30-mer peptide was simulated in a single unbiased instance at 400 K for 612 ns. Figure 3.1A shows the simulation box setup with the α -helical Opi1 30-mer AH centered in a cubic box surrounded by water and ions. Simulating this construct at these high temperatures led to immediate destabilization of protein secondary structures. A fully unfolded and almost stretched out 30-mer peptide is shown in Figure 3.1B.

The secondary structure was assigned using the STRIDE software. For every time step the structural ratio of one of three states, α -helix (H), β -sheet (B) and coil (C) was determined and plotted as a function of time (Figure 3.2). The 30-mer loses its α -helical secondary structure in the first tens of nanoseconds and adopts both coiled and β -sheet secondary structures for the remaining simulations. A single fully unfolded structure was selected by only considering structures with 29 non-structured residues, and

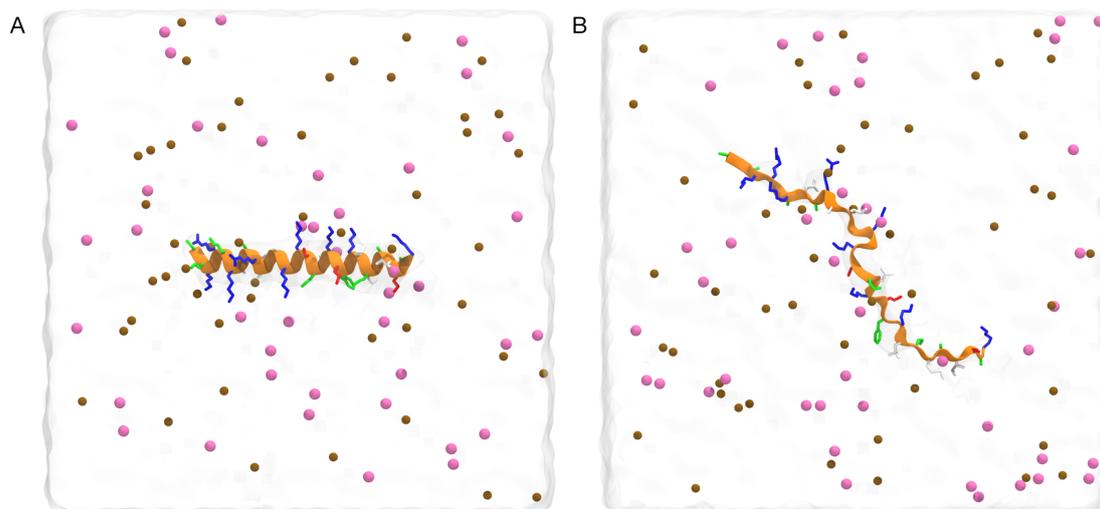


Figure 3.1: Simulation snapshots of Opi1 30-mer at 400 K. (A) Initial structure of Opi1 30-mer AH in the cubic simulation box surrounded by water and ions. (B) Unfolded structure after 270 ns of simulation at 400 K. Colored beads show sodium (brown) and chloride (pink).

N- and C-termini that are separated by at least 60 Å. This structure was used for the 30-mer simulations described in the following sections.

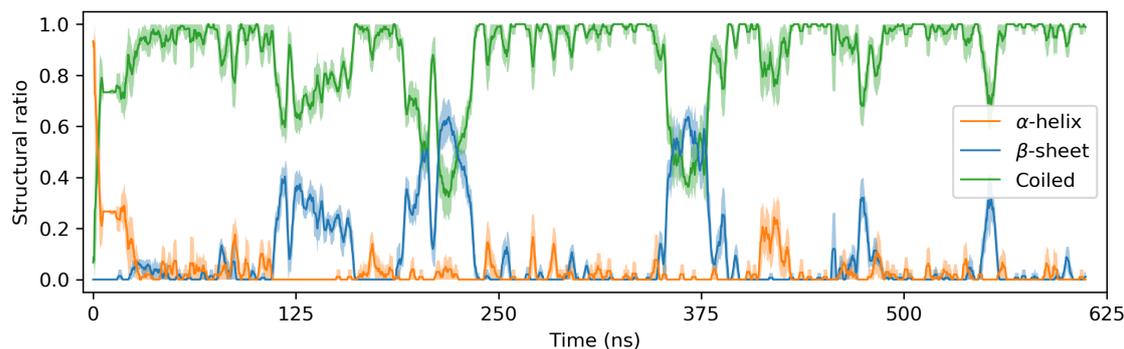


Figure 3.2: Structural composition of Opi1 30-mer peptide at 400 K. The secondary structure of each amino acid was evaluated at every timestep and ratios were calculated. The colored solid lines show α -helix (orange), β -sheet (blue) and coiled (green) secondary structures. All data was window averaged over 1.25 ns and shaded areas indicate the standard deviation.

3.3.2 Recruitment and tethering to membranes

To test whether a specific lipid composition is required to enhance insertion and folding of the Opi1 AH into the membrane, systems of different membrane compositions were investigated. All membranes were composed of 50% DOPC, with the remaining 50% of lipids divided between POPC, 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoethanolamine (POPE) and 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphate (POPA). While the 22-mer peptide was simulated only with PC and PA headgroups, the 30-mer peptide was also simulated in the presence of PE headgroups. The exact membrane compositions are listed in Table 3.4. Although membranes without PA headgroups do not bear any charges, the size of the headgroup affects the biophysical properties of the lipid bilayer [199]. PE headgroups are

the smallest uncharged GPL headgroup and contribute to the exposure of hydrophobic patches on the membrane surface.

Table 3.4: Membrane compositions used for binding simulations with unfolded Opi1 peptides. The 22-mer simulations were performed with compositions A and D, whereas the 30-mer peptide simulations were performed with all four membrane compositions.

#	A	B	C	D
	50% DOPC	50% DOPC	50% DOPC	50% DOPC
	50% POPC	30% POPE	30% POPE	30% POPC
		20% POPC	20% POPA	20% POPA

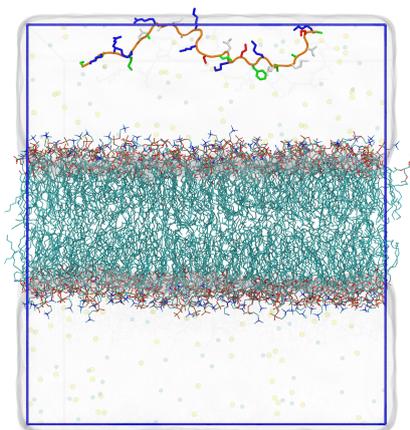


Figure 3.3: Initial structure of Opi1 30-mer binding simulations. The unfolded Opi1 AH (orange) was placed at 50 Å from the membrane center. Lipids are shown as sticks in cyan (carbons), red (oxygens), ocher (phosphate) and blue (nitrogen). Ions are shown as transparent beads and water as transparent surface. The borders of the simulation box are shown for clarity.

Lipid interactions of 22-mer Opi1 peptide

A simulation box with the 22-mer Opi1 peptide was set up to study its binding to lipid bilayers. An unfolded 22-mer Opi1 peptide was positioned 50 Å from the 70% PC and 30% PA membrane (no. 1 in Table 3.2). In addition to this single simulation, a total of 16 additional simulations were set up: six simulations in the presence and ten in the absence of 30% PA lipids (no. 2 and 3 in Table 3.2). The systems were set up using ten different conformations of the Opi1 22-mer peptide, some fully unfolded and others with a remaining α -helical secondary structure. The results were previously described [182], but will be summarized again in the following paragraph.

In all simulations the peptides were recruited to the lipid-water interface in less than 100 ns. The initial recruitment of the peptides is independent of its secondary structure and appears to only depend on the membrane charge. The peptides lysine and arginine residues mediated the ionic interactions with the phosphate moieties in the PA lipid headgroups. Once recruited to the membrane, the 22-mer peptide did not dissociate from the lipid-water interface in the presence of PA headgroups. In the absence of PA, the peptides continuously associated and dissociated from the membrane. Tethering to the membranes was only observed in the six systems with membranes containing PA lipids. The two

hydrophobic residues L113 and I117 tethered the 22-mer peptide to the membrane. Either of the residues was able to tether the peptide to the membrane and no preference for one over the other was observed. All 22-mer peptides lost their α -helical secondary structure, if any was present at the beginning of the simulation.

Lipid interactions of 30-mer Opi1 peptide

Additional simulations with the Opi1 30-mer were set up to test whether the above observations are biased by the sequence and/or length of the peptide. In addition to the two previous membrane compositions, two new compositions containing PE headgroups were used (Table 3.4). In every simulation box a lipid bilayer of a different composition was set up. The unfolded Opi1 30-mer AH from the previous section (Section 3.3.1) was placed at a distance of 50 Å from the membrane center (no. 4–7 in Table 3.2). The box was solvated with water and 0.15 M NaCl (Figure 3.3). For every membrane composition five independent repetitions were set up, for a total of 20 systems. Each system was equilibrated separately and run as production simulation for up to 4.5 μ s at 310 K. The minimal distance between the C_α atoms of the unfolded peptide and the membrane center was calculated and is shown for every membrane composition in Figure 3.4.

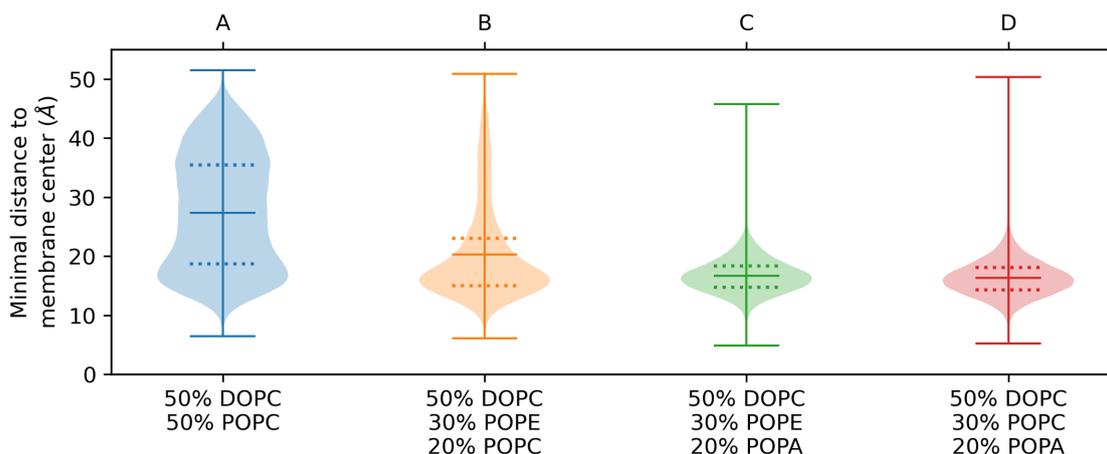


Figure 3.4: Distance distributions of backbone C_α atoms to membrane center for four different membrane compositions. Only the minimum distance out of all 30 backbone C_α atoms was considered for every frame. The distributions of each membrane composition (A–D) contains data for all five replicas. The membrane compositions are listed in Table 3.4. The mean and extreme values are shown with horizontal lines. The 25% and 75% quantiles are indicated as dotted lines.

On average, the 30-mer peptide has the highest distance to 100% PC membranes (Figure 3.4A). The addition of 30% PE leads to a reduction of the mean distance from 27.4 Å (100% PC) to 20.3 Å (Table 3.5). Further addition of PA to the membrane again decreases the mean distance to 16.7 Å. Surprisingly, running a simulation with only PC and PA headgroups reveals that the sole presence of PA is sufficient to keep the Opi1 AH close to the lipid-water interface (Figure 3.4D). The high distance of more than 40 Å observed for all membrane compositions is a result of the initial set up of the systems. As described above, the peptide was initially placed at 50 Å from the membrane center.

The minimal distances of the individual trajectories are shown as a function of time in

Table 3.5: Mean, minimal and maximal distances between C_α atoms of unfolded Opi1 AH and membrane center for all four membrane compositions (A-B) as listed in Table 3.4.

#	Mean distance (Å)	Min. distance (Å)	Max. distance (Å)
A	27.4	6.5	51.5
B	20.3	6.1	50.9
C	16.7	4.8	45.8
D	16.4	5.2	50.4

Figure 3.5. The Opi1 AH constantly associates and dissociates to membranes with 100% PC lipid headgroups in three of the five systems (Figure 3.5A). In two cases, the 30-mer peptide tethers itself to the membrane and remains at close distance (blue and orange traces in Figure 3.5). In systems with 30% PE the 30-mer peptide associates and dissociates from the membrane at first, but remains tethered in all five independent repetitions after 3 μ s. The addition of PA headgroups leads to the immediate recruitment of the AH to the lipid-water interface in all five independent repetitions (Figure 3.5C). Dissociation is not observed with this membrane composition containing lipids with both PE and PA headgroups. Similarly, and as already described above, the sole presence of lipids with PA headgroups is sufficient to recruit the AH to the lipid-water interface in tens of nanoseconds and keep it close to the membrane for the whole simulation.

Tethering to the membrane via hydrophobic residues

Visual inspection of the trajectories revealed that the 30-mer peptide tethers itself to the membrane with either L113 or I117, similar to the 22-mer peptide described above. PA containing membranes recruit the unfolded Opi1 AH in dozens of nanoseconds to the lipid-water interface, where the peptide is kept close by electrostatic interactions. Most GPLs are partially charged, whereas PA bears 1–2 negative charges on its small headgroup. The smaller PE headgroups expose both the partially charged phosphate moiety and hydrophobic parts of the bilayer. These partial charges recruit the peptide to the membrane and hydrophobic residues can insert into the membrane core when transient hydrophobic defects are exposed at the lipid-water interface. To quantify the preference of one hydrophobic residue over the other, the distance between the residues' C_α atom and the membrane center was calculated. The distance density for the two hydrophobic residues L113 and I117 in four different membrane compositions is shown in Figure 3.6.

In 100% PC membranes the two residues are predominantly found in the untethered state and a big distance from the membrane center (Figure 3.6A). In one case the unfolded peptide does indeed tether itself to the membrane, as can be seen by the L113 density around 20 Å. As described in the previous section, the addition of PE reduces the average distance between the unfolded 30-mer peptide and the membrane. The effect of PE can also be observed in the distance densities of L113 and I117 (Figure 3.6B). Here, both residues are able to insert into the membrane and thus tether the unfolded peptide (distances less than 20 Å). I117 is able to tether itself at closer distances than L113. The addition of lipids with PA headgroups further decreases the average distance between the unfolded 30-mer

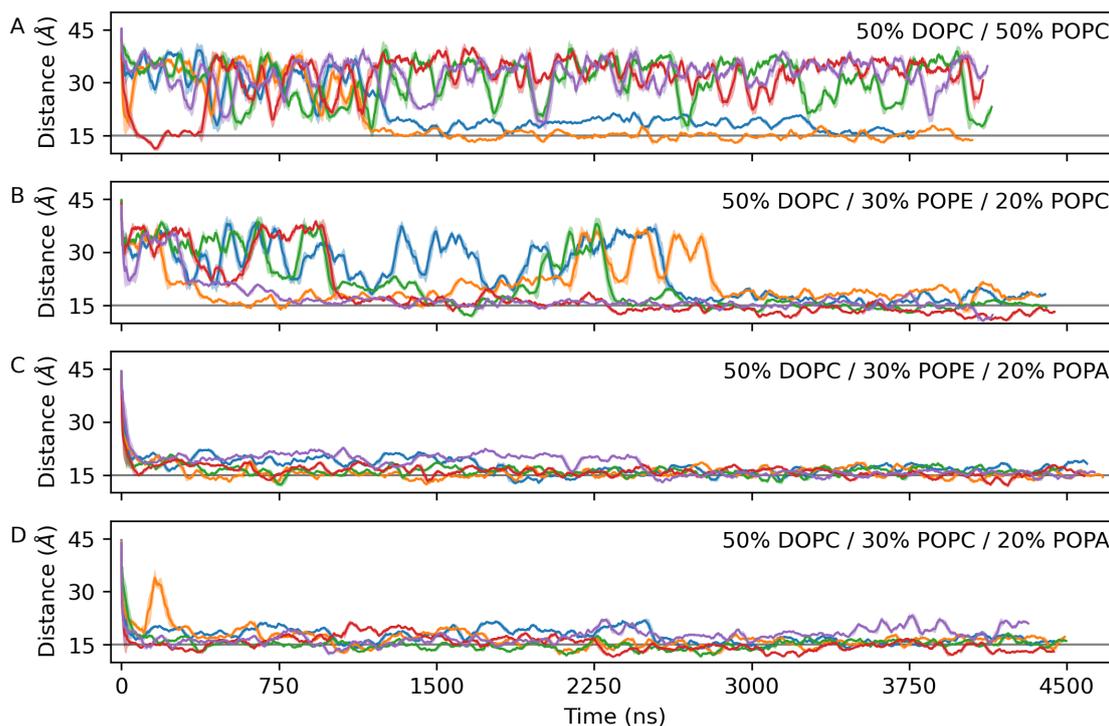


Figure 3.5: Distances of backbone C_{α} atoms to membrane center for four different membrane compositions as a function of time. (A) The Opi1 AH continuously dissociates from membranes with 100% PC headgroups. (B) Addition of PE headgroups allows the peptide to tether at the lipid-water interface. (C) The peptide is recruited and tethered immediately in presence of PE and PA headgroups. (D) The sole presence of PA headgroups is sufficient to tether the peptide to the membrane. In each panel, five traces from independent simulations are shown as colored lines. Solid colored lines show the 50 frame average and shaded areas indicate the standard deviation. The horizontal gray line indicates a distance of 15 Å.

peptide and the membrane. At the same time the presence of PA headgroups decreases the average distance of L113 to 25 Å (Figure 3.6C). The same holds true for I117, where a bimodal distribution can be observed. Whereas the average distance between unfolded peptide and the membrane does not change in a membrane only composed of PC and PA headgroups, the distance distribution of L113 and I117 does (Figure 3.6D). The distance distribution of L113 changes to a bimodal one, with distances similarly close as seen for I117. I117 shifts to higher average distances and its distribution looks more similar to that of L113. Therefore, in a mixed PC/PA membrane the preference of one over the other residue disappears, while I117 is found closer to the membrane more often in PE containing membranes.

Experimental validation of I117 tethering

To validate the functional role of L113 and I117 in membrane tethering, *in vitro* experiments using a truncated Opi1 construct were performed. The experimental set up and results were already described earlier [182]. In brief, the 180 residue construct was overexpressed in *E. coli* cells and subsequently purified. The protein quality was verified by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) and size exclusion chromatography.

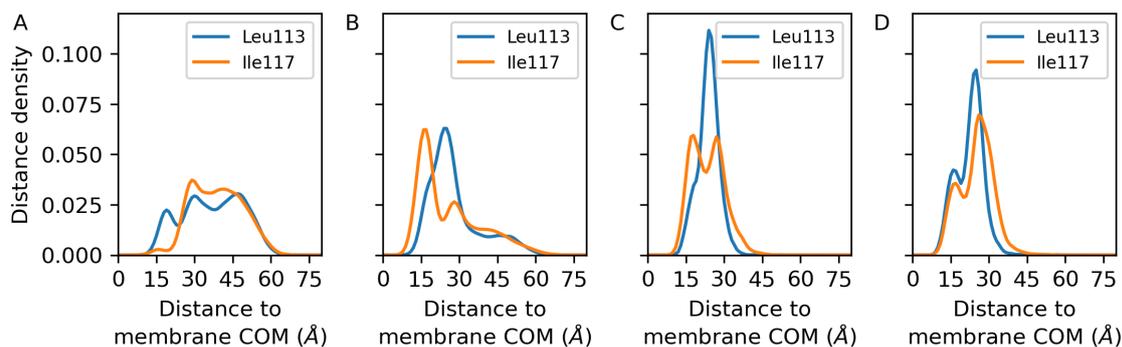


Figure 3.6: Distance distribution of L113 and I117 C_{α} atoms to the center of membranes of various compositions. (A-B) The two residues are found at large distances in membranes containing only PC or some amounts of PE lipids. (C-D) Inclusion of PA lipids is sufficient to tether both hydrophobic residue to the membrane. The exact membrane compositions are listed in Table 3.4.

Liposomes were extruded with either 100% PC or 80% PC and 20% PA lipids. The purified Opi1 construct was mixed with liposomes and incubated to allow for protein-lipid binding. Subsequently, a sucrose gradient was layered on top, followed by centrifugation in an ultracentrifuge at 240 000 $\times g$. The density difference between the sucrose gradient and the liposomes forced the liposomes to float to the top of the gradient. Unbound protein does not float to the top and is thus expected to be found at the bottom of the gradient. By separating the gradient in four equally sized parts it was possible to estimate the amount of protein that was bound to liposomes.

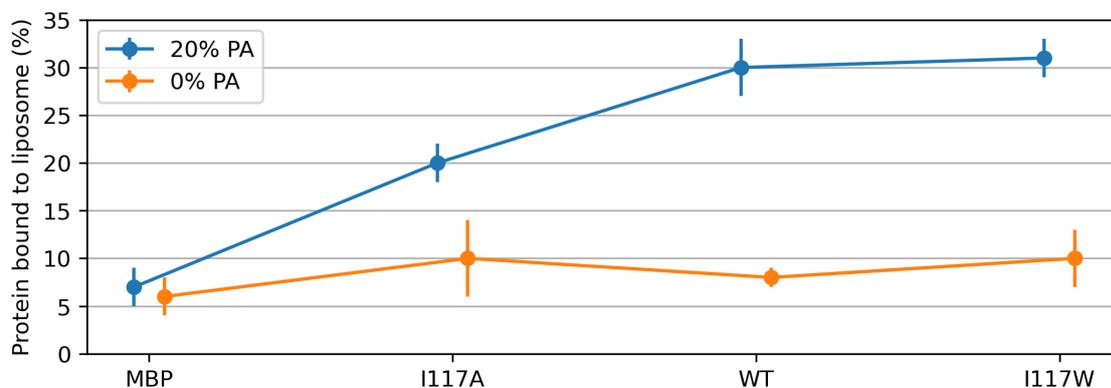


Figure 3.7: Bulky residues favor tethering of Opi1 AH in lipid membranes. Flotation assays of the Opi1 AH with membranes containing 20% PA (blue) or 0% PA (orange). Maltose-binding protein (MBP) was used as a negative control. Mutating I117 to alanine leads to a binding reduction in 20% PA membranes, whereas the I117W mutation does not show significant differences. The binding to uncharged (0% PA) membranes is unaffected by the mutations. Each data point shows the average out of 3 experimental repetitions with the error bars showing the standard deviation. The data was previously published in [182]. The figure was reproduced for this thesis.

Figure 3.7 shows the quantification of the *in vitro* experiments described above. MBP was used a negative control to account for its unspecific binding of the proteins to liposomes. 5–7% of bound protein can be attributed to unspecific binding to liposomes, as shown with the MBP data independent of the membrane composition. In the absence of PA the three Opi1 constructs show a small increase of bound protein in comparison to the MBP control.

The two mutants do not show any differences to the wild type construct. In contrast, the presence of 20% PA shows a clear difference of bound protein depending on the size of the residue found at position 117. 30% of the wild type construct is bound to liposomes, but only 20% of the I117 to alanine (I117A) mutant. This decrease in binding can be attributed to the hydrophobicity of the residue and its potential to anchor the protein in the membrane. Intriguingly, mutations to the most hydrophobic residue tryptophan (I117W) do not lead to a significant increase in binding capacity.

3.3.3 Secondary structure formation at lipid-water interfaces

The Opi1 30-mer peptide forms secondary structures when close to lipid-water interfaces. The visual inspection from the previous section has revealed that the unfolded peptide tethers itself to the membrane through its hydrophobic residues L113 and/or I117. The tethering does not require the peptide to be fully unfolded and can occur after partial secondary structure formation.

Transient secondary structure formation at membranes of varying compositions

The secondary structure composition of all five repetitions in each of the four membrane compositions from the previous Section 3.3.2 was calculated to quantify the different observed structures (Figure 3.8).

In the absence of either PE or PA headgroups, the 30-mer peptide shows a preference to predominantly form β -sheet structures (Figure 3.8A). In these simulations (repetitions 1, 4 and 5 of A) more than 50% of residues form β -sheets. This secondary structure formation is reversible, allowing the peptide to explore further conformations. The quantification in Figure 3.8 does not account for the difference of the observed structures. Figure 3.9 shows snapshots from these simulations, with the exact frames indicated as vertical line in Figure 3.8. Visual inspection of the β -sheet structure, as indicated in the first repetition in Figure 3.8A, shows a coiled conformation of the Opi1 30-mer, folded onto itself (Figure 3.9). In contrast, the fourth repetition in Figure 3.9A clearly shows a β -sheet that is not tethered to the membrane but kept close at its lipid-water interface through electrostatic interactions. The lack of secondary structure as shown in the second repetition in Figure 3.8A represents the conformation and tethered state that the unfolded 30-mer is usually found in throughout most of these independent simulations (Figure 3.9A, second repetition). In this case, the peptide is tethered to the membrane through L113 and Y127.

The addition of lipids with PE headgroups reduces the propensity of the 30-mer peptide to form β -sheets (Figure 3.8B). An interesting “circular” conformation can be seen in the fifth repetition in Figure 3.9B, where the unfolded peptide appears to be folded upon itself in a circular fashion, while tethered to the membrane through I117. The formation of an α -helical secondary structure can be observed predominantly in PA containing membranes (Figure 3.8C-D). Additionally it is observed at the beginning of the first 100% PC simulation (Figure 3.8A, first column). In most of the cases the α -helix forms at the C-terminal end of the Opi1 30-mer peptide, as illustrated in the snapshots in the fourth and fifth repetitions in

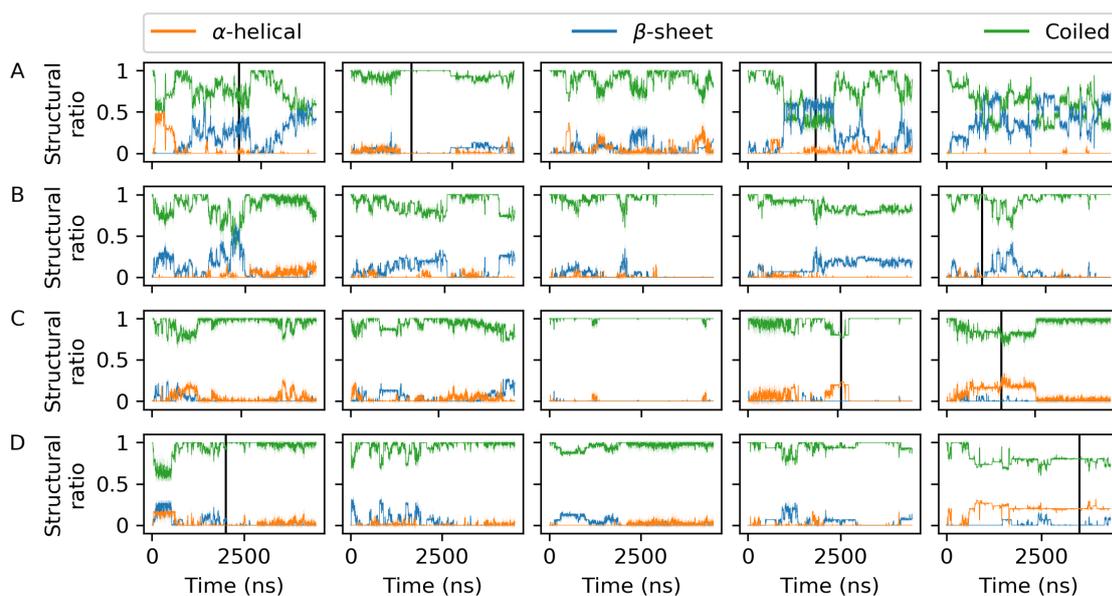


Figure 3.8: Structural composition of Opi1 30-mer peptide in simulation boxes with membranes of varying lipid compositions. Every row shows a single membrane composition with columns indicating one of the five independent repetitions. The exact membrane compositions are listed in Table 3.4. (A) Membranes with 100% PC show a high percentage of β -sheet secondary structures. (B) The addition of 20% PE reduces the amount of observed β -sheet secondary structures. (C) Membranes containing both PE and PA show some propensity of α -helices. (D) The presence of PA headgroups is sufficient to reduce the structural ratio of β -sheet and increase the α -helix propensity. The secondary structure of each amino acid was evaluated at every timestep. The colored solid lines show α -helix (orange), β -sheet (blue) and coiled (green) secondary structures. The data was window averaged over 1.25 ns and shaded areas indicate the standard deviation. Vertical black lines indicate the frames that were used for snapshot renders in Figure 3.9.

Figure 3.9C and the fifth repetition in Figure 3.9D. This C-terminal helix always includes Y127. In the fourth and fifth repetitions of simulation C, the C-terminal helix is found outside of the membrane, whereas it embeds into the lipid-water interface for hundreds of nanoseconds in the fifth repetition (Figure 3.9D). The 30-mer peptide is held close at the lipid-water interface after formation of the short 2-turn α -helical stretch by the peptides charged arginine and lysine residues. The insertion of the short 2-turn α -helix happens in an instant as soon as a suitable hydrophobic patch appears on the membrane surface. The insertion of this short α -helix is not stable, the helix dissociates and unfolds after some time. The snapshot in the first repetition of D shows another conformation where the peptide is collapsed onto itself, without any apparent secondary structure (first column of Figure 3.8D and first repetition of Figure 3.9D). These observations highlight the propensity of the Opi1 30-mer to explore different conformations at lipid-water interfaces. Furthermore, folding of the AH at membrane surfaces does not necessarily include the insertion of the folded helix into the membrane. The lipids in the membrane must be pushed aside by the protein to allow for proper insertion.

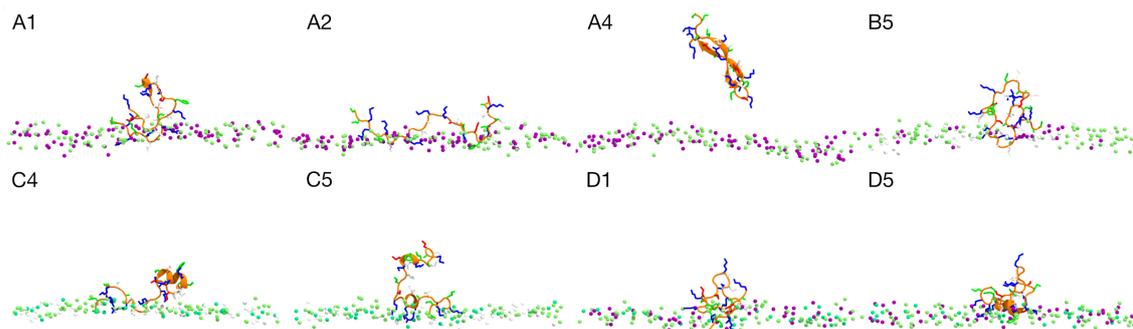


Figure 3.9: Snapshots of Opi1 30-mer peptide at the lipid-water interface of membranes with different lipid compositions. The letters of the labels in the top left corners differentiate membrane compositions as in Table 3.4, whereas the numbers indicate the simulation repetition the snapshot was taken from (columns in Figure 3.8). The Opi1 30-mer AH is colored as in Figure 3.3. Lipid phosphate moieties are shown as colored beads: DOPC (green), POPC (magenta), POPE (white) and POPA (lime green). Water and ions are not shown for clarity.

Folding and insertion of a short α -helical turn

The stable insertion of a short α -helix was observed in a single simulation (no. 1 in Table 3.2) and already described earlier [182]. The simulated 22-mer peptide is first recruited to the lipid-water interface, after which its two hydrophobic residues L113 and I117 tether it to the membrane. After 1.8 μ s of simulation, I117 is pushed out from the membrane and an α -helix spanning 1.5 turns is formed between residues L113 and I117, and subsequently embeds into the membrane. Figure 3.10 shows a representative structure of the partially folded AH after 2.5 μ s. The two hydrophobic residues I113 and L117 are shown to tether the short α -helix at the lipid-water interface, whereas K112 and R115 snorkel towards the solvent. The α -helical secondary structure does not continue past A118 which is followed by K119, G120 and K121. These residues, especially G120, appear to slow down the folding of the remaining 22-mer peptide.

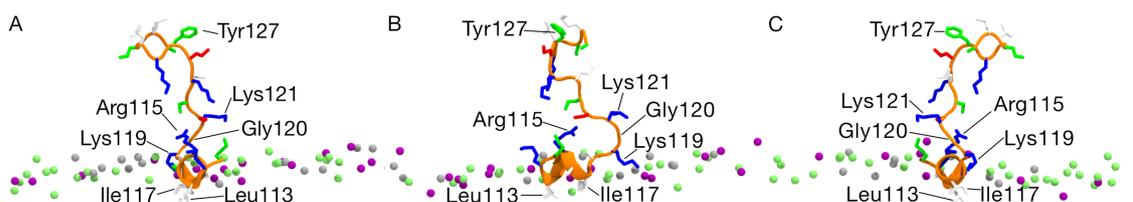


Figure 3.10: Partially folded Opi1 22-mer AH from unbiased simulations. The panels show the same snapshot after 2.5 μ s of simulation from different angles. (A) Front view looking onto the N-terminus. (B) Side view from N- to C-terminus (left to right, respectively). (C) Back view looking onto the C-terminus. Residues of interest are annotated in every panel.

The dihedral angles ϕ and ψ are plotted as a function of time in Figure 3.11 and follow the whole 18 μ s simulation. The dihedral angles of a proteins backbone adopt distinct ϕ and ψ angles in α -helices: -90 deg to -35 deg and -70 deg to -15 deg, respectively. The ϕ dihedral angles of L113–A116 adopt values of around -60 deg initially in the simulation, whereas their ψ angles adopt positive values. All five residues are α -helical after 2.5 μ s. Once formed, the α -helix remains folded for a total of 10 μ s. After around 12 μ s of simulation time the short α -helix dissociates from the membrane and subsequently unfolds in solution.

Intriguingly, most ϕ dihedrals remain close to the α -helical values, whereas the ψ dihedrals are again found in values above 0 degrees. This observation highlights that insertion of the Opi1 AH is not permanent. The biological function of Opi1 requires the AH to dissociate from the ER membrane upon depletion of PA lipids.

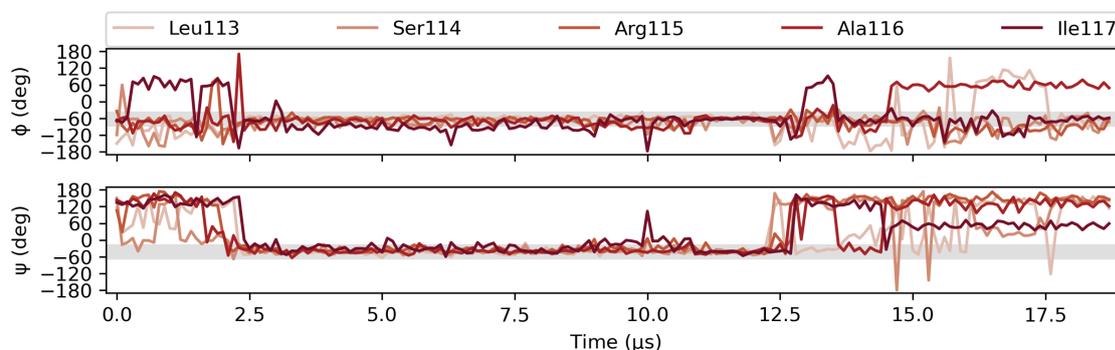


Figure 3.11: Backbone dihedral angles ϕ and ψ showing the folding of a 1.5 turn α -helix spanning residues L113–I117 in the Opi1 22-mer peptide. A 5 residue α -helix is formed and stably inserted into the membrane for 10 μ s from 2.0 μ s to 12.0 μ s. Each line shows the dihedral angle of a different residue. The gray area indicates the common ϕ (-90 deg to -35 deg) and ψ (-70 deg to -15 deg) dihedrals found in α -helices.

3.3.4 G120 mediates the AH folding and unfolding

Although initial folding of the α -helix was observed after 2 μ s, subsequent folding of the Opi1 AH did not occur after an extensive atomistic equilibrium simulation. While the lack of conformational sampling might be an issue, a detailed look at the amino acid sequence of the Opi1 AH offers another explanation. Folding of the first 1.5 turns stops right before the three residues 119–K GK–121, in the following referred to as “K GK motif”. The glycine amino acid is known to be a so called “helix breaker” because of its high conformational flexibility [200] and therefore unlikely to be found in an α -helical structure. The bare presence of G120 suggests a functional role in the Opi1 AH. And as described earlier (see Chapter 2), the folded and membrane-inserted AH has a tendency to transiently kink at G120. Additionally, the two lysine residues flanking G120 must end up on different sides of the AH after folding. This allows to question: how would such a sequence fold into a helical turn? When folding on the membrane, one of the lysine residues has to be dragged through the lipid-water interface. Another possibility is that each turn must form outside the membrane and subsequently slide into the interface.

Similarity analysis of amino acid sequences

The full length 404 amino acid sequence of Opi1 was used to search similar sequences in the NCBI non-redundant protein sequences (nr) database. A total of 92 sequences were considered from the class of *Saccharomycetes* after filtering. The exact filtering procedure is described in the methods section. The sequence similarity is shown in Figure 3.12. The “K GK motif” is conserved across all 92 sequences. In a few cases, the glycine is substituted by a serine residue and instead glycine is found at position 122. Lysines are found at both

positions 119 and 121 and can be substituted with arginines. The high sequence similarity suggests a functional role of the “KGK motif”.

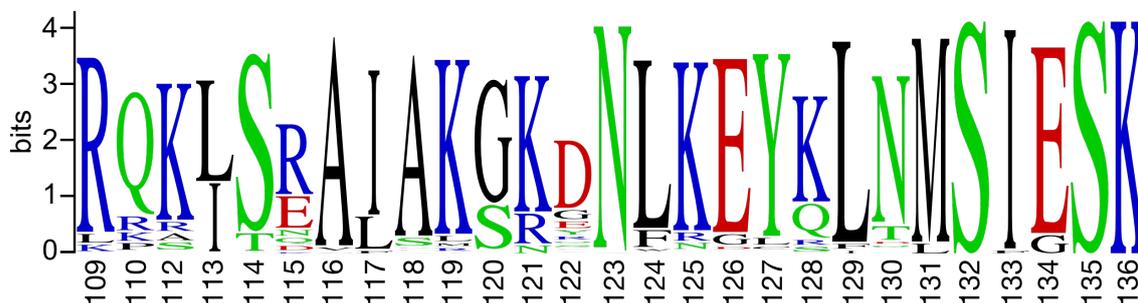


Figure 3.12: Sequence similarity of Opi1 AH and similar sequences. The letter size represents the conservation of a residue at the corresponding position. The consensus sequence is the same as used for all simulations (Table 3.1). One letter codes of hydrophobic residues are shown in black, positively charged in blue, negatively charged in red and the remaining polar residues in green. The figure was generated with WebLogo and adapted manually [195].

Experimental validation of G120 function

The functional role of G120 was assessed by Dr. Harald Hofbauer using a simple *in vivo* experimental setup. As mentioned in the main introduction of this work, *S. cerevisiae* is able to perform *de novo* synthesis of inositol. This process is controlled by a feedback loop directed by Opi1. Preventing the protein from regulation will lead the cells to produce inositol. As accumulation of inositol is toxic, cells will secrete unused inositol into the environment. A second tester strain, auxotrophic for inositol, can be grown alongside with *S. cerevisiae* to visualize the secretion. This tester strain grows in a red halo around yeast colonies that secrete inositol. The exact experimental description can be found elsewhere [157]. Briefly, *S. cerevisiae* was cultured on minimal media lacking inositol. After a day of growth, the tester strain was sprayed on top of the cultures and incubated for two more days. In addition to wild type *S. cerevisiae*, strains with point mutations at position 120 were used to test the functional role of G120 in the Opi1 AH.

Figure 3.13 shows photographs of the different strains cultivated on inositol lacking media. The wild type strains have a functional Opi1 protein, do not produce an excess of inositol and therefore do not secrete it to the outside. Increasing the hydrophobicity of the AH by substitution of G120 for a tryptophan, the biggest and most hydrophobic natural amino acid, leads to a clearly visible red halo around the yellow *S. cerevisiae* colonies (Figure 3.13). This visualizes that the Opi1 AH is retained at the ER membrane after PA lipid depletion. As mentioned in the introduction of this work, abundance of inositol depletes the PA levels in the ER membrane and should lead to the translocation of the Opi1 AH into the nucleus, where it downregulates the transcription of proteins involved in the *de novo* synthesis of GPLs. A tyrosine was introduced at position 120 to test if the hydrophobicity is of importance. Tyrosine is more polar than tryptophan and similar in size. Figure 3.13 shows a faint halo around the colonies indicating the secretion of inositol by *S. cerevisiae*. This result shows that hydrophobicity is not the sole explanation for a PA-independent tethering of the Opi1 AH to the ER membrane. To further exclude the

effect of hydrophobicity, glycine was mutated to alanine. The two amino acids differ by the substitution of glycines' single hydrogen atom side chain to a methyl group (CH_3). This substitution results in a minor hydrophobicity increase, but reduces the conformational flexibility and therefore increases the chances of an α -helical secondary structure. Indeed, the *in vivo* experiment shows a small red halo around the colonies, similar to that of the G120Y strain.

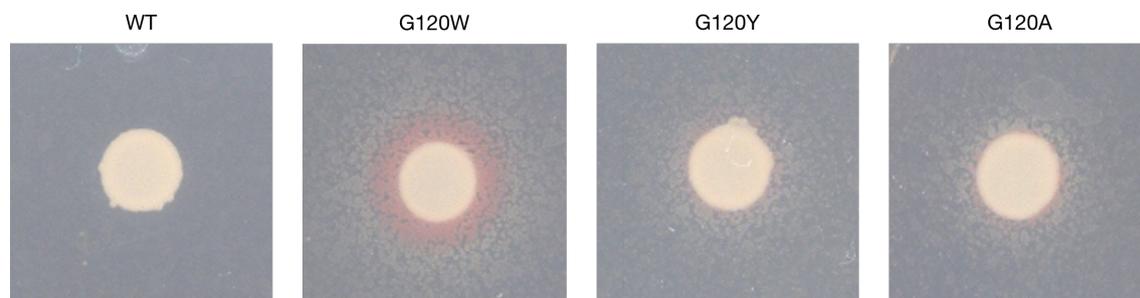


Figure 3.13: Four different *S. cerevisiae* strains grown on inositol lacking media. Point mutations in the Opi1 AH lead to the secretion of inositol into the environment. A tester strain sprayed on top is visible as a red halo around the *S. cerevisiae* colonies (big yellow circles). No halo is visible for the wild type (WT). The G120W mutation shows the biggest red halo, whereas both the G120Y and G120A mutations show only a faint red halo. The experiments were planned and performed by Dr. Harald Hofbauer.

These results highlight the importance of G120 for the functional role of the Opi1 AH. The conformational flexibility of glycine makes its unfavorable to be in an α -helical conformation. As described before, the folded and membrane-inserted AH has a tendency to kink at G120 (see Chapter 2). This observation, in combination with the results above, suggests that G120 acts as a switch to control the functionality of Opi1.

Tethering and folding of G120A at lipid interfaces

Additional simulations were set up to test if point mutations at position 120 help the formation of α -helical secondary structure. First, five repetitions of an unfolded 22-mer Opi1 AH with the G120A point mutation were set up in a box containing 70% PC and 30% PA lipids (no. 8 in Table 3.2). Visual inspection revealed the same membrane recruitment and tethering behavior as observed in simulations of the wild type peptide. The unfolded peptide is recruited to the membrane in dozens of nanoseconds and kept close through electrostatic interactions. Subsequent insertions of L113 and/or I117 tether the peptide at the lipid-water interface. Figure 3.14 shows the structural composition of the 22-mer peptide across five independent simulations with a length of almost $15 \mu\text{s}$ each. The structural ratio is very similar to wild type simulations in the same membrane composition (Figure 3.8D).

The AH forms short stretches of α -helical turns outside the membrane. Visual inspection of the secondary structure formation in these simulations reveals the folding of different parts of the AH. Figure 3.15 shows the time series of the dihedral angles ϕ and ψ for three independent repetitions of the G120A simulations for the residues A120-L124. This region is found in an unfolded conformation in the majority of the simulations. Nevertheless, different parts of the 22-mer peptide explores α -helical conformations for short amounts

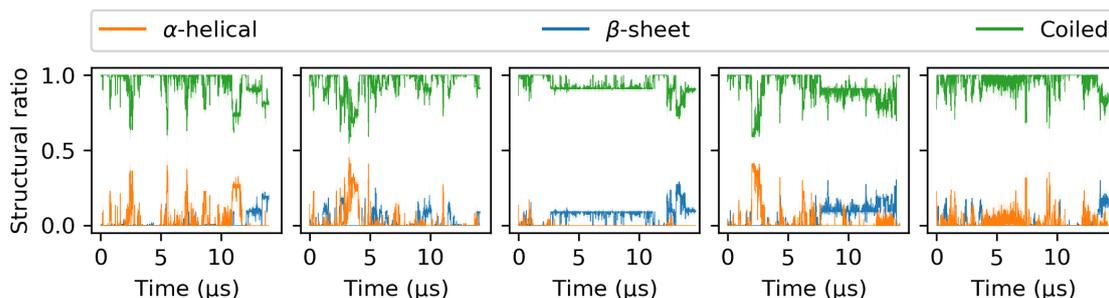


Figure 3.14: Structural composition of 22-mer peptide with the G120A point mutation. Each panel shows an independent simulation, with all five simulations accumulating a total time of 71.5 μs . Partial formation of an α -helical secondary structure can be observed in some simulations. Similarly, a few residues adopt a β -sheet conformation. The secondary structure of each amino acid was evaluated at every timestep. The colored solid lines show α -helix (orange), β -sheet (blue) and coiled (green) secondary structures. The data was window averaged over 1.25 ns and shaded areas indicate the standard deviation.

of time. A short α -helix is formed between residues A120-L124 and subsequently inserted into the membrane at around 11 μs (Figure 3.15A). Indeed, this is the region of interest whose folding and membrane insertion has not been observed before. Unfortunately, the insertion is not permanent and the short α -helix is excluded from the membrane after less than 1 μs . The reason appears to be an internal conflict of the peptides' residues to embed into or interact more closely with the membrane. Although A120 and L124 are both part of the α -helix and, together with L113, are inserted into the membrane, the hydrophobic residue I117 is not (Figure 3.15A). Instead, it faces away from the lipid-water interface and is fully exposed to the solvent. Its propensity to insert into the membrane seems to put a strain on the partially folded peptide and ultimately leads to the unfolding. In a different simulation a short α -helix is formed from residues K121-Y127 after 3 μs outside of the membrane (Figure 3.15B). Here, the 22-mer peptide is tethered to the membrane by L113 and all hydrophobic residues on the short α -helix point towards the membrane, but never insert into it. This α -helix does not encompass A120 and was also observed in the extensive wild type simulations described earlier (Section 3.3.3). In the third simulation, a slightly longer α -helix is formed from A120 to Y127. The dihedral angles ϕ and ψ clearly reveal what is easily missed during a visual inspection: the α -helix is left-handed. This conformation is stable for more than 3 μs , but never inserts into the membrane. Instead it stands upright on the membrane. The hydrophobic residues appear to compete with the charged lysines on the opposite side for a closer contact to the lipid-water interface. Again, this short α -helix does not insert into the membrane over the course of its lifetime. As there is no experimental data to support a left-handed helix, it can be assumed that this might be an artifact of the CHARMM36 force field, as reported by Best *et al.* [189].

Validation of double mutation G120A/K121A simulations

A single point mutation at position 120 increases the α -helical propensity in this region. Now that the experimental observations were confirmed by MD simulations, additional simulations with a second point mutation were set up. As mentioned earlier, the folding of

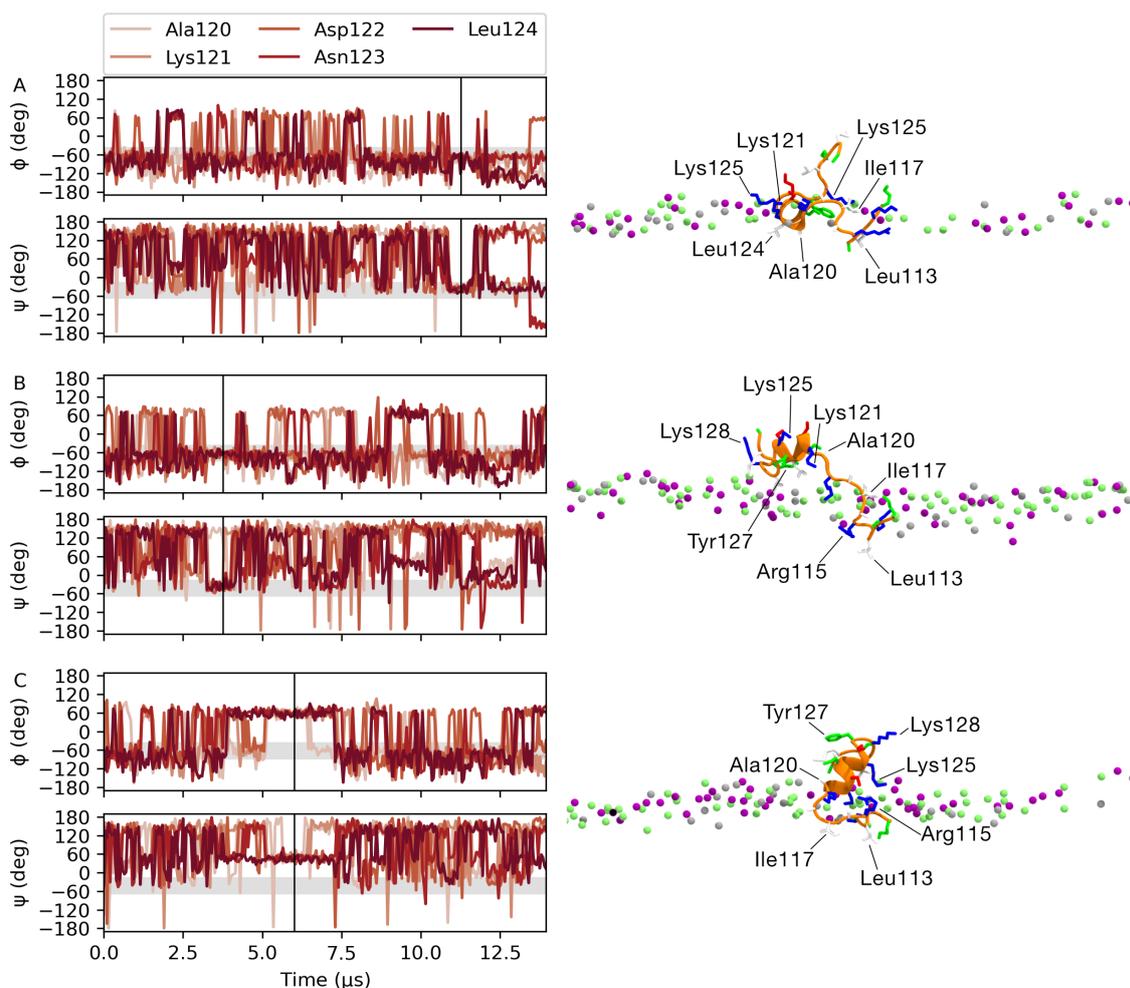


Figure 3.15: Dihedral angles ϕ and ψ as a function of time for three independent simulations of the unfolded 22-mer peptide with a G120A point mutation. (A) A short α -helix is formed after 11 μ s from A120–L124. (B) C-terminal α -helix from K121–Y127 formed outside of the membrane after 3.7 μ s. (C) Left-handed α -helix forms for more than 3 μ s outside of the membrane. Vertical lines indicate the origin of the render on the right side. The panels show repetitions 1, 2 and 5 from system no. 8 in Table 3.4. The data was sampled every 50 ns.

the AH must also overcome the formation of an α -helix at residue K121. In the partially folded and inserted 22-mer peptide (Figure 3.10) the folding stops just before the “KGK motif”. To include K121 in the α -helix it must pass through the lipid-water interface during helix-formation, or the helical turn must form outside of the membrane and slide into it to insert. To test if K121 does indeed slowdown the folding of the AH, two independent simulations with a 22-mer G120A/K121A double mutant peptide were set up (no. 9 in Table 3.4). The structural ratio of secondary structures is shown in Figure 3.16. Similar to all previous simulations, the 22-mer peptide forms α -helical structures for parts of its polypeptide chain. The lifetimes of these structures ranges from hundreds of nanoseconds to multiple microseconds. β -sheet secondary structures were also formed by the G120A/K121A double mutant.

Visual inspection of the trajectory reveals a folded α -helix around the two mutated residues. Figure 3.17 shows the dihedral angles ϕ and ψ as a function of time over the

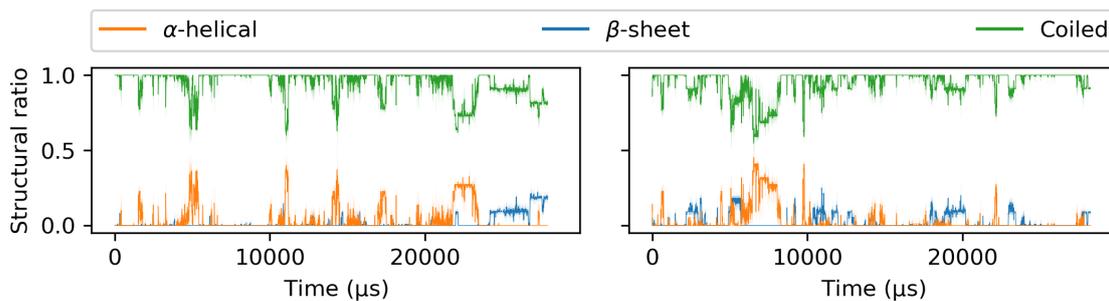


Figure 3.16: Structural composition of 22-mer peptide with the G120A point mutation. The two panels show the structural composition as a function of time, covering a total simulation time of 31.5 μs . Partial formation of an α -helical secondary structure can be observed in some simulations. Similarly, a few residues adopt a β -sheet conformation. The secondary structure of each amino acid was evaluated at every timestep. The colored solid lines show α -helix (orange), β -sheet (blue) and coiled (green) secondary structures. The data was window averaged over 1.25 ns and shaded areas indicate the standard deviation.

whole trajectory for a single repetition. The 22-mer peptide explores various unfolded conformations, before folding a short α -helical stretch around residues K119-A121 after 5 μs for some hundred nanoseconds (Figure 3.17). Intriguingly, the peptide manages to fold a longer α -helical stretch in the same region one microsecond later. At first the α -helix stretches S114-A120, with L113 close to the lipid-water interface and additional tethering by Y127, L129 and M131 (Figure 3.17A).

The short α -helix is propped up against the membrane. The electrostatic interactions between the charged membrane surface and the peptides' lysine residues force the α -helix to face the membrane with its hydrophilic side and expose most hydrophobic residues to the solvent (Figure 3.17B). After several hundreds of nanoseconds L113 manages to reinsert into the membrane and properly tether the peptide (Figure 3.17C). Unfortunately the proper insertion of this folded α -helix is not observed in the trajectory. Again, it appears as if the charged residues have a tug-of-war against the hydrophobic residues and hinder insertion. Ultimately, the α -helix fully unfolds after a total lifetime of 6 μs .

The single and double mutations at positions 120 and 121 of the 22-mer peptide reveal its propensity to fold into an α -helix. These results are in agreement with the experiments that were described at the beginning of this section. They point to the hypothesis that G120A acts as a control element in the AH. It slows down the folding and insertion of the Opi1 AH, to ensure that there is indeed an abundance of PA lipids on the ER membrane.

3.3.5 Folding of the amphipathic helix using enhanced sampling

Previous simulations of the 22-mer have shown that the AH can stay inserted in the lipid-water interface for tens of microseconds (Chapter 2). Complete folding of the AH has not been observed in the extensive simulations described in the previous sections. The folding of proteins can be thought of as a rugged energy surface. Local minima might be separated by high energy barriers that will rarely be crossed with thermal fluctuations. With limited sampling of unbiased MD simulations, rare events such as folding of proteins are rarely observed. Different enhanced sampling techniques were developed to overcome

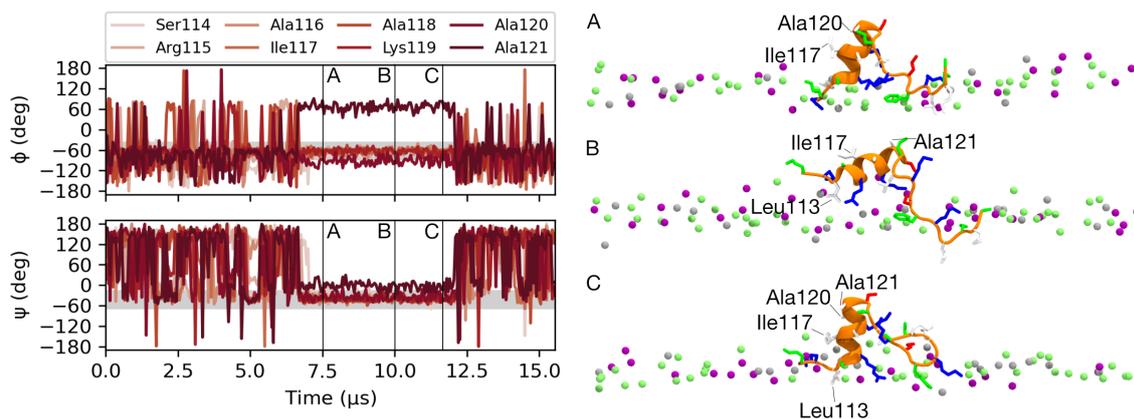


Figure 3.17: Dihedral angles ϕ and ψ as a function of time for a single simulation of the unfolded 22-mer peptide G120A/K121A double mutant. The vertical lines (A-C) in the time series indicate the origins of the renders on the right side. (A) An α -helix spanning residues S114–A120 forms outside the membrane after 7.5 μ s. (B) L113 tethers the helix to the membrane in an inverted orientation after 10 μ s. (C) The α -helix stands upright on the membrane with L113 tethered into the lipid-water interface. The panels show a single repetition of system no. 9 in Table 3.4. The data was sampled every 50 ns.

these energy barriers and thus speedup the exploration of the conformational space. In the following, enhanced sampling techniques were tested for their capability to fold the remaining residues of the partially folded 22-mer wild type peptide from a previous section (Section 3.3.3). The same partially folded starting configuration after 7 μ s was chosen for all biased simulations.

Biasing the secondary structure of Opi1 AH

ABMD adds a ratchet-and-pawl restraint acting along a specified CV [196,197]. This means that the system is allowed to evolve into the direction that increases the CV naturally and no external force is applied. If the system tries to evolve such that the CV decreases, a harmonic potential is used to keep the system in its current state. Using the analogy of a rugged energy surface being a mountain: in ABMD the system is chaperoned to climb the top, but kept from falling down hill by external influences. The choice of a CV is very important and Pietrucci and Laio have introduced a CV for the efficient exploration of α -helical and β -sheet secondary structures [201]. This CV will be referred to as *alpharmsd* in the following. The RMSD between a six residue stretch of the simulated structure and an ideal α -helix is calculated. The RMSD is transformed to a value between 0 and 1 with a switching function and summed over all possible six residue blocks. A high CV value should indicate the successful folding of the AH. The reference *alpharmsd* CV for a 22-mer α -helical peptide was determined to be 14.5. This is the target value which the system should have after successfully biasing.

Five biased simulations were set up with the *alpharmsd* CV for different values of κ : 0.5, 5, 50, 100 and 500 kJ mol^{-1} . All residues of the 22-mer peptide (K112–M131) were considered with the harmonic restraint during the ABMD simulations. Figure 3.18 shows the *alpharmsd* CV as a function of time for all five values of κ in panel A, whereas panel B shows the instantaneous bias potential applied at a given time. The low value

of $\kappa = 0.5 \text{ kJ mol}^{-1}$ is sufficient to fold an α -helix from K122–K128 after 40 ns, but not strong enough to have the “K GK motif” folded as well. The final structure resembles that of two separate helices, with one embedded into the membrane and the second one perpendicular to it. Increasing the κ value to 0.5 and 5 kJ mol^{-1} results in a left-handed helix spanning A121–M131 and G120–M131, respectively. Both of these left-handed α -helices are again found outside of the membrane. Setting $\kappa = 100 \text{ kJ mol}^{-1}$ collapses the unfolded region of the 22-mer peptide into an α -helical conformation from residues K119–K125. At the same time, a short single turn left-handed helix is formed starting at E126. Only $\kappa = 100 \text{ kJ mol}^{-1}$ is able to fold a proper α -helix, which inserts into the membrane after formation. This simulation was run for a total of 720 ns with the bias constantly acting on the CV. This inserted and fully folded AH was then run for another 1.5 μs without any bias in a conventional MD simulation. The AH remained fully folded and membrane-embedded throughout this unbiased MD simulation.

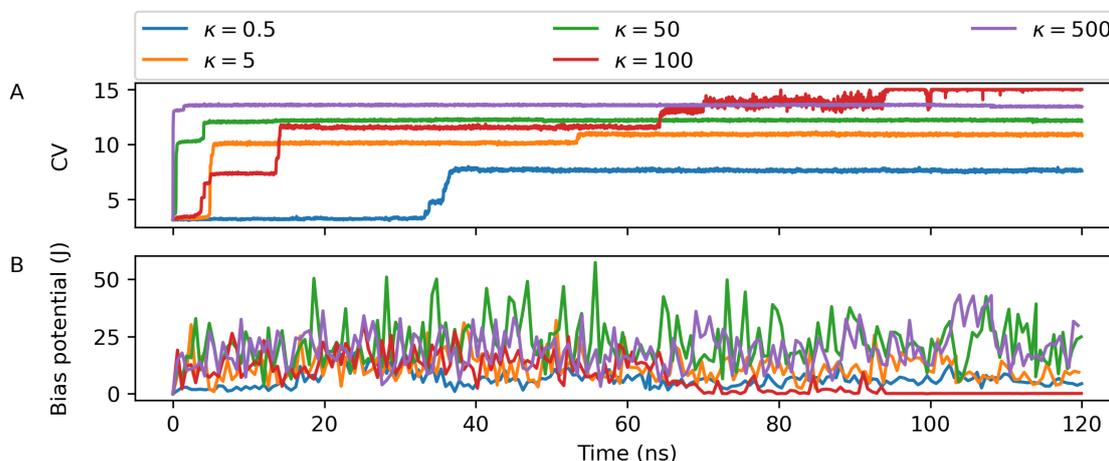


Figure 3.18: Observables from ABMD simulations for different values of κ . The (A) collective variable and (B) bias potential are plotted as a function of time for different values of κ . Only the simulation with $\kappa = 100 \text{ kJ mol}^{-1}$ is able to completely fold and embed the AH. The others result in left-handed α -helices.

Dihedral angle biases of a single turn

A second biasing approach was used to try and fold the AH. Although the *alpharmsd* CV was successful in folding the AH, it also produced unexpected left-handed α -helices. One way to overcome this problem is to specifically bias on the peptides’ backbone dihedral angles. The ϕ values, and to some degree also the ψ values, are separated far enough between left-handed and right-handed α -helices. The folding of α -helices is favored to proceed from N- to C-terminus [202]. In some of the previous simulations folding was sometimes observed to start in the middle of the sequence. Even after insertion of this partially folded AH into the membrane (Figure 3.15A), the folding would have needed to continue both toward the N- and C-terminus for completion. To test if a bias on the dihedral angles is sufficient to fold, an ABMD simulation was set up and a harmonic potential was applied on the dihedral angles of A118, K119, G120 and K121 one after the other with $\kappa = 500 \text{ kJ mol}^{-1}$. The bias

for the ϕ and ψ dihedral angles of a given residue was turned on followed by a waiting time of 1 ns before the next bias. Figure 3.19 shows the time series of the ϕ and ψ dihedral angles, as well as four snapshots of the folding. The bias for A118 is immediately turned on, bringing the ϕ dihedral into an α -helical region. Subsequently, the biases for K119 and G120 are turned on. The time series shows that the ϕ dihedral of K119 first undergoes a reorientation and falls into the α -helical region, correlated with the ψ dihedral of G120. The short α -helical turn in the “KKG motif” forms outside of the membrane (Figure 3.19A) and must rearrange the position of its lysines (Figure 3.19B), before it is able to start inserting (Figure 3.19C) into the membrane (Figure 3.19D). The C-terminal N122 residue transiently adopts an α -helical conformation at 25 ns, but does not keep it without any biases. The insertion or folding of this partially folded AH does not continue. K121 is seen to prop up the α -helix against the membrane, causing the AH to tilt (Figure 3.19D). The simulation was run for more than one microsecond, after which the bias on the four residues was turned off. The “KKG motif” immediately unfolded without any external biases. This points to the observation that the α -helix is propped up against the membrane, has no bulky hydrophobic anchor and is not properly inserted at its folded C-terminus. As further folding was not observed in the microsecond long ABMD simulation and folding of a single loop was not sufficient to keep it in the α -helical conformation, another simulation with a more sophisticated biasing scheme was set up.

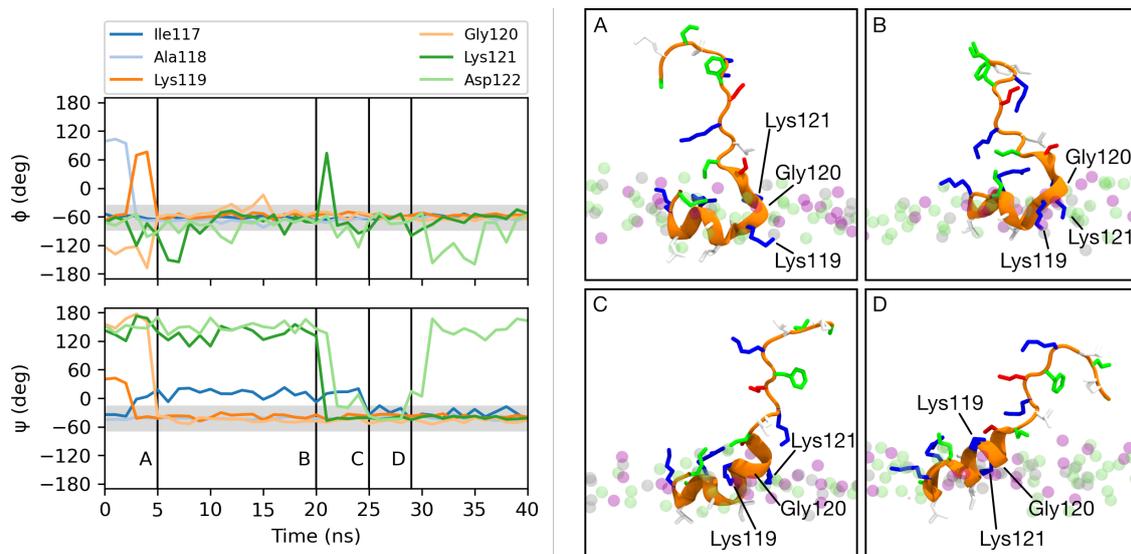


Figure 3.19: Snapshots from ABMD folding simulation of AKGK in 22-mer peptide. The left panels show the dihedral angles ϕ and ψ as a function of time for the first 40 ns of the simulation. Vertical lines in the time series correspond to the snapshots in the panels on the right (A–D). Key residues are annotated and lipid headgroups are shown as transparent colored beads.

Blockwise biases on dihedrals of separate turns

A blockwise biasing scheme was used to fold the remaining AH starting at the partially folded 22-mer peptide. In light of the observation that the folding of the previously described partially folded 22-mer peptide proceeded from N- to C-terminus, a biasing scheme was set up to fold each helical turn separately. In brief, the ϕ and ψ dihedral angles of the amino

acid residues of every turn were biased in short succession using steered MD [198]. This method acts on each dihedral with a specified force and pushes it into the specified direction. After the bias on each turn was in place, a waiting time of several dozen nanoseconds held the turn folded. The dihedral biases on the next turn were turned on after the previous waiting time. A total of three blocks were defined: (i) A118–K121, (ii) D122–K125 and (iii) E126–K128. Each block contains one hydrophobic amino acids that can anchor the AH on the membrane. In between blocks a waiting time of 80 ns was observed, as well as an additional waiting time of 1 ns between each residue inside a block. Figure 3.20 shows the dihedral angles ϕ and ψ as a function of time.

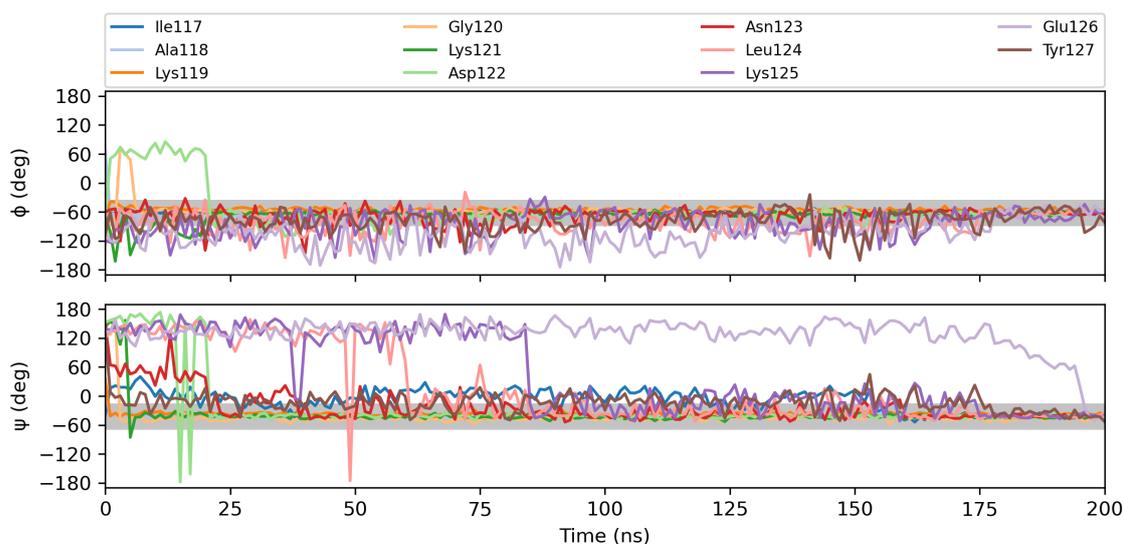


Figure 3.20: Dihedral angles from 22-mer steered MD simulation show folding of the AH. The dihedral angles ϕ and ψ are plotted as a function of time for I117–Y127. The dihedral biases were switched on in blocks of three to four residues at a time. ψ dihedrals of A118–G120 adopt the expected values immediately, whereas I117 takes more than 150 ns to fold. The gray area indicates the common ϕ (-90 deg to -35 deg) and ψ (-70 deg to -15 deg) dihedrals found in α -helical structures.

The ψ dihedral angles of A118–K121 immediately converge to the α -helical region after their respective bias is turned on. D122 spontaneously adopts the helical conformation after 25 ns, followed by N123 and L124 that also adopt the conformation without any biases after 60 ns. The bias of the second block was turned on after 80 ns and immediately pushed the ψ dihedrals of K125 into an α -helical conformation. Finally, Y127 and E126 adopt the α -helical conformation after 175 ns. Although the dihedral angles of I117 were not biased, its time series are also plotted in Figure 3.20. The ψ dihedral indicates a kink in the α -helical structure and only relaxes into the fully folded α -helix after 150 ns. Snapshots of the folding are shown in Figure 3.21. The panels A–C visualize how the C-terminal helix is formed almost turn-by-turn. After I117 adopts a fully α -helical conformation, the folded C-terminal helix inserts into the membrane in less than 10 ns (Figure 3.21C and D). The remaining residues Y127 and K128 folded after the insertion.

Biasing on the separate dihedral angles ϕ and ψ in distinctive helical blocks results in a properly folded α -helix. It has to be noted that the helix did not form as planned. It was

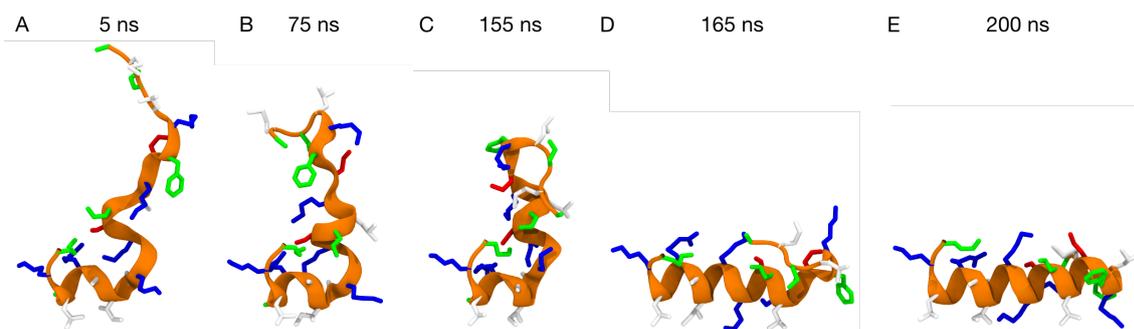


Figure 3.21: Snapshots from steered MD folding simulation of 22-mer Opi1 AH. Each panel shows a snapshot at a different point of the trajectory. (A) The turn around the “KKG motif” is formed. (B) The second turn (D122–L124) forms and (C) continues to E126. (D) I117 adopts an α -helical conformation and tilts the helix onto the membrane. (E) The C-terminus of the AH folds into an α -helix. Lipids and solvent are not shown for clarity.

envisioned that each turn would form separately and then insert into the membrane to elongate any previously folded helix. Only after insertion the next turn was supposed to fold, insert and thus again elongate the growing helix. The results described above show that the α -helix is mostly formed outside of the membrane and insert into the lipid-water interface in a single quick motion. It remains unclear, if this is the actual process of the Opi1 AH membrane insertion or if another pathway can also lead to the fully folded state.

3.4 Discussion

The conformational dynamics of the Opi1 AH was studied with atomistic MD simulations in addition to *in vivo* and *in vitro* experiments. Based on the abundance of PA, Opi1 maintains the lipid homeostasis of GPLs in *S. cerevisiae*. It uses its AH to recognize and interact with PA. Simulations at high temperature have shown that the Opi1 AH is not folded in solution, consistent with experimental observations [157]. Similarly, other AHs have been described to be unstructured in solution. For example, both the AH of Atg8 and mellitin fold only upon membrane binding [203–207].

When Opi1 is not bound to the membrane its AH must decide if it is worth tethering, folding and inserting into the lipid bilayer. Simulations with different membrane compositions have revealed a difference in binding preference of the AH. With sufficient membrane charges the AH is recruited to membranes and kept close long enough for its hydrophobic residues L113 and I117 to tether the unfolded peptide at the lipid-water interface. Without any membrane charge, thus in the presence of only PC headgroups, the unfolded peptide is transiently recruited to the membrane, but fails to properly tether itself. The addition of PE headgroups allows the AH to tether to the membrane more often. PE lipids are considered to be conical lipids that prefer a negative curved membrane, as compared to cylindrical PC lipids that prefer positive curvature. Previous studies have shown that conical lipids lead to lipid-packing defects in lipid bilayers [16, 22, 208–210]. Lipid-packing defects are areas with a low lipid density exposing parts of the hydrophobic membrane core. It is possible that the introduction of PE lipids exposes more packing-defects on the surface and thus leads to a more stable tethering of the Opi1 AH. The exchange of PE for PA lipids reveals that

not only the lipid-packing defects, but also the general membrane charge can be important for tethering. PA headgroups are smaller than PE and hold a negative charge of 1- in the simulations presented in this thesis. Experimental studies have shown that the charge of PA increases from 1- to 2- when interacting with lysine or arginine residues [211]. The Opi1 AH is lined with both lysine and arginine residues, providing sufficient interaction partners for the PA headgroups.

After the AH is tethered to the membrane, it must either fold and insert or dissociate. The simulations in this thesis have shown the propensity of the Opi1 AH for partial folding at lipid-water interfaces. This observation highlights the urge of the peptide to form an α -helical conformation and hide its hydrophobic residues inside the hydrophobic core of a membrane. All folding events did not lead to a stably inserted and fully folded α -helix. In one case a short two turn α -helical conformation was stabilized by membrane insertion for $\sim 10 \mu\text{s}$, but was excluded from the membrane and ultimately unfolded. No full folding of the AH was observed in the extensive unbiased simulations reported in the previous sections. The folding pathways of the designed WALP peptide, as well as if the antimicrobial peptide mellitin have been studied extensively with biased and unbiased MD simulations [212–215]. Spontaneous insertion of these peptides has been observed, whereas proper folding required the use of biased MD simulations to overcome free energy barriers. An exception is seen with mellitin, which is able to fold on the lipid-water interface without any external biases [213]. Its partial and full folding were observed at high-temperature simulations in less than $2 \mu\text{s}$. Specific differences, as the primary sequence of the peptide, as well as the membrane environment can have an impact on the folding kinetics. Charged residues are found mostly at the termini of the mellitin peptide, whereas they are found throughout the AH of Opi1. The presence of these charged residues can have a severe effect on the folding on charged lipid-water interfaces.

In the MD simulations folding stopped right before the “KGK motif” suggesting a biological function of some of these three residues. The sequence similarity analysis revealed that most of the AH is conserved in *Saccharomyces*, including the “KGK motif”. To test if these theoretical observations are indeed of relevance, *in vivo* and *in vitro* were performed, confirming the functional role of G120. The single point mutation G120A increases the stability of the AH, and keeps it bound to the ER membrane even after PA lipid depletion. Simulations of single and double point mutations at positions 120 and 121 in the 22-mer peptide revealed that the “KGK motif” indeed bears a functional role. The results highlight the functional duality of G120 in the Opi1 AH. The folding of the AH is slowed by the presence of G120, to allow the protein to decide if embedding is necessary, depending on the PA abundance. The charges of PA, as well as the packing-defects allow for tethering and insert of the AH into the lipid bilayer. This leads to a “fully functional” form of the AH. In this situation, G120 switches its function and aids to destabilize the α -helical structure of the Opi1 AH. Here, the abundance of PA lipids will determine if the AH will stay embedded. When PA is depleted this destabilization of the secondary structure will lead to the dissociation from the membrane and subsequent unfolding in solution. No literature is available on the functional modulation of AHs by glycines.

While initial folding can happen in hundreds of nanoseconds, the complete folding of the AH takes a long time. Only with sufficient time and the presence of enough membrane charge will the unfolded peptide have a chance to form a complete α -helix and subsequently insert into the membrane. The fully folded α -helix was shown to have a specific lipid sensing function when completely inserted into the membrane (see Chapter 2) [157,183]. The present work shows that G120 acts as a control element that either (i) delays the process of folding before the Opi1 AH can start sensing the PA lipid abundance or (ii) destabilizes the α -helical fold of the AH to enable its dissociation from the membrane, after the depletion of PA lipids from the membrane leaflet. Ultimately, the attempts to fold the AH using enhanced sampling methods have revealed additional complications. The full folding of the AH is possible outside of the membrane, while parts of it are already anchored. It remains unclear if the AH folds in a turn-by-turn fashion by inserting each turn after the other into the membrane. As the simulations in this work have shown, turns can fold separately outside of the membrane, but do not insert by themselves. Instead, the turns form a longer α -helix first and then embed the full helix into the membrane. This work has established a possible mechanism of AH folding on lipid-water interfaces using enhanced sampling methods.

Structural dynamics of the HIV-1 glycoprotein

The investigations of the dynamics of different HIV-1 glycoproteins comprise computational work that was performed at the Max Planck Institute of Biophysics, Frankfurt am Main, Germany. All simulations, data analysis and interpretations were carried out by myself under the supervision of Dr. Roberto Covino (Frankfurt Institute for Advanced Studies, Frankfurt) and Prof. Dr. Gerhard Hummer (Department of Theoretical Biophysics, Max Planck Institute of Biophysics, Frankfurt).

4.1 Introduction

1.7 million people have acquired the human immunodeficiency virus (HIV) in 2019, while 940000 have died due to the acquired immune deficiency syndrome (AIDS) [216]. Although these numbers have declined by 23% in the past decade, finding a cure is still an important goal. The structural and functional characterization of HIV proteins can help the development of novel vaccines and treatment options [217–219].

In order to penetrate the cell's defenses, HIV predominantly targets activated CD4⁺ T cells [220–223]. The envelope (Env) glycoprotein sits on viral membranes and facilitates the fusion between the viral and target cell membrane [53, 224, 225]. Two distinct subunits compose the Env protein, a soluble surface domain, gp120, and a membrane-bound domain, gp41, which is tethered to the viral membrane by one TMD [226]. The two protomers are initially synthesized as a precursor polypeptide gp160, and are produced via proteolytic cleavage. gp120 mediates the interaction with the CD4 receptor and its co-receptors CCR5 or CXCR4, after rearrangement to the open state from a previously pre-fusion mature state [226]. These interactions lead to membrane fusion, which is thought to be mediated by gp41. After membrane fusion, Env ends up in the post-fusion state. Env is the only protein on the viral membrane that is recognized by the human immune system. Available vaccines bind to the membrane-proximal extended region (MPER) of gp41, which precedes the TMD on its N-terminus [227–229].

HIV-1 Env structures were solved in a trimeric conformation both pre- and post-fusion [230–235], whereas the oligomeric state of the TMD remains under debate. Two recent publications reported the NMR structure of the TMD in 1,2-dimyristoyl-sn-glycero-3-phosphocholine (DMPC) bicelles, describing different oligomeric states in identical experimental setups [236, 237]. Dev *et al.* described a C₃ symmetric homotrimeric structure, that is characterized by a coiled-coil arrangement at the N-terminus and C-terminal polar interactions [236]. Instead, Chilivery *et al.* presented a continuously folded 32-residue

monomeric α -helix, validating their results via extensive NMR and DEER EPR measurements, as well as analytical ultracentrifugation [237]. Two more publications add to the puzzle: one identifies a gp41 conformation with trimeric TMD, but monomeric MPER interacting with the membrane surface [238], whereas the other reports a C_3 symmetric trimeric arrangement of MPER-TMD, where MPER forms two distinct α -helices separated by a kink [239].

The TMD of gp41 has two specific characteristics: (i) its sequence is too long to be accommodated in the hydrophobic core of its host membrane and (ii) it displays a highly conserved arginine, R696, which experiments locate at the center of the membrane [236]. Cationic residues are conserved at this position and mutational screenings of R696 reveal a loss of viral membrane fusion [236, 240–243]. MPER, preceding the TMD on its N-terminus, is well conserved and also important for membrane fusion [244]. It interacts with cholesterol-rich membranes through its cholesterol recognition/interaction amino acid consensus (CRAC) and could act as an AH [245–248]. The structure and dynamics of MPER are of great interest, because it is the only region in gp41 that is recognized by two bnAbs, 4E10 and 10E8, which bind at the junction between MPER and TMD [246]. These antibodies show a neutralization of up to 98% and have helped to identify one Achilles' heel in the HIV-1 Env glycoprotein.

Aim. Here, extensive all atom MD simulations are used to study the dynamics of the monomeric and trimeric conformations of gp41. The aim was to identify the differences in their arrangements in the membrane. The results highlight specific MPER conformations that recover known broadly neutralizing HIV-1 antibody (bnAb) epitopes. The accessibility of these epitopes is studied in the absence and presence of the N-terminal Env glycoprotein.

4.2 Methods

Throughout this chapter the sequences from HIV-1 Clade D were used. All residue numbers are referred to by the HXB2 numbering [249]. All simulations were performed with the CHARMM36m force field [179] using the TIP3P water model [250]. All simulations were run using GROMACS 2016 [165].

4.2.1 Membrane composition

Simulations were set up with symmetric lipid bilayers representing viral-like compositions (Table 4.1) and membranes completely composed of DMPC. In viral-like membranes 1-palmitoyl-2-linoleoyl (PLxx) lipids were included. These contain one acyl chain with two unsaturated bonds, and replace plasmalogen lipids (6 unsaturated bonds) that are enriched in viral membranes [251]. Parameters for plasmalogen lipids were not available at the time of writing of this chapter. A variety of different lipid species and sterols was used for the viral-like compositions: cholesterol, 1-palmitoyl-2-linoleoyl-sn-glycero-3-phosphocholine (PLPC), 1-palmitoyl-2-linoleoyl-sn-glycero-3-phosphoethanolamine (PLPE), POPE, POPE, sphingomyelin (SSM) and DOPC.

Table 4.1: Lipid composition used for viral-like membrane bilayers in atomistic simulations. The columns represent the total number of lipids in both leaflets.

Lipid	Number of lipids	Lipid fraction (%)
Cholesterol	96	32
PLPC	52	17
PLPE	48	15
POPE	48	15
POPC	40	13
SSM	18	6
DOPC	6	2

4.2.2 Preparation of membrane based structures

UCSF CHIMERA [162] was used to model the 50-residue segment of the gp41 TMD (672-WFDITNWLWYIRIFIIIIVGSLIGLRIVFAVLSLVNRVRQGYSPFSFQTLL-721) into an α -helix and align it along the z-axis for all monomeric TMD simulations. Similarly, a 61-residue segment (656-NEKDLELDKWASLWNWFDITNWLWYIRIFIIIIVGSLIGLRIVFAVLSLVNRVRQGYSPFS-716) was used to model the initial α -helix for monomeric MPER-TMD simulations. The R696 rotamers were generated using the Dunbrack 2010 library [252], choosing the χ angles such that the arginines point in a parallel, orthogonal or anti-parallel direction with respect to the membrane normal (Table 4.2). The two published trimeric structures (PDB IDs: 5JYN and 6E8W) were used for simulations of the trimers [236, 239]. In all simulations N- and C-terminus were acetylated and amidated, respectively.

Table 4.2: Side chain dihedral angles used for the three R696 rotamers with their accompanying names.

Name	χ_1	χ_2	χ_3	χ_4
up	-70.7	177.5	-180.0	172.6
middle	-179.2	175.2	178.1	-178.6
down	-179.4	65.3	178.1	173.5

4.2.3 Preparation of soluble HIV-1 gp140 structure

The crystal structure of HIV-1 gp140 (PDB ID: 5CEZ [253]) was used to model the trimeric structure. The crystal structure 5CEZ contains a single unit that was used to construct the C_3 symmetric trimer. First, all partially resolved glycans were removed from the structure in the single unit. A total of 26 residues were not resolved in the crystal structure across three missing regions, two in the V1 loop and one in the V4 loop. These residues were modeled using MODELLER [254]. The missing residues are listed in Table 4.3.

Table 4.3: Missing residues in HIV-1 gp140 model that were constructed with MODELLER.

Region	V1	V1	V4
Residues	D140–G152	N185–S187	N398–S410

After modeling the missing loops, the trimeric protein was created with UCSF CHIMERA

by mirroring the single unit to establish C_3 symmetry. The protein was embedded in a 14.9 nm^3 simulation box, solvated with TIP3P water and sodium and chloride ions using CHARMM-GUI [172, 178]. The system was energy minimized using steepest descent and subsequently equilibrated in a single step for 25 ps with an integration time of 1 fs. The Nosé-Hoover thermostat was used to establish a temperature of 310 K with a characteristic time of 1 ps [152, 255, 256]. For the 665 ns production simulation the Velocity Rescale thermostat [170] and Parrinello-Rahman barostat [171] were used. The thermostat kept a constant temperature of 310 K with a characteristic time of 1 ps, whereas the barostat established a pressure of 1 bar with a characteristic time of 5 ps.

4.2.4 Determination of HIV-1 gp140 stability

The stability of the HIV-1 gp140 protein structure (simulation #9) was quantified by calculating the RMSD and root-mean-square fluctuation (RMSF). Before calculations of either of these two quantities, the instantaneous structure was superimposed onto the starting structure using a rigid-body alignment.

Root-mean-square deviation

The RMSD measures the deviation of atomic coordinates of an instantaneous structure to a reference structure and averages this deviation over all particles N . Therefore, the RMSD is a function of time t :

$$\text{RMSD}(t) = \sqrt{\frac{1}{N} \sum_{i=1}^N |r_i(t) - r_i^{\text{ref}}|^2} \quad (4.1)$$

where r_i is the position of the i th atom in euclidean space and r_i^{ref} is the position of the i th atom in the reference structure.

Root-mean-square fluctuation

The RMSF measures the deviation of the instantaneous position of the j th particle and its reference position as an average over the trajectory of time T :

$$\text{RMSD}_j = \sqrt{\frac{1}{T} \sum_i^T |r_j(i) - r_j^{\text{ref}}|^2} \quad (4.2)$$

The RMSF of every side chain in the HIV-1 gp140 structure was calculated by considering only the C_α backbone atoms. To calculate the RMSD of the structure, only C_α backbone atoms with a RMSF $\leq 3 \text{ \AA}$ were considered.

4.2.5 Calculation of the hydrophobicity profile

Every residue in the sequence of the gp41 TMD was assigned a hydrophobicity value according to the scale defined by Kyte and Doolittle [257]. For every residue i the hydrophobicity

score was computed by considering $i - (n - 1)/2$ residues on each side, with the window size $n = 19$. The final score for residue i is the sum of all hydrophobicity values of the considered residues. In addition, the hydrophobicity of 18-residue windows was assessed with HeliQuest [258].

4.2.6 Setup of membrane systems

CHARMM-GUI was used to set up membrane systems containing either the monomeric or trimeric structure of gp41 (MPER-)TMD [172, 175–178]. All structures were centered at the origin, with the N-terminus pointing in the positive z-direction. Simulations were set up with a symmetric viral-like lipid bilayer (Table 4.1) with 154 lipids per leaflet, as well as a symmetric DMPC model lipid bilayer with ~ 150 lipids per leaflet. Ions were added for a total concentration of 0.15 M, with KCl for the two trimer simulations in viral-like bilayers (systems #6) and NaCl for all other simulations (systems 1–5 and #7–8). The systems were first energy minimized and subsequently equilibrated. Each system was energy minimized using steepest descent and subsequent equilibration was done in several steps. In the first step, the system was equilibrated in a canonical (NVT) ensemble with an integration time step of 1 fs for 25 ps, maintaining a temperature of 310 K. During this phase, position restraints of $1000 \text{ kJ mol}^{-1} \text{ nm}^{-2}$ were applied to lipid heavy atoms. Restraints on backbone heavy atoms of the protein were lowered from $4000 \text{ kJ mol}^{-1} \text{ nm}^{-2}$ to $2000 \text{ kJ mol}^{-1} \text{ nm}^{-2}$. The Berendsen barostat was used to keep a constant pressure of 1 bar in the isothermal-isobaric (NPT) equilibration for the first 25 ps, with a time step of 1 fs [151]. The subsequent equilibration steps were performed for 300 ps with a time step of 2 fs. Here, lipid position restraints were decreased from $400 \text{ kJ mol}^{-1} \text{ nm}^{-2}$ to $0 \text{ kJ mol}^{-1} \text{ nm}^{-2}$ and protein position restraints were decreased from $1000 \text{ kJ mol}^{-1} \text{ nm}^{-2}$ to $200 \text{ kJ mol}^{-1} \text{ nm}^{-2}$. Trimeric simulations in DMPC bilayers underwent an additional equilibration of 50 ns, where positions restraints were applied on all protein heavy atoms or backbone atoms. Restraints were decreased in subsequent steps, as described above. During the production simulations no restraints were applied on the system. The Velocity Rescale thermostat [170] and Parrinello-Rahman barostat [171], with semiisotropic pressure coupling, were used to keep a constant temperature of 310 K and pressure of 1 bar, applying them on the protein, membrane, and solvent with characteristic times of 1 ps to 5 ps, respectively. Throughout the simulation non-bonded interactions were cutoff at 1.2 nm. All HIV-1 glycoprotein simulations reported in this chapter are listed in Table 4.4.

4.2.7 Calculation of tilt angles

The angle of the TMD to the membrane normal was calculated to quantify the TMDs tilt. The COM of two reference atom groups were defined as two points A and B (Table 4.5). The vector \vec{v} spanning through BA forms a tilt angle with the membrane normal. The time series was window averaged with a window size of 25 ns. Similarly, the tilt angle of the R696 side chain was calculated by defining the points A and B as the side chains' C_α and C_ζ atoms, respectively.

Table 4.4: Simulated systems with HIV-1 glycoprotein structures reported in this chapter. The simulation time is listed for each repetition of a single membrane composition or construct.

#	Description	PDB ID(s)	Membrane	Time (μ s)	Box (nm^3)
1	Monomer TMD ¹	–	Virus-like	6.4	10.0×10.0×10.0
2	Monomer TMD ²	–	Virus-like	6.4	10.0×10.0×10.0
3	Monomer TMD ³	–	Virus-like	6.4	10.0×10.0×10.0
4	Monomer MPER-TMD ¹	–	Virus-like	4.0	10.0×10.0×10.0
5	Monomer MPER-TMD ³	–	Virus-like	4.0	10.0×10.0×10.0
6	Trimer TMD	5JYN	Virus-like	2×10.0	13.0×13.0×9.3
7	Trimer MPER-TMD ⁴	6E8W	DMPC	3×3.0	9.8×9.8×10
8	Trimer MPER-TMD ⁵	6E8W	DMPC	3×3.0	9.8×9.8×10
9	HIV-1 gp140	5CEZ	–	0.65	14.9×14.9×14.9
10	HIV-1 gp140 connected to gp41 MPER-TMD	5CEZ 6E8W	DMPC	4.0	15.0×15.0×15.0
11	HIV-1 gp140 connected to gp41 MPER-TMD with bound 4E10 bnAb	5CEZ 6E8W 4XAW	DMPC	0.69	22.9×22.9×22.6

¹ Initial orientation of R696 side chain in *up* conformation.

² Initial orientation of R696 side chain in *middle* conformation.

³ Initial orientation of R696 side chain in *down* conformation.

⁴ Protein heavy atoms restrained during additional 50 ns of equilibration.

⁵ Protein backbone heavy atoms restrained during additional 50 ns of equilibration.

4.2.8 Distances of N- and C-termini

The N- and C-terminal distances were calculated as a function of time, to quantify the instability of the trimeric structure in systems #6, #7 and #8. The C $_{\alpha}$ atoms of L679 and S703 were defined as the N-terminal and C-terminal reference points, respectively. The pairwise distance between all three protomers was calculated, window averaging the resulting time series with a window size of 10 ns.

4.2.9 Density calculations

The mass densities of different atom groups were calculated over the whole trajectory using `gmx density`. The trajectory was centered around the membrane before each calculation.

Table 4.5: Atom groups used to calculate the tilt angle of helical segments in gp41 TMD and gp41 MPER-TMD simulations. The center of mass of all backbone atoms from group *A* and *B* were used as reference points for the respective group. The system numbers are the same as in Table 4.4.

#	Type	<i>A</i>	<i>B</i>
1–2, 4	N-terminal TMD	I682-I686	L692-R696
1–2, 4	C-terminal TMD	V698-L702	R707-G711
3, 5	Full TMD	I682-G690	S703-G711
6	Full TMD	Y681-V689	L702-Q710
7–8	Full TMD	I682-G690	S703-G711

All atoms were binned into 200 equally sized spatial bins according to their z coordinate, averaging all positions over the trajectory.

4.2.10 Number of hydrogen bonds

The `HydrogenBondAnalysis` module from MDAnalysis [180,181] was used to calculate the number of hydrogen bonds between all amino acids and water molecules. Hydrogen bonds were defined with two geometric criteria: the euclidean distance d between the acceptor and hydrogen atom is $d = 1.3\text{\AA}$, and the angle ω between donor-hydrogen-acceptor is $\omega \geq 130^\circ$.

4.2.11 Alignment of antibody crystal structures

Rigid-body RMSD alignment [259] was used to superimpose the antibody-bound and co-crystallized gp41 epitope structures onto the corresponding regions of gp41 MPER-TMD in the atomistic MD simulations (#7–8). The bnAbs were superimposed onto every frame and every protomer in the 3 μ s trajectories of all six simulations. The bnAb crystal structures used in this study are listed in Table 4.6.

Table 4.6: Epitopes of bnAbs superimposed onto gp41 MPER-TMD structures.

bnAb	PDB ID	Crystal epitope	Simulation epitope	Reference
4E10	4XAW	N671–K686	N671–I686	Irimia <i>et al.</i> , 2016 [260]
10E8	5T6L	N119–D131	N671–R683	Irimia <i>et al.</i> , 2017 [261]
10E8	4G6F	E659–R685	L660–F685	Huang <i>et al.</i> , 2012 [262]

4.2.12 Preparation of antibody structures

The 4E10 bnAb crystal structure (PDB ID: 4XAW [260]) was used for atomistic simulations. All non-protein atoms were removed from the structure, including the co-crystallized glycerol backbones. Not all residues were resolved in this crystal structure and required modeling. While the antigen-binding fragment (Fab) of antibodies has a conserved fold, the paratope and its CDRs vary in length. This allows the CDRs to adapt and recognize the various structural features of their epitopes. To make these variable regions comparable, Kabat *et al.* developed the Kabat numbering scheme [263]. This scheme allows to align the sequences of antibodies using the constant regions by reusing the residue numbers in the CDRs. For example, the 4E10 bnAb is characterized by multiple residues with the residue number 100: TRP100, GLY100A, TRP100B, etc. In the 4XAW structure various residues were missing that are found in the CDRs. The following heavy chain residues required modeling: E1, G99, W100, G100A, W100B, L100C, S129, S130, K131, S132, T133, S134. The residues W100, G100A, W100B and L100C are part of the loop that was shown to interact with the membrane [260]. The co-crystallized epitope was kept in the structure to allow superimposition onto the corresponding MPER region, but was removed in the simulation with the bnAb docked to its epitope underneath the soluble part of the Env glycoprotein (system #11). The bnAb structure was not energy minimized or equilibrated

separately before being used in the other systems.

4.2.13 Preparation of combined HIV-1 structures

To construct a model containing HIV-1 gp120 and gp41, the latter without its CTD, multiple structures were combined. The C-terminal pairwise distances between the C_α atoms of N675 in the gp140 model (simulation #9) were calculated and compared with the N-terminal pairwise distances of I675 C_α atoms in the gp41 MPER-TMD simulations (systems #7–8). The MPER-TMD simulation with the best overlapping distances to the HIV-1 gp140 model was chosen for the next modeling step.

The membrane patch required extension in the x and y directions. The membrane edges of systems #7–8 measured 9.8 nm, whereas the HIV-1 gp140 simulation (#9) was set up with an edge length of 14.9 nm box to fit the bigger soluble part of the Env glycoprotein. The membrane of system #7B was extended by considering its immediate periodic boundary images in x and y directions, resulting in an edge length of 20 nm. The patch was finally cropped to an edge length of 14.8 nm and water and ion atoms were added. To re-establish PBCs, a short 100 step energy minimization using “soft-core” potentials was used [264,265]. This procedure is able to minimize the energy in a system with overlapping atoms by removing the singularity in the Coulomb and Lennard-Jones potentials at very close interatomic distances. Afterwards a normal steepest descent energy minimization was performed, followed by three additional equilibration steps. First, the system was heated to 310 K over 50 ps using the Berendsen thermostat with a characteristic time of 1 ps. Next, the Velocity-Rescale thermostat was established for the same amount of time using the same settings. Afterwards semiisotropic pressure coupling was established to keep a constant pressure of 1 bar with the Parrinello-Rahman barostat and a characteristic time of 5 ps. The same procedure was repeated for system #11 by considering the already extended membrane patch from system #10 and extension to a final edge length of 22.9 nm. The same minimization and equilibration procedure was performed as described above.

To connect the gp41 MPER-TMD structure in the extended and equilibrated membrane, the water and ion molecules were removed from the box. The principal axis of the equilibrated HIV-1 gp140 structure (simulation #9) was aligned along the membrane normal, and the structures heptad repeat 1 region (HR1) α -helices were centered along the principal axis of the gp41 TMD in the MPER-TMD model. For system #10 the HIV-1 gp140 construct was translated by 10 Å away from the lipid-water interface. In system #11 the construct was translated by 10 Å in both x and y directions, as well as tilted by $\sim 45^\circ$. Only this translation and tilt of the HIV-1 gp140 model allowed the placement of the 4E10 bnAb onto its corresponding MPER epitope in system #11. All missing residues in the 4E10 bnAb (PDB ID: 4XAW) were modeled (Section 4.2.12) and the bnAb was superimposed onto its corresponding MPER epitope (Section 4.2.11). Some lipids (less than five) needed to be removed from the system, as they were clashing with the inserted loop of the bnAb. For system #10 the N-terminal MPER α -helix was truncated by removing residues L660–N674, whereas in system #11 the C-terminal α -helices in the HIV-1 gp140 model were truncated, removing residues E654–D659. In both cases these removed residues

allowed the connection between the two models. The two structures were connected by unstructured loops using MODELLER. In a last step, the system was solvated with TIP3P water molecules and 0.15 M NaCl. In either case, the system was equilibrated using the same multi-step procedure as described previously in Section 4.2.6.

4.2.14 Contacts between antibody and its epitope

Contacts between the 4E10 bnAb and its MPER epitope were calculated for the whole trajectory of system #11. The definitions of the contacts were taken from Cardoso *et al.* [266] and can be found in Table 4.7. In every frame, the euclidean distance between the two atoms of interest was calculated.

Table 4.7: Putative hydrogen bonds and salt bridge contacts between 4E10 antibody and its MPER epitope as described by [266]. Superscript L and H indicate the light and heavy chains, respectively.

Peptide atom	Fab 4E10 atom	Distance (Å)
W670-O	S94 ^L -O _γ	3.4
N671-O _{δ1}	Y91 ^L -O	2.9
N671-N _{δ2}	S94 ^L -N	3.2
W672-N	S94 ^L -O _γ	3.2
W672-N _{ε1}	I56 ^H -O	3.2
N673-O _{δ1}	K32 ^L -N _ζ	3.4
T676-O _{γ1}	E95 ^H -O _{ε1}	3.0
T676-O _{γ1}	E95 ^H -O _{ε2}	2.8
K680-N _ζ	K100C ^H -O	2.7

4.3 Results

In the following section, the results of atomistic MD simulations of various HIV-1 gp41 constructs are described. The section starts with simulations of a monomeric gp41 TMD in a viral-like lipid bilayer and compares it with simulations of monomeric gp41 MPER-TMD. The transmembrane regions in both simulations are observed to tilt, such as to fit their hydrophobic core into the membrane. The section then continues to describe the dynamics of the trimeric gp41 TMD structure (PDB ID: 5JYN) in viral-like lipid bilayers, showing that these NMR structures are not stable in unbiased MD simulations. They rearrange at the C-terminus due to snorkeling of R696 towards the cytofacial leaflet. To assess if the experimentally used DMPC bilayers can stabilize the trimeric configuration, the gp41 MPER-TMD structure (PDB ID: 6E8W) is simulated in homogeneous DMPC bilayers using two different equilibration schemes. A rearrangement of these trimeric structures is observed. Afterwards, crystal structures of bnAbs are docked to showcase how the MPER recovers known bnAb epitopes in the unbiased MD simulations. To assess the accessibility of bnAbs to the MPER epitope in presence of the N-terminal Env glycoprotein, a crystal structure of HIV-1 gp140 (PDB ID: 5CEZ) is modeled on top of the gp41 MPER-TMD structure and simulated. A final simulation is described with a 4E10 bnAb (PDB ID: 4XAW) bound to its MPER epitope in an almost full length HIV-1 gp160 model. Implications on the accessibility of epitopes by bnAbs targeting the MPER are discussed.

4.3.1 R696 adapts two stable conformations in monomeric gp41 TMD

The influence of R696 on the conformation of the monomeric gp41 TMD was investigated. If not shielded, charged residues inside membranes tend to snorkel toward the lipid-water interface. A highly conserved arginine residue is found at the center of the TMD of HIV-1 gp41. To test whether R696 snorkels to either membrane leaflet, three independent simulations were set up. The monomeric gp41 TMD was positioned in the membrane such that R696 was found between the two leaflets. The simulations differed by rotamers defining the initial orientation of R696: towards the exofacial leaflet, perpendicular to the membrane normal, and towards the cytofacial leaflet (*up*, *middle* and *down*, respectively) (Figure 4.1A). Both R696 rotamers pointing upwards (red) and downwards (blue) are stable in their initial configuration for the total simulation time of 6.5 μ s, whereas R696 perpendicular to the membrane normal (green) is not stable and immediately points upwards during equilibration (Figure 4.1B).

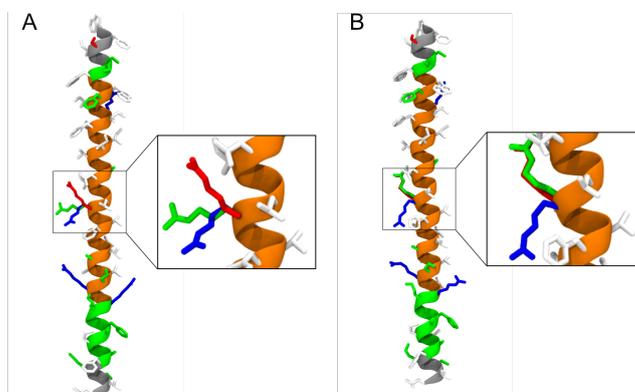


Figure 4.1: Orientation of R696 in a model lipid bilayer. The initial structure after energy minimization (A) shows three rotamers of R696, while the equilibrated structure (B) shows their stable orientation. The ribbon represents the helical fold of the TMD, with the gray area being the 50-mer used in the simulations reported here, the green area being the 40-mer used in two experimental studies [236, 237]. The orange area covers the R679–709 stretch, and was reported to form a consecutive α -helix [237]. Three possible rotamers of R696 are shown: up (red), middle (green) and down (blue). Water and lipid molecules are not shown.

4.3.2 The gp41 TMD exhibits two conformations depending on R696

The TMD adopts a tilted configuration in order to be entirely contained in the membrane. The three monomeric simulations from the previous section, with the TMD placed initially parallel to the membrane normal and with R696 roughly at the membrane center were each run for a total of 6.5 μ s. Immediately after equilibration the TMD tilts and R696 pulls water molecules into the membrane bilayer (Figure 4.2). R683 anchors the N-terminal part of the TMD to the lipid-water interface, both R707 and R709 anchor the C-terminal part. The N-terminal part of the TMD stays folded throughout the simulation, whereas the C-terminus unfolds. Overall, the TMD does not stay parallel to the membrane normal and tilts to accommodate larger regions of the α -helix inside the membrane.

While the R696 side chain is found almost at the center of the membrane in the *up* conformation, it is predominantly found at the cytofacial leaflet in the *down* conformation.

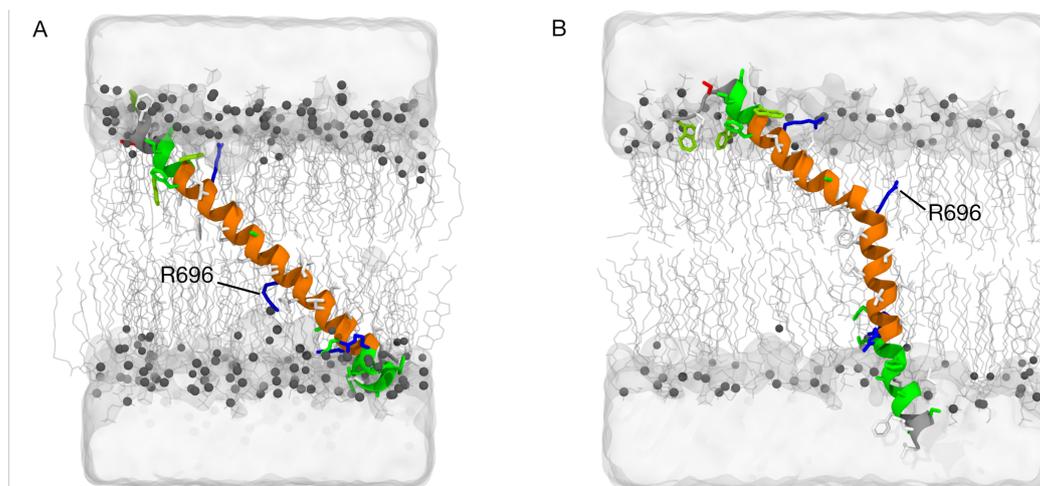


Figure 4.2: Stable configurations from two independent atomistic simulations with a different starting R696 rotamer. R696 either points downwards (A, system #3) or upwards (B, system #1), determining the orientation and overall placement of the TMD in the membrane. Water (gray surface) penetrates into the lipid bilayer in both conformations. The lipid bilayer represents only a small patch of the full membrane used in the simulation. Ribbon and side chain colors are the same as in Figure 4.1.

The tilt angle of the helix with respect to the membrane normal was calculated to quantify the differences between the two conformations. In the *down* monomer the angle was defined between the membrane normal and the vector spanning from the COM of C_{α} atoms of residues S703-G711 to I682-G690. The two monomers *up* and *middle* transiently kink at R696, leading to two separate α -helical regions. This kinking is stabilized by water molecules forming hydrogen bonds with the backbone atoms of the TMD.

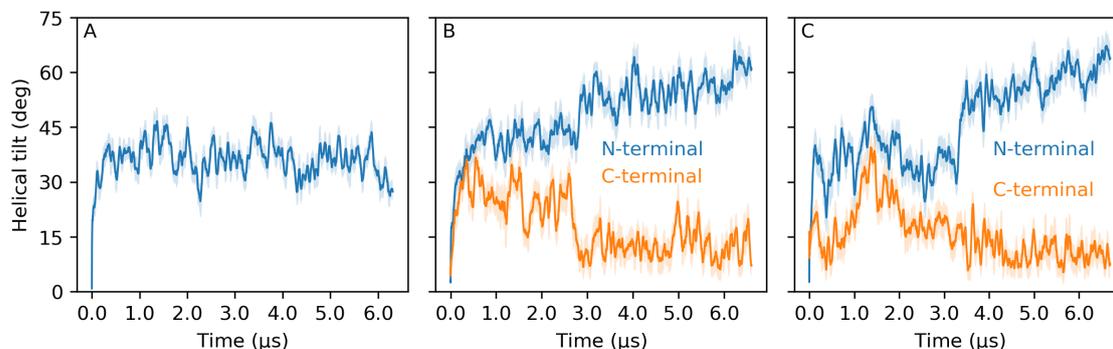


Figure 4.3: gp41 monomers adopt different tilt angles depending on the orientation of R696. The tilt angles of gp41 TMD α -helices from three independent simulations are shown as a function of time. (A, system #3) R696 in the initial *down* conformation. The center of mass of C_{α} atoms of residues S703-G711 to I682-G690 were used to calculate the tilt angle. (B, system #2) R696 in the *middle* and (C, system #1) in the *up* conformation. Here, two different tilt angles were calculated: L692-R696 to I682-I686 (N-terminal, blue) and R707-G711 to V698-L702 (C-terminal, orange). The time series (shaded area) were window averaged with a window size of 25 ns.

This interaction was quantified two by calculating tilt angles of the TMD to the membrane normal. The COM of C_{α} atoms of residues L692-R696 to I682-I686 were used at the N-terminus. At the C-terminal end residues R707-G711 to V698-L702 were considered. A tilt angle was calculated for both regions in the TMD. The monomer in the *down*

conformation displays an angle of $37.0 \pm 12.4^\circ$ to the membrane normal, whereas the N-terminal helical segments of the upwards pointing monomers display angles of $39.4 \pm 8.6^\circ$ and $45.6 \pm 7.2^\circ$, *up* and *middle*, respectively (Figure 4.3). The N-terminal segments display an angle similar to the one displayed by the monomer pointing downwards, whereas the C-terminal segments exhibit a smaller tilt angle.

4.3.3 Interactions of MPER with the membrane depend on TMD orientation

A longer α -helix containing both the sequence of MPER and TMD was simulated to test whether the orientation of R696 has an impact on the MPER. The TMD is preceded by a short 23-residue MPER domain that is rich in highly conserved tryptophan residues. Two simulations were set up embedding the MPER-TMD construct into the membrane, with R696 either pointing *up* or *down* as shown in the previous sections. The N-terminal residues N656-E662 start to unfold after $\sim 1.2 \mu\text{s}$ in both simulations, whereas the residual 54 residues (L663-S716) stay α -helical. The hydrophobic side of MPER, especially W666 and W670, are unable to embed into the membrane bilayer when R696 snorkels to the cytofacial membrane leaflet (Figure 4.4A). On the contrary, in simulations where R696 points to the exofacial membrane leaflet, a transient kink forms at residues N671 and W678 or R696, providing the N-terminal part of MPER with the flexibility to interact with the membrane (Figure 4.4B). Under these conditions the hydrophobic residues W666, W670, W672 and F673 are able to interact with the membrane at the lipid-water interface.

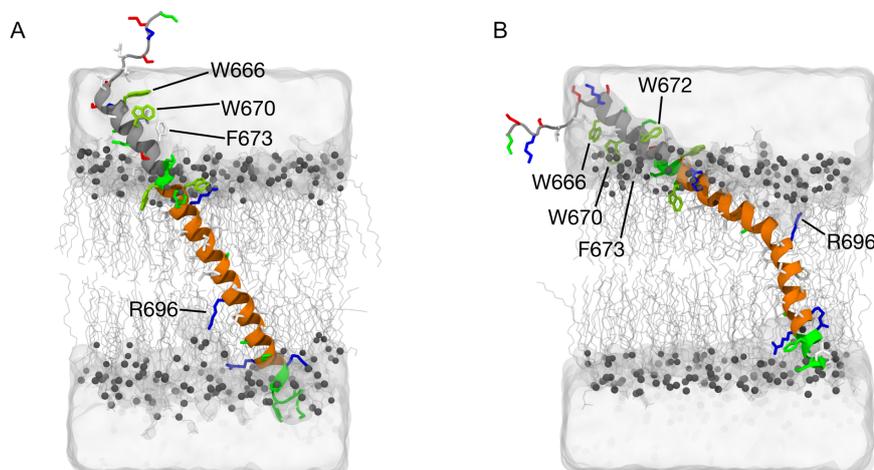


Figure 4.4: The orientation of R696 determines different conformations of gp41 MPER. When R696 points to the lower leaflet (A, system #4), the hydrophobic side of MPER is unable to interact with the membrane. If R696 points to the upper leaflet (B, system #5), MPER acts as an extension to the TMD and embeds its hydrophobic tryptophan residues. Colors are the same as in Figure 4.2.

The tilt angle of the TMD was calculated for both simulations with the same definitions as in the previous section (Figure 4.5). In the R696 *down* conformation the TMD adopts a similar tilt angle as shown for the TMD-only construct in the previous section (Figure 4.5A). In contrast, the construct with an R696 initially in the *up* conformation shows different tilt angles for its two α -helical sections. The N-terminal section tilts to an angle of around $\sim 50^\circ$ in the first $0.5 \mu\text{s}$ and converges to $\sim 45^\circ$ over the next $3.5 \mu\text{s}$ (Figure 4.5B). The C-terminal

section of the gp41 TMD is tilted by $\sim 45^\circ$ for the whole simulation. The tilt angles of R696 in the *up* conformation from the previous section show that the C-terminal section of the TMD is less and the N-terminal section is more tilted. The MPER-TMD construct with R696 in the *up* conformation contains an extended α -helical structure that requires it to insert into the membrane. This insertion forces the whole MPER-TMD construct to adopt these similar tilt angles.

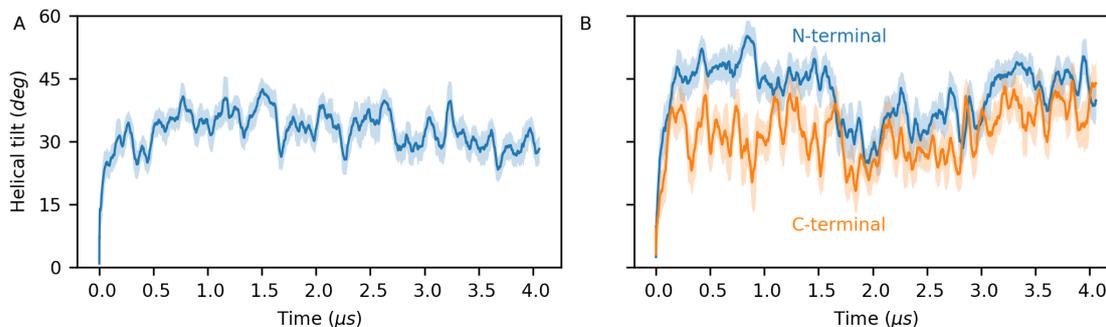


Figure 4.5: gp41 MPER-TMD constructs adopt the same tilt angle. The tilt angles of gp41 MPER-TMD α -helices from two independent simulations (A and B, systems #4 and #5) are shown as a function of time. (A) The TMD remains fully folded and does not kink. The tilt angle was calculated by spanning a vector between S703-G711 and I682-G690. (B) The TMD kinks over the course of the simulation. Two separate tilt angles were calculated using residues L692-R696 to I682-I686 (N-terminal, blue) and R707-G711 to V698-L702 (C-terminal, orange). The time series (shaded area) were window averaged with a window size of 25 ns.

4.3.4 Trimeric gp41 TMD rearranges in viral-like membranes

The trimeric NMR structure of gp41 TMD (PDB: 5JYN) [236] was simulated in a viral-like bilayer in two independent runs for 10 μ s each. In the initial starting structure the TMDs are arranged in a tight trimeric bundle with no apparent tilting. Three charged arginine R696 residues are oriented along the plane of the membrane (Figure 4.6A). In less than 100 ns of unrestrained atomistic MD simulation all three arginine residues rotate and point towards the cytofacial membrane leaflet (Figure 4.6B). The rearrangement of the arginines has three consequences: (i) water and lipid headgroups are pulled into the cytofacial leaflet, (ii) the C-terminal regions of the gp41 TMD protomers start to dissociate from one another and (iii) the helices adopt different tilt angles.

The orientation of R696 with respect to the normal of the membrane was calculated to quantify its rearrangement. The angle was defined to be formed by the vector spanning the C_α and C_ζ atoms of R696 and the membrane normal (Figure 4.7A). This angle increases in the first 100 ns and converges to a value of $120 \pm 8.5^\circ$ for both independent runs. The immediate increase of the angle highlights how the charged side chains of R696 snorkel towards the lipid-water interface.

To quantify the dissociation of the trimeric structure, the pairwise distances between the N- and C-termini of each protomer were calculated (Figure 4.8B). The N-terminal distances do not change over the course of the trajectory, whereas all C-terminal distances increase in the first 100 ns. These distances stay constant for the first 3 μ s of the simulation

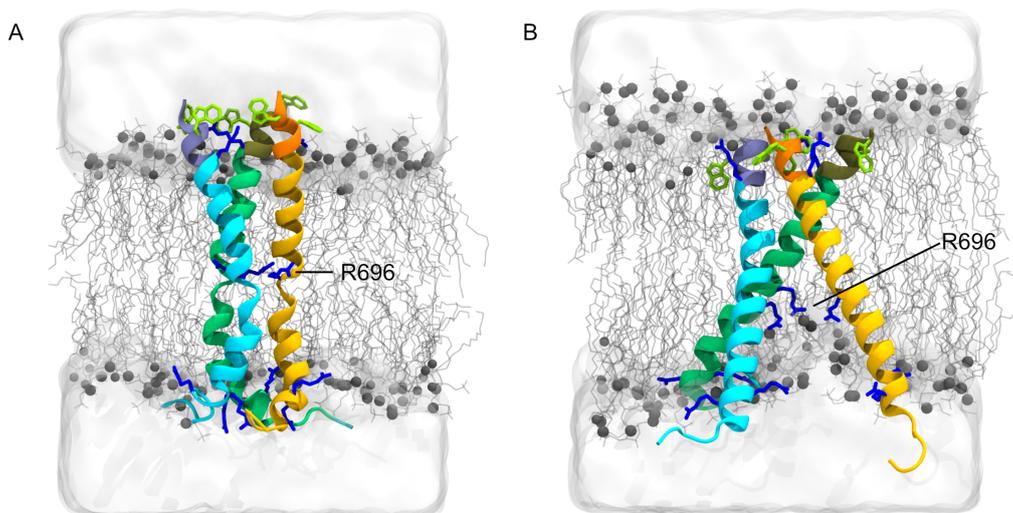


Figure 4.6: Trimeric gp41 TMD rearranges during atomistic simulations (system #6A). Snapshots of (A) the trimeric gp41 TMD structure (PDB: 5JYN) embedded into a lipid bilayer and (B) after an extensive equilibrium simulation of 10 μ s. Protein protomers are shown as ribbons, where darker shades indicate the MPER. Arginine (blue) and tryptophan (green) side chains are shown as sticks. Lipid headgroups are shown as gray beads, their acyl tails as sticks, and water as white transparent surface.

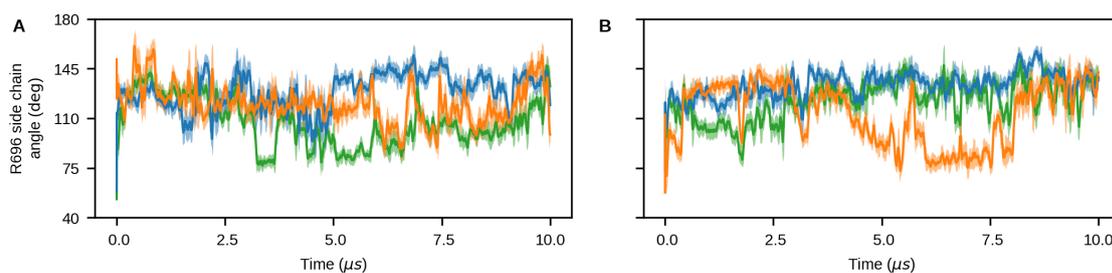


Figure 4.7: Angle between each R696 residue and the membrane normal for two independent systems A and B (systems #6). The side chains immediately tilt toward the cytofacial leaflet (high angles). Each arginine is plotted in a different color (green, blue and orange). The solid line shows the window average, while the shaded area represents the standard deviation of each window.

and start to increase thereafter.

The dissociation of the N- and C-termini is accompanied by an increasing tilt angle of each protomer (Figure 4.9). The vector spanning the COM of C_{α} backbone atoms from residues L702-Q710 to Y681-V689 was considered for the calculation of the tilt angle with respect to the membrane normal. A rapid increase of the tilt angle can be observed over the first 500 ns (Figure 4.9A and B). All three protomers display a different tilt angle in the following 2.5 μ s, with the smallest one at $13.4 \pm 4.0^{\circ}$ for protomer 2 (blue), while protomer 1 (green) displays an angle of $20.7 \pm 4.4^{\circ}$ and protomer 3 (orange) an angle of $29.5 \pm 4.5^{\circ}$ (Figure 4.9B). At 3 μ s another conformational change happens and all protomers start to converge to a tilt angle of $\sim 30^{\circ}$, consistent with that of a monomeric gp41 TMD described in the previous section.

These drastic changes in the tilt angle can be explained with changes in the trimeric configuration (Figure 4.10). The simulations start from the C_3 symmetric configuration

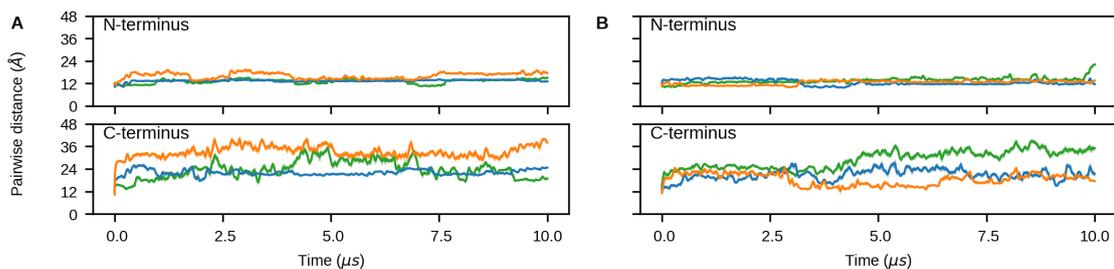


Figure 4.8: Pairwise N- and C-terminal protomer distances for two independent simulations (A and B, systems #6). The N-termini of all three protomers do not dissociate (top), whereas the C-termini move apart significantly at the beginning of the trajectory and after 3 μs (bottom). Distances are shown for three protomer pairs in green, blue and orange.

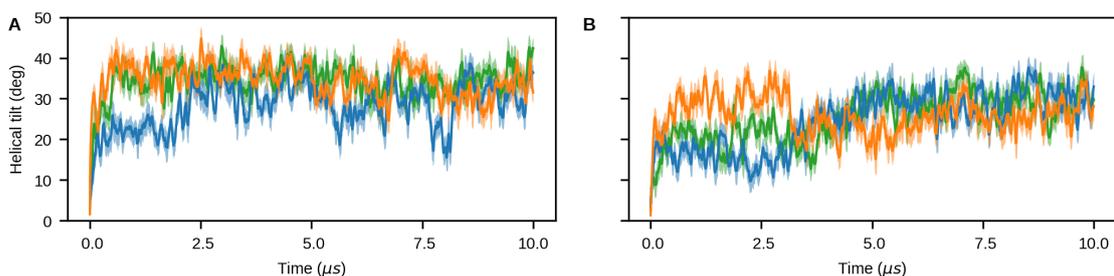


Figure 4.9: Tilt angle of the three α -helical protomers for two independent simulations (A and B, systems #6). Each protomer adopts a tilt of 20–40° after several microseconds. The three protomers are plotted in different colors (green, blue and orange).

(Figure 4.10, $t = 0.0 \mu\text{s}$). During the first 500 ns, the C-termini of the protomers start to dissociate from each other, leading to a higher tilt angle (Figure 4.10, $t = 2.5 \mu\text{s}$). After 3 μs another change in the configuration happens and the trimer starts to lose C_3 symmetry, due to reorganization of the N-terminal contacts (Figure 4.10, $t = 7.5 \mu\text{s}$). The contacts are not regained for the entire duration of the simulation (Figure 4.10, $t = 10.0 \mu\text{s}$).

4.3.5 R696 snorkels to the cytofacial leaflet in the trimeric assembly

The rearrangement of R696 is characterized by the snorkeling of its polar atoms to the lipid-water interface (Figure 4.11). While the C_α atom is found almost at the center of the membrane, the C_ζ atom moves along the z -axis towards one of the leaflets. Both water molecules and chloride ions interact with the polar atoms of R696, with the latter demonstrating a visible density at a distance of $\sim 10 \text{ \AA}$ (Figure 4.11).

Water molecules preferentially interact with the arginine residues present in the gp41 TMD. To quantify the exposure of R696 to water, the number of hydrogen bonds between every residue and the solvent was calculated (Figure 4.12).

The number of hydrogen bonds was calculated for every frame and then averaged over the whole trajectory and over all protomers. All four arginines form on average 9–12 hydrogen bonds with water molecules, whereas the remaining side chains at the N- and C-terminus of each protomer form at most 9 hydrogen bonds. Consistent with the observations from previous sections, R696 interacts with water molecules by formation of hydrogen bonds. The arginine residues R683 at the N-terminus, as well as R707 and R709

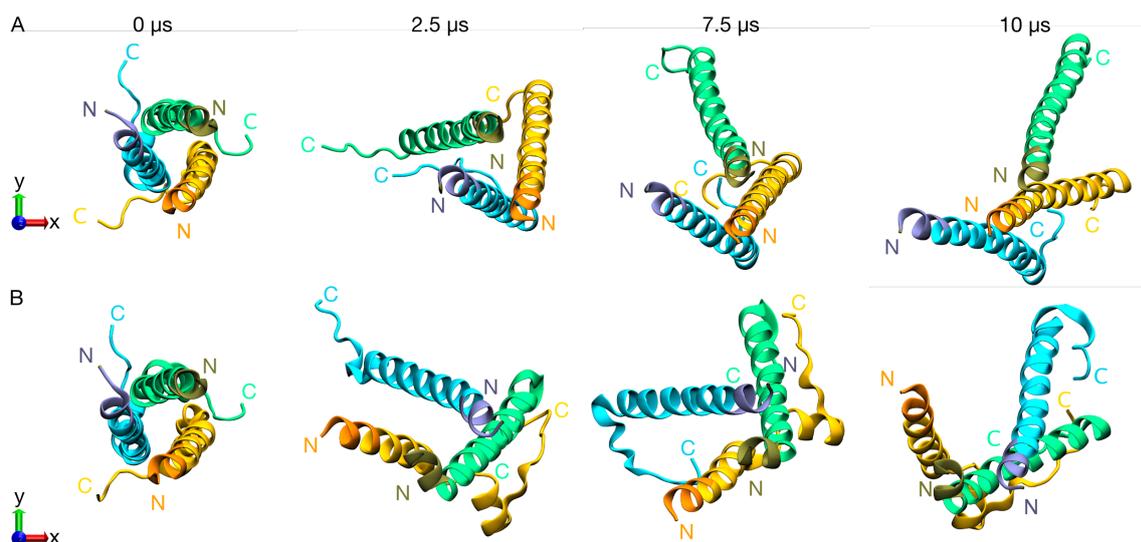


Figure 4.10: Snapshots at 2.5 μs intervals of trimeric gp41 TMD simulations embedded into lipid bilayers. The two rows show two independent simulations (A and B, systems #6) from the exofacial side of the membrane. The trimeric bundle starts with axial C_3 symmetry ($t = 0.0 \mu\text{s}$), dissociates at the C-terminus ($t = 2.5 \mu\text{s}$), completely loses C_3 symmetry ($t = 7.5 \mu\text{s}$) and does not regain contact for the remaining simulation ($t = 10.0 \mu\text{s}$). Ribbon colors are the same as in Figure 4.6. Lipid and water molecules are not shown.

at the C-terminus are found at the lipid-water interface and act as anchors for the TMD. Therefore, their involvement in hydrogen bonding with water molecules is expected.

4.3.6 Arginines restrict possible placements of TMD in lipid bilayers

Simulations with symmetric DMPC bilayers were set up to test if specific lipid properties stabilize the trimeric NMR structures. In the following sections the trimeric MPER-TMD structure (PDB ID: 6E8W) was used [239]. The possible positioning of the arginines found in the TMD were assessed. As described in the previous sections, the TMD is flanked by three arginine residues, R683 at its N-terminus, and R707 and R709 at its C-terminus. These residues tether the TMD at the lipid-water interface and limit its possible movement.

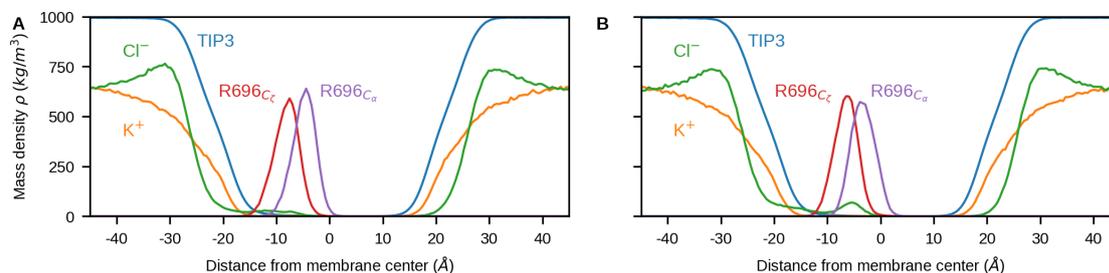


Figure 4.11: R696 snorkels to the cytofacial leaflet with its polar side chain. Mass density plot of selected molecules averaged for two independent simulations (A and B, systems #6) over a 10 μs each, and centered around the membrane. Water (blue), chloride (green) and potassium (orange) ions depict the soluble part of the system. The C_α (purple) and C_ζ (red) atoms of R696 are found inside the membrane, while C_ζ snorkels to the cytofacial leaflet. The following densities were multiplied by the given factor to become visible in the plot: chloride and potassium ($\times 100$), $R696_{C_\zeta}$ and $R696_{C_\alpha}$ ($\times 10000$).

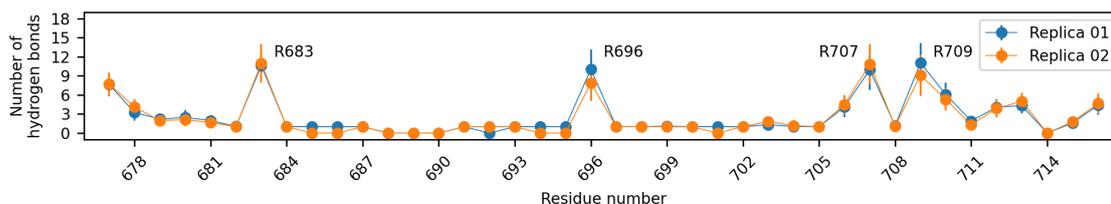


Figure 4.12: Arginine residues form hydrogen bonds with water molecules. The average number of hydrogen bonds formed by each residue of the TMD is shown with the accompanying standard deviation. The number of hydrogen bonds was calculated for each frame and then averaged over the whole trajectory of 10 μ s. All four arginines found in the TMD are annotated. The number of hydrogen bonds are shown for two independent simulations (orange and blue, systems #6).

In addition, a fourth and highly conserved arginine residue R696 is found at the center of the TMD. This residue always snorkels towards the lipid-water interface independent of the oligomeric state of the TMD, as shown in the previous sections. To test how the four arginines restrict a possible positioning of the TMD in a lipid bilayer, the hydrophobicity of the TMD was analyzed. Five different windows with a length of 18 amino acids were characterized with HeliQuest (Figure 4.13).

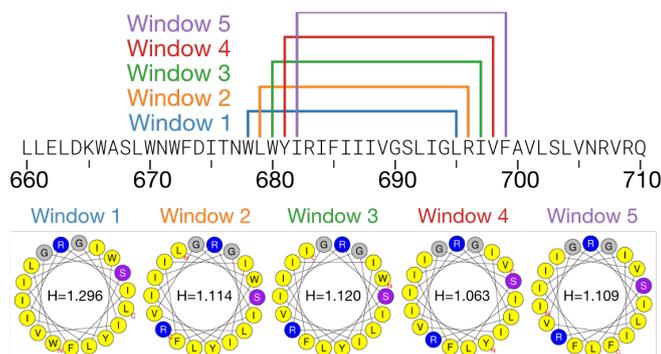


Figure 4.13: Hydrophobicity of HIV-1 Clade D sequence (L660-Q710) along 18 amino acid windows. Window 1 (W678-L695, blue) is the most hydrophobic stretch in the 51 amino acid long sequence, whereas window 3 (W680-I697, green) is less hydrophobic but contains both R683 and R696. The helical wheels were generated with HeliQuest [258].

The highest hydrophobicity is found in the first window, because it only includes a single arginine. Shifting the windows towards the C-terminus leads to the inclusion of R696, lowering the hydrophobicity. Moving the window further will ultimately lead to the inclusion of the two C-terminal arginine residues, R707 and R709. Therefore, the two viable options are to test the first window with only R683 at the exofacial leaflet or the third window with both R683 and R696 anchoring the TMD at the respective lipid-water interfaces. In both cases, R707 and R709 would be positioned outside the membrane and be completely exposed to the solvent. Figure 4.14 shows the insertion of the trimeric MPER-TMD construct (PDB ID: 6E8W) into a DMPC bilayer.

The bulk of the hydrophobicity is found after R683 and before R696, but the hydrophobic region continues towards the C-terminus just before R707. Two short simulations were set up with the trimeric gp41 MPER-TMD positioned in a DMPC membrane such that either R696 was put outside the membrane just below the cytofacial leaflet (window 1

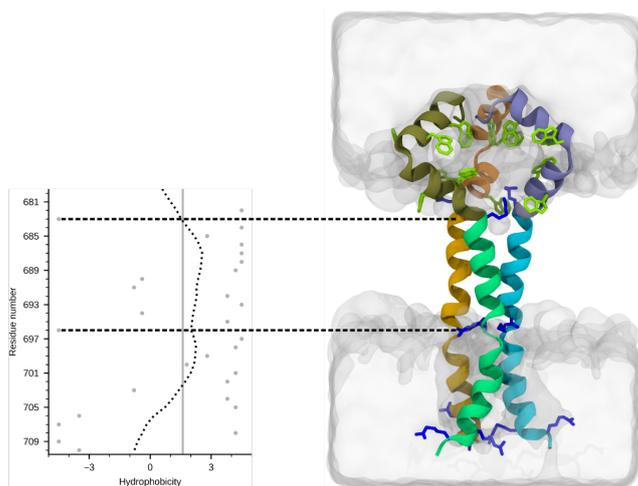


Figure 4.14: (Right) Possible insertion of trimeric gp41 MPER-TMD structure in a thin DMPC bilayer. (Left) Hydrophobicity profile of gp41 TMD. The hydrophobicity of individual amino acids (gray circles) was averaged with a window size of 19 amino acids (dotted curve). The solid vertical line represents a hydrophobicity value of 1.6, a threshold for identifying transmembrane domains. Horizontal dashed lines indicate residues R683 and R696.

in Figure 4.13) or directly at the lipid-water interface (window 3 in Figure 4.14 and Figure 4.14). Both of these structures equilibrated such that R696 is immediately pushed inside of the membrane, resulting in a similar relaxed structure as seen in the previous sections with the trimeric gp41 TMD constructs. It is therefore suggestive to argue that R696 indeed must be positioned close to or at the center of the membrane, because the hydrophobic mismatch of the TMD requires proper insertion of the helices. Ultimately, all further simulations were set up with R696 inside the lipid bilayer (systems #7–8).

4.3.7 Rearrangement of trimeric structures in DMPC bilayers

The dynamics of trimeric gp41 MPER-TMD structures were studied in symmetric DMPC bilayers. In the last section it has been established that the positioning of R696 outside the membrane is not suitable due to the hydrophobic mismatch of the TMD. Therefore, the trimeric structure was positioned such to include R696 inside of the membrane. A DMPC bilayer was used as model bilayer for these simulations. In the NMR experiments in which the gp41 TMD structures were solved, bicelles composed of DMPC and 1,2-dihexanoyl-sn-glycero-3-phosphocholine (DHPC) were used. DMPC is considered to form the bilayer and DHPC shields the hydrophobic sides of the membrane with its short acyl chain. Due to the PBC in MD simulations a DMPC model membrane sufficiently represents the bilayer part of a bicelle. Figure 4.15 shows snapshots of a trajectory in 200 ns increments. The thin DMPC membrane leads to a rapid dissociation of the TMD C-termini after less than 200 ns. A conical shape is formed over the course of the total 1 μ s simulation time, with the three R696 residues pointing towards the cytofacial leaflet. The phosphate moieties of lipid molecules are pulled upwards into the core of the membrane.

A total of six gp41 MPER-TMD simulations were set up with the TMD placed inside a DMPC bilayer. To check if the relaxation of the structures leads to the observed dissociation, two different equilibration schemes were applied to three simulations each. The two schemes

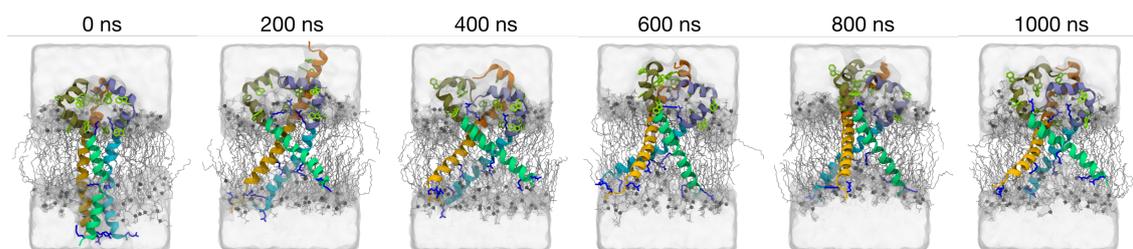


Figure 4.15: HIV-1 gp41 MPER-TMD trimer rearranges into conical trimers. Snapshots taken between 0 and 1000 ns from a simulation of the gp41 MPER-TMD trimer in a DMPC bilayer (system #7A). The C_3 symmetry of the initial structure is lost after hundreds of nanoseconds. Lipids are pulled towards the membrane center by the charged R696 residues.

differ by the atoms on which restraints are applied during the equilibration. All heavy atoms were restrained for an additional 50 ns equilibration step in one scheme (Figure 4.16A), whereas only the backbone heavy atoms were restrained in the other (Figure 4.16B). This means that in the latter the heavy backbone atoms of the side chains were allowed to relax without any restraints. Independent of the equilibration scheme, the trimeric gp41 MPER-TMD structures rearrange into either conical or tent-shaped trimers, as shown in Figure 4.16. Conical trimers are characterized by their symmetry and almost equidistant protomers, as showcased in the first two panels of Figure 4.16A. In contrast, tent-shaped trimers have two close and one distant protomer (Figure 4.16B, second panel). Again, in all six simulations R696 are found roughly at the center of the membrane and snorkel towards the cytofacial leaflet.

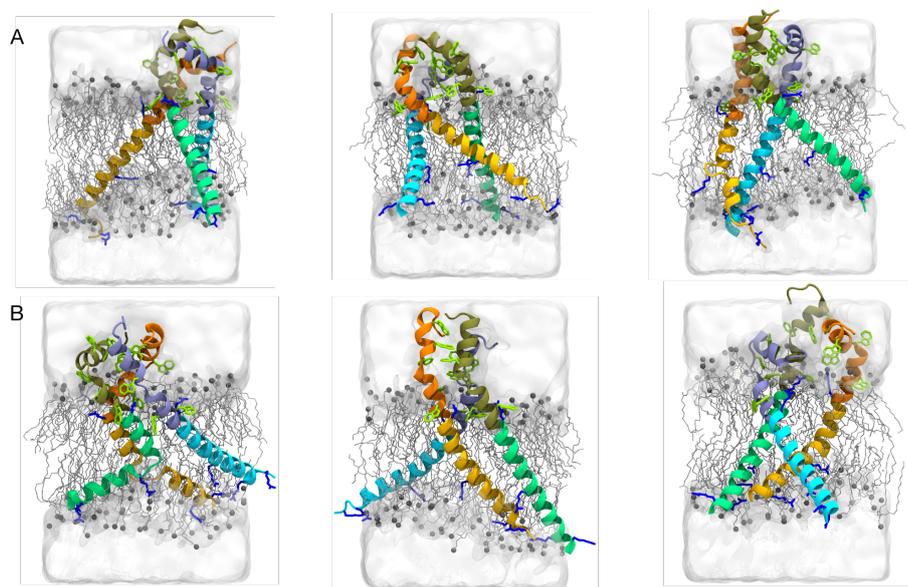


Figure 4.16: Snapshots of equilibrated structures after 1 μ s from six independent MD simulations of gp41 MPER-TMD in DMPC bilayers. The trimers are either found to be conical or tent-shaped. The systems were independently relaxed with restraints on (A) all heavy atoms (systems #7) or (B) heavy backbone atoms (systems #8).

The initial structure of MPER consists of two distinct short α -helices interrupted by a short turn [239]. The N-terminal helix consists of L663–N671 and connects to the second helix (I675–R683) through the turn formed by W672–D674. An additional kink

at R683 separates the second MPER helix from the TMD. In all simulations performed with this gp41 MPER-TMD construct, the C-terminal kink at R683 disappears and a long, continuous α -helix is formed between the C-terminal MPER helix and the TMD. This behavior is similar to that observed in the simulations of the monomeric TMD and MPER-TMD constructs that form a long, continuous and uninterrupted α -helix, as well as the observations reported by Chiliveri *et al.* [237].

Figure 4.17 shows the pairwise distances of protomers' TMD N- and C-termini. As stated earlier no clear difference can be observed for the two equilibration schemes (Figure 4.17A and B). In all six simulations the N-terminal distances approach values of 20 Å and stay in close proximity to another. In contrast, the C-termini dissociate more visibly and obtain distances between 50 Å to 70 Å. The dissociation starts in the first hundreds of nanoseconds and does not proceed after 1 Å. The distinction between conical and tent-shaped trimers, as made in the preceding paragraph, can be clearly explained with the C-terminal pairwise distances. The first and second columns highlight the protomers being almost equidistant and forming a conical trimers (Figure 4.17A). The third columns show how the distance of two pairs is big, but one is small (Figure 4.17A and B). In these cases a tent-shape is formed by the gp41 MPER-TMD trimers.

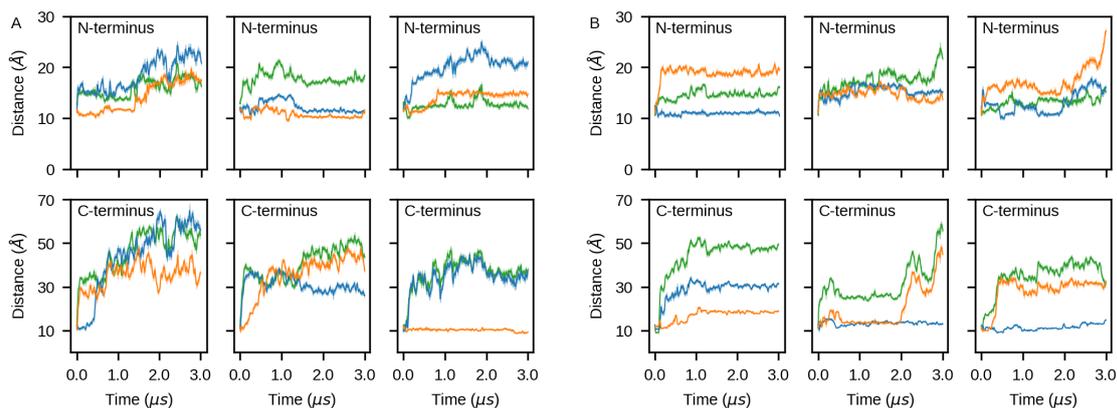


Figure 4.17: Pairwise distances between the N-termini (top rows) and the C-termini (bottom rows) of gp41 MPER-TMD shown as blue, orange, and green lines for three protomer pairs. (A) All heavy atoms (systems #7) and (B) heavy backbone atoms (systems #8) were relaxed with restraints during equilibration. Columns represent three independent simulations.

4.3.8 Formation of MPER epitopes accessible by bnAbs

The MPER domain explores conformations of identified bnAb epitopes in the MD simulations of gp41 MPER-TMD. Multiple bnAbs were identified to target the MPER domain. In particular, at least two bnAbs are known to interact with the membrane: 4E10 and 10E8 [260, 261]. Crystal structures of the bnAb in complex with the epitope also include glycerol backbones of GPLs, indicating a certain orientation and proximity of the bnAb to the membrane. To identify whether the previously described simulations of gp41 MPER-TMD recover the epitopes of these bnAbs, three crystal structures were chosen and studied. A single 4E10 structure (PDB ID: 4XAW [260]) and two 10E8 structures (PDB IDs: 5T6L and 4G6F [261, 262]) were used. Each bnAb recognizes a different linear motif as its epi-

tope. For 4XAW and 5T6L its a short α -helix of either 13 or 16 residues, whereas 4G6F recognizes two distinct α -helices separated by a kink. The epitopes are listed in Table 4.6. Figure 4.18 shows the three co-crystallized gp41 MPER-TMD fragments as pink ribbons superimposed onto the corresponding regions in the gp41 MPER-TMD simulations. Surprisingly, the unbiased MD simulations manage to recover the epitopes and present them at the lipid-water interface.

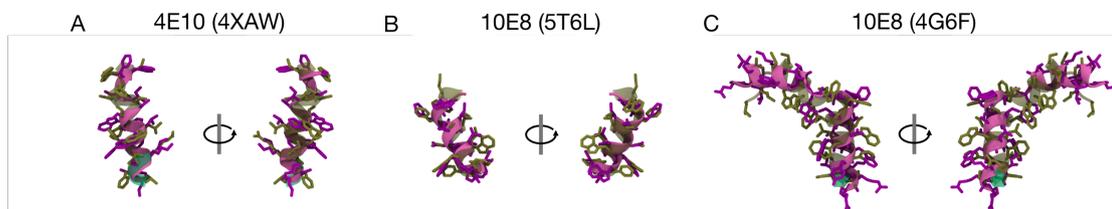


Figure 4.18: Superimposition of the co-crystallized gp41 MPER-TMD fragments (pink ribbon and magenta sticks) onto the corresponding regions (brown-green ribbons and sticks) from MD simulations. Each bnAb is shown from two angles. (A) 4E10 epitope N671-K686 (PDB ID: 4XAW) superimposed onto N671-I686. (B) 10E8 epitope N119-D131 (PDB ID: 5T6L) superimposed onto N671-R683. (C) 10E8 epitope E659-R685 (PDB ID: 4G6F) superimposed onto L660-F685.

The co-crystallized epitopes bound to the bnAb structures were superimposed onto the corresponding regions of gp41 MPER-TMD using rigid-body root-mean-square distance alignment. All of these three co-crystallized bnAb epitopes were superimposed onto every gp41 MPER-TMD protomer in each frame of the six $3\ \mu\text{s}$ simulations. Only one protomer per simulation recovered the epitope, if at all. Possible candidates were first selected by lowest RMSD and subsequent visual checks to avoid clashes between the bnAb's Fab and the MPER domain. Figure 4.19 shows three snapshots from different gp41 MPER-TMD simulations each with a bnAb superimposed onto its corresponding epitope. The simulated structures present the epitopes such that each bnAb is (i) able to interact with the corresponding gp41 MPER-TMD region and (ii) bury its lipid-interacting residues in the lipid-water interface of the membrane.

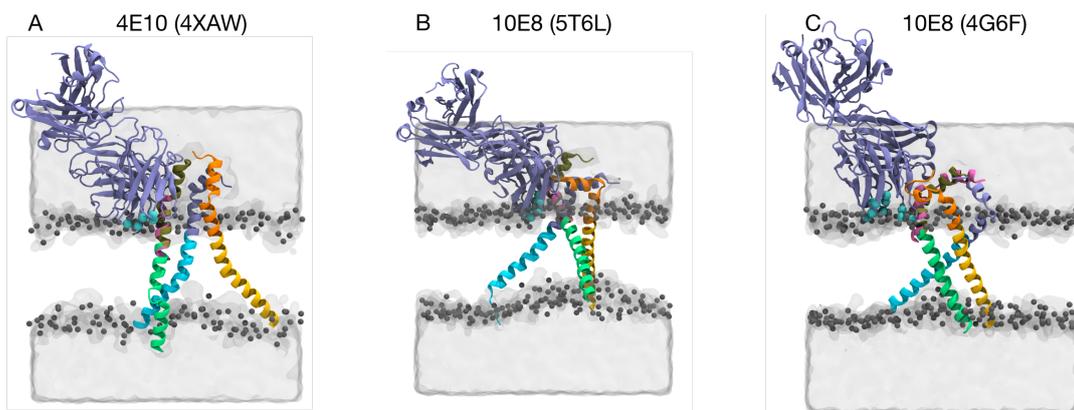


Figure 4.19: Snapshots of bnAb epitopes superimposed onto the corresponding MPER-TMD segment from gp41 trimers. (A) 4E10 (PDB ID: 4XAW), (B) 10E8 (PDB ID: 5T6L), and (C) 10E8 (PDB ID: 4G6F). Colors are as in Figure 4.6, with the bnAb in blue and the MPER epitope in pink. C_{α} atoms of lipid-interacting residues in the bnAb are shown as cyan spheres.

The positioning of each bnAb is slightly different, even though the light and heavy

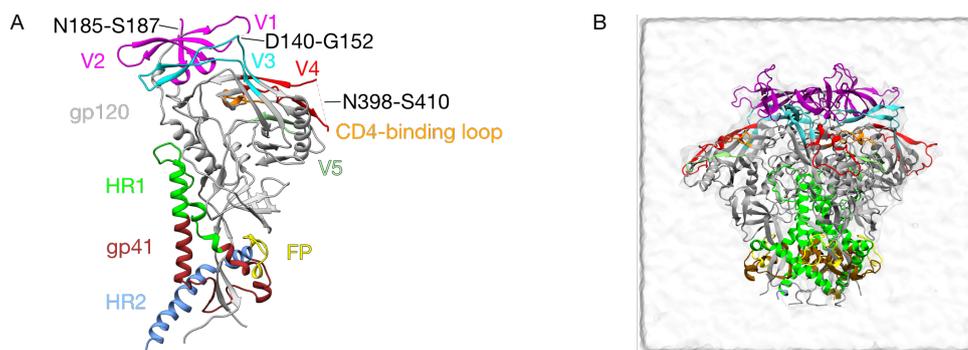


Figure 4.20: (A) Missing residues in 5CEZ crystal structure of HIV-1 Env gp140. The two regions gp120 (gray) and gp41 (brown) are shown as ribbons with their domains in distinct colors. gp120: V1 and V2 (magenta), V3 (cyan), CD4-binding loop (orange), V4 (red) and V5 (lime). gp41: FP (yellow), HR1 (green) and HR2 (blue). The structure of three stretches (two in V1, one in V4) are unresolved. The residue sequences are annotated accordingly. (B) Fully modeled HIV-1 Env surrounded by water (white surface) after equilibration, system #9. Ions are not shown for clarity. Colors are the same as in (A).

chains are identical. For example, the 10E8 bnAb almost lies flat on the membrane (Figure 4.18B). In contrast, the other two bnAbs, and most prominently the second 10E8 bnAb, stand upright on the membrane at a very high angle (Figure 4.18A and C). These differences in orientation pose an important question: are the MPER-TMD epitopes accessible to antibodies at all times or must the N-terminal gp120 protein undergo some global conformational change?

4.3.9 HIV-1 gp140 simulations in solution

A single atomistic MD simulation of HIV-1 gp140 was run in solution to check its stability, before assessing if the epitopes are accessible to bnAbs in a full-length construct. More than 16 Env trimer structures are available to date with a resolution of $<6 \text{ \AA}$ [267]. All of these structures are a chimera of gp120 and gp41, called gp140. They contain the full length gp120 sequence and parts of gp41, but lack most of the MPER, TMD and CTD. The biggest part of the trimeric fold is conserved across all of them. The structure 5CEZ [253] was chosen for the herein described simulation because it (i) shows a C_3 symmetry allowing an easier set up, (ii) has only a few unresolved residues and most importantly (iii) the HR1-HR2 regions in the core of Env are fully resolved. In some of the available structures the HR1-HR2 regions are unresolved and thus require extensive modeling. Instead, the 5CEZ crystal structure is only missing a total of 26 residues shared between two variable regions V1 and V4 at its apex (Figure 4.20A).

The HIV-1 Env gp140 model is stable in water and keeps its general fold. The unresolved residues were modeled as unstructured loops with UCSF CHIMERA and subsequently the trimer was assembled (Figure 4.20B). The trimeric gp140 model was simulated in solution at 310 K for a total of 650 ns. Visual inspection did not reveal any structural rearrangements. To quantify these observations, the RMSF of every side chain was calculated (Figure 4.21A). Most residues show a RMSF of less than 3 \AA , whereas specific regions have higher fluctuations. Ignoring the N- and C-terminus of the model, the RMSF reports on

the flexibility of the V1, V2, V4 and V5 variable loops, as well as on the major fluctuations of the fusion peptide (FP). In addition, the HR1 is seen to fluctuate as well. All residues with a RMSF $< 3 \text{ \AA}$ were selected for the RMSD calculations (Figure 4.21B). The structure show some rearrangements of 2 \AA in the first tens of nanoseconds and then converges to a RMSD of 3.5 \AA .

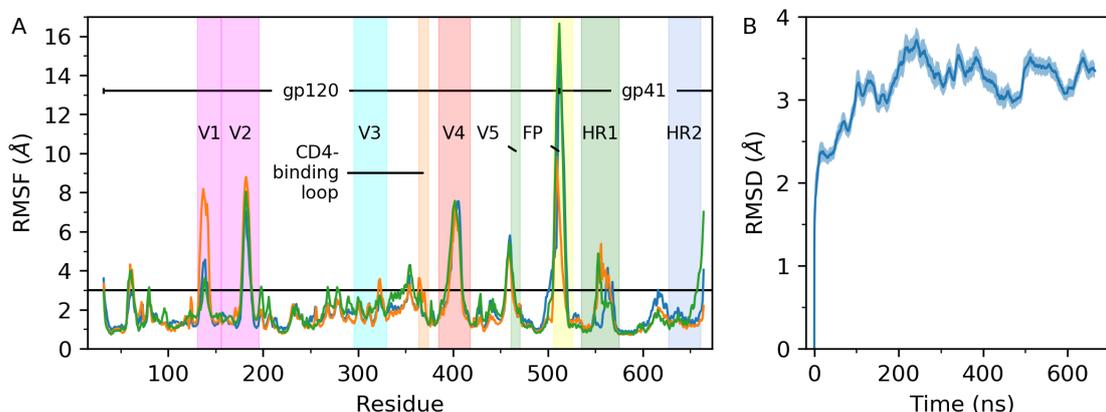


Figure 4.21: Observables from atomistic HIV-1 Env MD simulation in solution. (A) RMSF as a function of the residue number for the total simulation time of 665 ns. The RMSF of every protomer was calculated separately (blue, orange and green lines). The separate regions are annotated using similar colors as in Figure 4.20. The horizontal line at 3 \AA indicates a threshold used in the other panel. (B) RMSD as a function of time for all residues with a RMSF ≤ 3 . The instantaneous structure was superimposed onto the starting structure to calculate the RMSD.

4.3.10 Construction of an almost full-length HIV-1 model

The model of Env glycoprotein gp140 was connected with an equilibrated model of gp41 MPER-TMD. The available structures of HIV-1 gp140 and gp41 MPER-TMD overlap in four residues at the structures' C- and N-termini, respectively. The secondary structures of these residues is α -helical in the gp140 model, while it is unstructured (L660-E672) and α -helical (L663-D664) in the gp41 MPER-TMD model. Unfortunately, the tertiary structure poses another restraint on these structures. While the C-terminal α -helices in the gp140 construct are separated by about 30 \AA , the N-terminal α -helices in the gp41 MPER-TMD NMR structure point towards each other and are rather close. In all six $3 \mu\text{s}$ simulations of gp41 MPER-TMD the N-terminal helices stay close, but most importantly lose their initial C_3 symmetry. Given that the C-terminal α -helices of gp140 remain C_3 symmetric after more than 650 ns simulation, it is impossible to simply merge both models. Instead, it was necessary to truncate the gp41 MPER-TMD models at the N-terminal MPER α -helix (L660-N674). This removed the N-terminal α -helix of MPER, as well as the short turn connecting the two α -helices. Each of the six independent gp41 MPER-TMD simulations was considered for a distance analysis to find the best structure to use for modeling. This allowed to determine which, if any, of the simulations is suitable to be chosen as the one to connect the two models. Figure 4.22A shows the pairwise N-terminal distances between the C_α atoms of I675 of each protomer. Only two independent simulations in systems #7 (#7A and #7B) are shown. These are the only systems with

suitable distances. In comparison, #7B shows pairwise distances between 20 Å to 30 Å, whereas the N-terminal MPER in system #7A all have distances of 15 Å or less. As the C-terminal helices in the HIV-1 gp140 model all display a distance of around 30 Å (Figure 4.22B), system #7B was chosen as being suitable to connect the two models.

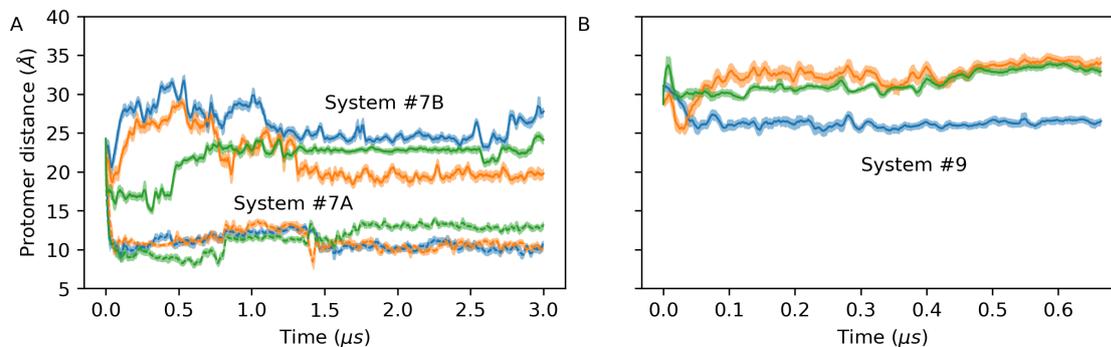


Figure 4.22: Terminal distances in HIV-1 gp140 and gp41 MPER-TMD constructs. (A) Pairwise distances between the N-termini of three protomers in gp41 MPER-TMD simulations from two independent simulations (solid and dashed lines). The distances were calculated between the C_{α} atoms of I675. (B) Pairwise distance between the C-termini in a HIV-1 gp160 simulation. The distances were calculated between the C_{α} atoms of N656. Each pairwise distance is shown in a separate color (blue, orange and green).

After removing the N-terminal MPER α -helix (L660-N674) the equilibrated HIV-1 gp140 model was put at 10 Å distance above the membrane. The models' HR1 α -helices were centered along the principal axis of the gp41 TMD. The C-terminal ends of the HIV-1 gp140 model are tilted at approximately $\sim 45^{\circ}$ facing away from each other (Figure 4.20B). The model was rotated such that the N-terminal ends of the gp41 MPER-TMD model were between the gp140 C-terminal ends. UCSF CHIMERA was used to model the now missing 15 residue stretch for every protomer separately. The final models were chosen visually to mitigate clashes between the now connected protomers and the modeled linkers. Figure 4.23A shows the final equilibrated construct in a simulation box, with a truncated and thus partially unfolded MPER, as well as a membrane embedded TMD in a DMPC bilayer.

The structure was simulated for a total of 4 μ s. The overall structure is stable in the first $\sim 0.9 \mu$ s, with the exception of the TMDs conformation. It initially starts in a conical configuration, adopts a transient tent-shaped configuration in the first hundreds of nanoseconds and then again rearranges into a conical configuration in the first microsecond of the simulation. The Env glycoprotein remains close to the membrane and undergoes only minor tilting from side to side, remaining completely upright for the most part (Figure 4.23A). After $\sim 0.9 \mu$ s all tryptophan residues in the MPER domain insert into the membrane leading to a major deformation (Figure 4.23B). The membrane forms a sinusoidal wave, with gp41 TMD anchored at its maximum. This deformation is not permanent and the membrane flattens after less than 100 ns. Multiple membrane deformation events were observed in the simulation of the HIV-1 Env construct in DMPC lipid bilayers.

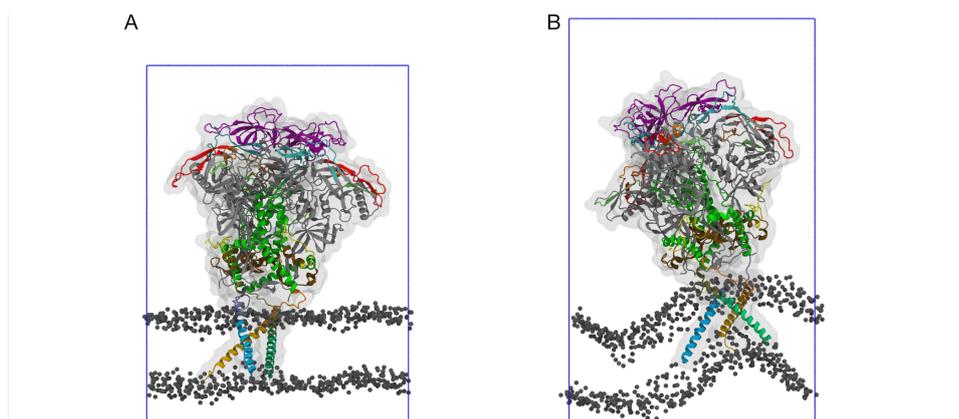


Figure 4.23: Snapshots of fully modeled HIV-1 Env construct connected to the membrane anchored TMD through the MPER domain, system #10. (A) Equilibrated model at the beginning of the MD simulation. (B) Inserted Trp residues (not pictured) cause a deformation of the membrane after hundreds of nanoseconds. Colors are as in Figures 4.6 and 4.20. A gray surface surrounds the protein. Water is not shown for clarity. The blue lines indicate the periodic boundaries of the simulation box.

4.3.11 Accessibility considerations of bnAbs to MPER

The short distance between the Env glycoprotein and the membrane raise the question how a bnAb can interact with its epitope when wedged between both the Env and the membrane. To understand what conformation the Env glycoprotein of HIV-1 has to adapt such that a bnAb can interact with its epitope, another attempt was made to build a model containing the bnAb, HIV-1 gp140 and gp41 MPER-TMD structures. First, a 4E10 bnAb (PDB ID: 4XAW) was superimposed onto its corresponding epitope in a gp41 MPER-TMD simulation, as described in a previous section. Comparing Figure 4.19 with Figure 4.23 highlights that the bnAb might not be able to reach its epitope when the Env glycoprotein is present in its upright position above the MPER-TMD. Thus, the gp140 model was again centered above the MPER-TMD and subsequently shifted by 10 Å in both x and y directions, as well as tilted by $\sim 45^\circ$. Next, residues E654–D659 in the gp140 model were removed. This allow for a proper connection between the HIV-1 models, while keeping the complete bnAb epitope in the gp41 MPER-TMD model intact. The two structures were connected by unstructured loops using MODELLER. A render of the full model embedded into a DMPC bilayer is shown in Figure 4.24.

The system was simulated for a total of 687 ns. Over the course of the simulation the bnAb does not maintain contact with its MPER epitope. This process is permanent and the bnAb does not regain contact with the epitope. Figure 4.25 shows the distances of putative hydrogen bonds and crystal contacts described in the publication by Cardoso *et al.* [266]. Two of the nine distances start at the reference values found in the crystal structure. The others are up to 12 Å apart. N671 is the only residue that regains contact for several nanoseconds at the end of the trajectory. All other contacts keep their distance and some go up to 18 Å.

Figure 4.25 highlights a problem with the initial contacts formed by the superimposed bnAb with its MPER epitope. The starting structure of gp41 MPER-TMD are arguably not the best fit, because they do not represent the distances observed in the crystal

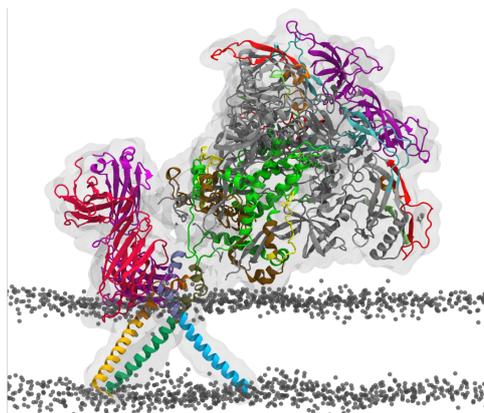


Figure 4.24: Snapshot of fully modeled HIV-1 Env construct connected to the membrane anchored TMD through the MPER domain with an 4E10 bnAb (PDB ID: 4XAW) bound to its MPER epitope, system #11. Colors are as in Figure 4.23. The two chains of the bnAb's Fab are shown in red (heavy chain) and magenta (light chain). A gray surface surrounds the protein. Water is not shown for clarity.

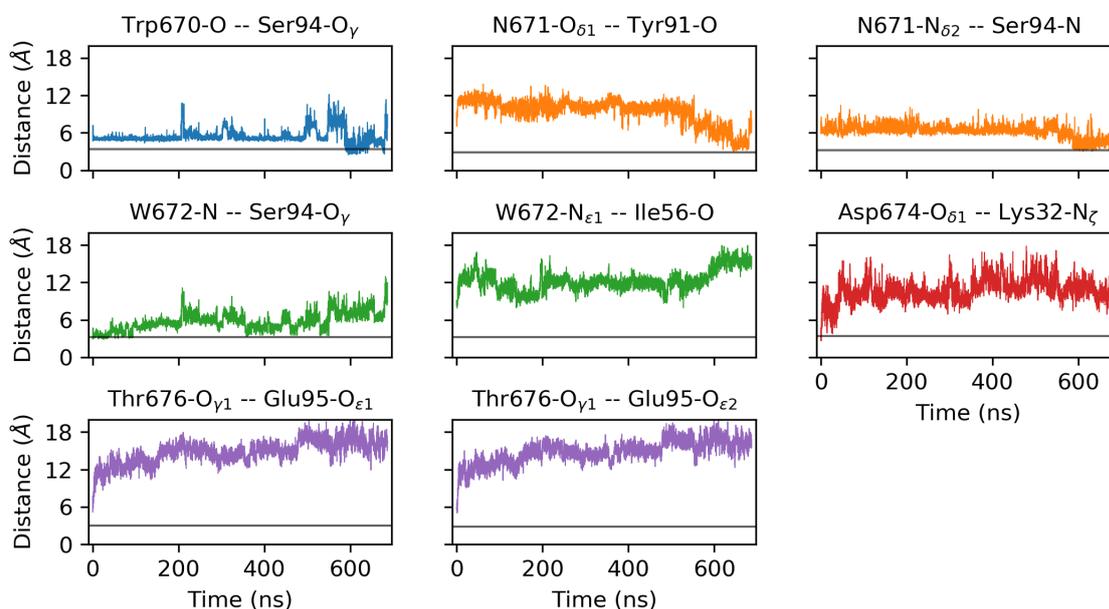


Figure 4.25: Distances of putative hydrogen bonds and crystal contacts between bnAb and its MPER epitope. Every panel shows a different distance between the MPER epitope (left label) and the bnAb (right label). The putative hydrogen bonds between the bnAb 4E10 and its MPER epitope were reported by Cardoso *et al.* [266]. Horizontal lines indicate the reference values as obtained by Cardoso *et al.* from the crystal structure.

structure. Nevertheless, the chosen gp41 MPER-TMD structure fit other model building constraints the best. First, it showed a reasonable RMSD to the co-crystallized epitope in the 4E10 crystal structure. Second, the N-terminal MPER α -helices do not clash with the superimposed bnAb structure. Third, the bnAb embeds properly into the membrane, as described by Irimia *et al.* [260]. Last, even though the bnAb stands almost upright with its Fab fragment, it was possible to fit and connect the HIV-1 gp140 model to MPER. In a future approach, one could try to run biased simulations to properly force the formation of the bnAb epitope such as to satisfy all putative hydrogen bonds and crystal contacts. In a subsequent simulation, one would expect the bnAb to maintain these hydrogen bonds and thus remain close to its epitope.

4.4 Discussion

The HIV-1 Env glycoprotein is a homotrimeric protein composed of different subunits. Its pre-fusion conformation resembles a mushroom-like shape, with a short stalk anchored in the membrane. The post-fusion conformation requires a major rearrangement of the gp41 fold found in the soluble part of Env. Two experimental studies report on the different oligomeric states of the gp41 TMD, being either monomeric or trimeric [236,237]. In the present work, the structural dynamics of different HIV-1 constructs were investigated using extensive atomistic molecular dynamics simulations.

The gp41 TMD was shown to be stable as a monomer, adopting two different conformations that are defined by the orientation of R696. In all simulations, the monomeric TMD tilts and allows R696 to snorkel towards the lipid-water interface. This tilted conformation is favored both by the charged R696 at the center of the membrane, and by the unusual length of the TMD, with almost 40 consecutive α -helical residues. The latter observation is consistent with Chiliveri *et al.*, who report a consecutive folded 32-residue α -helix [237]. These conformations are not only important for the TMD, but also determine the conformation of the N-terminally preceding MPER in monomeric gp41 MPER-TMD constructs. The TMD of gp41 is preceded by MPER at its N-terminus. Its secondary structure is α -helical and it is the only part of gp41 that is recognized by available antibodies [235,242], as shown by X-ray structures with the bnAbs 4E10 and 10E8 [260–262]. Intriguingly, the antibody-bound region is also part of the CRAC motif, which is important for interactions between gp41 and cholesterol [246]. MPER contains a high density of tryptophan residues that could embed into the membrane. Interestingly, simulations of MPER attached to the TMD reveal that the orientation of R696 locks the configuration of the TMD and in this way determines the specific orientation of MPER with respect to the membrane. When R696 snorkels towards the exofacial leaflet, MPER acts as an extension to the TMD and is able to embed into the upper leaflet of the membrane. Instead, when R696 snorkels to the cytofacial leaflet, the hydrophobic side of MPER faces away from the membrane. No unfolding of the MPER-TMD junction is observed in the simulations, whereas Kwon *et al.* have solved a MPER-TMD trimer with an unstructured turn spanning the CRAC motif (W678-K683) [238].

The results in this chapter have further shown that the proposed trimeric TMD assembly

(PDB ID: 5JYN) is unstable in a viral-like membrane, adopting a conic conformation after hundreds of nanoseconds. Dev *et al.* proposed that all three R696 are perpendicular to the membrane normal in the trimeric gp41 TMD structure [236, 268]. In contrast, the simulations show that these residues snorkel towards the lipid-water interface, irrespective of the oligomeric state. Dev *et al.* also proposed that the charged guanidinium group of R696 is shielded inside a hydrophobic pocket formed by L692, L695, and I697 of the neighboring protomer. However, experimental observations by the same group show that the C_α atom of R696 is located at the center of the membrane, while the polar side chain is accessible for water-hydrogen exchange [236, 268]. These two observations are consistent with the simulations reported in this chapter. R696 C_α is located roughly at the membrane center, whereas R696 C_ζ snorkels to the lipid-water interface. Chiliveri *et al.* speculate that the unusually long α -helix tilts in the membrane in response to the TMDs hydrophobic mismatch. The trimeric structure of Dev *et al.* show an untilted helical bundle with respect to the membrane normal. Simulations in this chapter revealed that the monomeric gp41 TMD adopts a tilt angle of $\sim 30^\circ$ in a few hundred nanoseconds, while the protomers in the trimeric conformation take $3\ \mu\text{s}$ to equilibrate and adopt a similar tilt angle. The tilt angles reported for the gp41 TMD trimer by Oakes *et al.* [242] are consistent with the pre-equilibrated values from gp41 TMD simulation during the first $3\ \mu\text{s}$ in viral-like bilayers, but their short simulations do not capture the equilibrated tilt angles.

The trimeric gp41 TMD structure is not stable in MD simulations performed in a viral-like lipid bilayer. The trimeric conformation quickly disassembles, and each protomer partially unfolds. This is consistent with Hollingsworth *et al.* [243], who report similar instabilities despite using position restraints throughout their atomistic molecular dynamics simulations. The proposed trimeric structure of gp41 TMD [236] could be stabilized by a specific protein:lipid ratio, or a specific lipid bilayer composition. The uncertainty about the oligomeric state of gp41 TMD is further complicated by the difficulty to use *in vivo*-like membrane structures and compositions in experimental studies, which often rely on the use of bicelles. These, in combination with varying protein:lipid ratios, can potentially cause differences in observations with the herein reported MD simulations from this chapter. Furthermore, it is possible that the gp41 TMD is found in a variety of different oligomeric states during its life cycle. The unusual length of the TMD might play an important role in facilitating the transition along the secretory pathway, overcoming the effect of the conserved charge R696, which is buried deep in the transmembrane region of the membrane and would otherwise block the exclusion of the TMD from the endoplasmic reticulum [138]. Kwon *et al.* summarize three possible oligomeric conformations of gp41 MPER-TMD [238]: (i) trimeric TMD with monomeric MPER bound to the membrane surface [238], (ii) trimeric TMD with trimeric MPER [236, 237, 239] and (iii) monomeric TMD with trimeric MPER [269]. All of these may represent a conformation that gp41 undergoes during the viral life cycle, enabling the elaborate evasion of host defense mechanisms.

The possible placements and conformations of the trimeric gp41 MPER-TMD in a model DMPC bilayer were studied, to assess the effects of a thin hydrophobic membrane core on the unusually long TMD. In all six simulations the trimeric TMD structures

rearranged into conical or tent-shaped trimers. The structures fell apart quicker than in the viral-like bilayers. This is to be expected, as the latter are more rigid due to the inclusion of a large percentage of cholesterol and sphingolipids. Intriguingly, the MPER α -helices recovered the epitopes for bnAb. The superimposition of the 4E10 and 10E8 bnAbs onto their corresponding epitopes revealed the striking resemblance. These observations led to the question, if and how bnAbs are able to access their epitopes in the presence of the N-terminal Env glycoprotein in the pre-fusion state. A model of the soluble Env domain was built using a crystal structure of HIV-1 gp140, showing the stability of the structure in solution. Afterwards, this structure was connected with a previously simulated gp41 MPER-TMD model. These simulations revealed that the distance between the soluble Env glycoprotein and the MPER domain at the lipid-water interface is not sufficient to allow a bnAb to interact with its epitope. These observations are supported by a recent study by Li *et al.* [270] where subtomogram averaging reveals a 12 Å gap between the lipid-water interface and the gp120 glycoprotein. To overcome this limitation, a final model was built with a 4E10 bnAb docked onto its corresponding MPER epitope with a soluble, titled HIV-1 Env glycoprotein connected to the gp41 MPER-TMD construct. Unfortunately, the contacts formed by the crystal structure of 4E10 with its co-crystallized epitope were not fully recovered in the initial set up of this system, as well as the MD simulation. The modeling of this complete structure required the fulfillment of various constraints, making it hard to find a suitable structure that also satisfies the putative hydrogen bonds and crystal contacts in the crystal structure [266].

The undulations observed in the simulations of the almost full-length HIV-1 gp160 construct might be a simulation artifact. The thin DMPC bilayers are easier to distort than, i.e., the viral-like membrane containing both sphingolipids and cholesterol that make it more stiff. As no experimental studies discuss deformed membranes on HIV-1 virions it is hard to argue that the observed effect is not an artifact.

The simulations of HIV-1 constructs in this chapter highlight different aspects of the viral Env glycoprotein. The conserved residue R696 might carry out an important functional role, given that all simulations show R696 to cause local membrane defects. These defects could facilitate fusion between viral and target cell membranes. A recent molecular dynamics study of the gp41 TMD trimer reported that water penetrates into the membrane core [243]. Two metadynamics studies also reported water penetration due to snorkeling of R696 [271, 272]. Different configurations of R696 could also act as a switch to stabilize alternative structural conformations of gp41. These conformations might provide a different interface for the four arginine residues R683, R696, R707 and R709, but also control the conformation of the MPER and its interaction with the membrane. The recognition of MPER by bnAbs might be affected by both the oligomeric state of the TMD, as well as its conformation. Furthermore, a variety of glycans cover the surface of the Env glycoprotein and shield it from the immune system [273–276]. No glycans were included in the models built in this chapter, but they might further complicate the accessibility of the MPER epitope by the respective bnAb.

Optimization of molecular dynamics simulation performance

This chapter describes the efforts to develop a software package that optimizes the resource usage of supercomputers. It contains computational work that was performed at the Max Planck Institute of Biophysics, Frankfurt am Main, Germany. Dr. Jürgen Köfinger derived the equations for Amdahl's law. This work has been published in the article: M. Gecht, M. Siggel, M. Linke, G. Hummer, J. Köfinger. MDBenchmark: A toolkit to optimize the performance of molecular dynamics simulations. *J Chem Phys*, 153(14):144105, 2020. This chapter contains discussion with Dr. Max Linke, Dr. Sebastian Kehl, Dr. Markus Rampp and Dr. Klaus Reuter. All simulations, data analysis and interpretations were carried out by myself under the supervision of Dr. Jürgen Köfinger (Department of Theoretical Biophysics, Max Planck Institute of Biophysics, Frankfurt) and Prof. Dr. Gerhard Hummer (Department of Theoretical Biophysics, Max Planck Institute of Biophysics, Frankfurt).

5.1 Introduction

Usage of MD simulations has grown over the past decade and as a result its users are likely unfamiliar with the implementation details of the underlying software. Various software packages, so called MD engines, have emerged to orchestrate the calculation of MD simulations, e.g., ACEMD [277], Amber [278], CHARMM [279], Desmond [280], GROMACS [165], LAMMPS [281], NAMD [282] and OpenMM [283]. These software were designed to be efficient at computing the time evolution of particles in a simulation box, crunching numbers for millions of consecutive steps. Supercomputers are commonly used to perform the expensive computations required by MD simulations. To run a computational job, the user can specify numerous parameters in different permutations to control the behavior of underlying hardware and software. A common desire is to optimize the time-to-solution (TTS). Here, the solution is an extensive atomistic simulation composed of 10^4 atoms to 10^7 atoms and dozens to hundreds of microseconds simulation time.

Users new to MD simulations might miss the details that are required to run efficient simulations on supercomputers. The wrong choice of parameters can notably degrade the simulation performance, measured in ns/day, or increase it only by a small margin. A user might decide to request additional nodes for a simulation and gain a net increase in simulation performance. While this optimizes TTS locally and is a short time gain for the user, it diminishes global cluster-wide efficiency. This behavior will, in the long run, lead to less simulation throughput, a higher electricity demand and cost, and ultimately to a worse carbon footprint. It is therefore in the interest of every supercomputer user to optimize

their usage of hardware resources, and, at the same time, keep their TTS to a minimum.

Previously, Kutzner *et al.* have determined the performance-to-price (P/P) ratio of many central processing unit (CPU) and graphics processing unit (GPU) architectures and simultaneously scanned the number of MPI ranks, OpenMP threads and separate PME ranks with GROMACS 4.6 [284]. In their follow up study, the authors use GROMACS 2018 with more recent hardware and show that the previous observations also hold true when considering the costs for energy and cooling [285]. Overall, they conclude that a single node equipped only with CPUs or at most one GPU yields the best P/P ratio. Additionally, four to eight CPU cores are sufficient to perform optimal computations with Nvidia GeForce GTX 1080 Ti or more recent GTX 2080 cards. While Kutzner *et al.* are, for the most part, interested in running many “short” simulations to perform exhaustive sampling of phase space, many groups are interested in accumulation of considerable simulation time for individual MD simulations with a low TTS.

Aim. Here, multi-node benchmarks with varying numbers of MPI ranks and OpenMP threads were performed to determine the optimal performance. To ease this process, a standalone software package, implemented as command-line interface (CLI), to conveniently set up, run and analyze benchmarks of MD simulations was developed. MDBenchmark was created to streamline and ease the process of finding the optimal run parameters and settings for any simulation and hardware stack. It is capable of submitting a simulation to all common queuing engines, scale it over a different number of nodes, toggle the usage of CPUs and/or GPUs, and also tune the ratio of MPI ranks over OpenMP threads. To showcase the functionality, an exhaustive scaling study of 23 MD simulations of varying sizes (4×10^4 atoms to 4×10^6 atoms) and system compositions is reported. Certain ratios of message passing interface (MPI) ranks to open multi-processing (OpenMP) threads are identified that produces the best performance for a range of system sizes.

5.2 Methods

5.2.1 Additional background on simulations

Modern compute clusters are made up of compute nodes. Each node usually contains one CPU and can optionally contain GPUs as well as gigabytes of dedicated random-access memory (RAM). The nodes are connected through a network to exchange data. This exchange allows the nodes to distribute calculations and compute them in parallel on multiple nodes at the same time. Modern CPUs are made up of dozens of physical cores. Each of these cores can perform computations independent from the other cores. A single physical core can often perform two computations at the same time, a feature called “hyperthreading” [286]. When it is enabled, the number of physical cores is virtually doubled. For example, for each physical core two “logical cores” are introduced. Two different interfaces MPI and OpenMP are implemented to use these resources efficiently: MPI spawns processes, which will be referred to as ranks in the following. Each MPI rank can be composed of all cores in a single node or only a subset of the cores. In contrast, computational threads are created by OpenMP. These threads are made up of a subset of

cores that are allocated to a single MPI rank. It is important to note that OpenMP threads share the same memory, whereas MPI ranks do not. The numbers of MPI ranks to OpenMP threads have an influence on the performance of GROMACS simulations [284, 285].

To run a simulation on a computer cluster, the user must first submit a computational job to its queuing system. To do this, the user has to configure a submission script to launches the MD engine. In this submission script the correct parameters for the specific queuing system have to be defined by the user manually. Before starting a simulation, the simulation box is divided into separate domains in a process called domain decomposition to run on parallel architecture [287]. Each domain is isolated from the others and only information at the borders of every domain is communicated with the others. A single MPI rank is in charge of the calculations for a single domain. This rank then also handles the communication between its neighboring domains (and therefore ranks). Also, to distribute the load inside a domain each MPI rank spawns multiple OpenMP threads. The product of the number of ranks n_{ranks} and the number of threads n_{threads} must equal the number of logical cores available in the compute node. When hyperthreading is deactivated the term “logical cores” is identical to the physical cores. When hyperthreading is activated, the number of physical cores are doubled to obtain the logical cores.

5.2.2 General procedure

GROMACS 2018.8 was used to run all simulations reported in this chapter [165]. The numbers of MPI ranks, n_{ranks} , and OpenMP threads, n_{threads} , were varied to determine the optimal performance $P(N)$ for a given number of nodes N . On CPU-only nodes GROMACS was allowed to dedicate around 25 % of available MPI ranks for PME calculations. On mixed CPU-GPU nodes no individual PME ranks were specified. GROMACS 2018 was compiled with external MPI libraries and OpenMP support, with AVX_512 SIMD instructions enabled at compile time using GCC 8.3 and CUDA 10.1. Nodes with two Intel[®] Xeon[®] Gold 6148 CPUs (2.40 GHz) were used for all benchmarks. Mixed CPU-GPU nodes were equipped with two NVIDIA Quadro RTX 5000 graphic cards each. All nodes were connected with a 100 Gb/s OmniPath interconnect. Benchmarks were run on specific numbers of MPI ranks and OpenMP threads for a total wall time of 15 minutes, if not specified otherwise. SLURM was used as the queuing system for all benchmarks on the available MPCDF “cobra” supercomputer. Example submission files for both CPU-only and mixed CPU-GPU nodes can be found in the appendix.

5.2.3 Systems description

The setup of each system can be found in the corresponding reference, if it was already published. In general, the systems were set up with a 1.2 nm cut-off for nonbonded interactions in the real space and a grid spacing of 0.12 nm for PME. An exception is formed by the systems #6, #12, #16, #20, #21 where a real space cut-off of 1.0 nm was used with a grid spacing of 0.16 nm. Dynamic load balancing was used in all benchmarks to tune the cut-off for nonbonded interactions and PME grid spacing automatically. Therefore, the listed values serve only as lower and upper bounds. In addition, the systems #0 to #21

were run with an integration time step of 2 fs, whereas system #22 used 4 fs and doubled the mass of all hydrogens [288].

5.2.4 Fitting Amdahl’s law to performance data

Amdahl’s law was used as a simple model to fit the performance of the herein presented scaling studies [289]. The law captures the speed-up $S(N)$ of parallel computations as a function of the number of nodes N using the following formula:

$$S(N) = \frac{1}{1 - p + \frac{p}{N}} \quad (5.1)$$

where p is the part of the code that benefits from the parallelization, as represented by a fraction.

Amdahl’s law can be used to express the performance $P(N) = S(N) \times P(1)$ as

$$P(N) = \frac{P^{\max}}{1 + \frac{p}{N(1-p)}} \quad (5.2)$$

where the maximal possible performance $P^{\max} = \lim_{N \rightarrow \infty} P(N)$ is obtained by

$$P^{\max} = \frac{P(1)}{1 - p} \quad (5.3)$$

The speed-up that is defined by the performance ratio of going from $N = 0$ to $N = 1$, is called “ideal scaling” and is given by:

$$P^{\text{id}}(N) = NP(1) \quad (5.4)$$

Equation (5.4) can be expressed as the following:

$$P^{\text{id}}(N) = N(1 - p)P^{\max}. \quad (5.5)$$

5.3 Results

In this section, the results from extensive MD simulation benchmarks are presented and discussed. At the beginning the software package MDBenchmark is introduced. The software is then used to showcase its capabilities to benchmark MD simulations and enhance the performance by varying certain parameters. First, the optimal run time for a benchmark is defined. Afterwards, the optimal choice of MPI ranks and OpenMP threads is estimated at first for two exemplary systems. Amdahl’s law is used to fit the simulation performance data to estimate the maximum performance. The previous observations are then tested for all 23 systems of varying sizes. At the end of the section, the option of running multiple simulations on a single CPU-GPU node is explored.

5.3.1 Implementation details of MDBenchmark

Four main functions can be accessed with the MDBenchmark CLI (Figure 5.1). Using the `GENERATE` command all benchmark parameters for the benchmark(s) can be defined by the user (Figure 5.1A). A input file of an equilibrated MD simulation has to be provided, for example a `.TPR` file for GROMACS. The user then defines the MD engine and version to use, onto how many nodes to scale the simulations, whether MDBenchmark should choose CPU-only and/or mixed CPU-GPU nodes and the number of MPI ranks. MDBenchmark will automatically check if the requested MD engine is available using the “Environment Modules” system [290]. This feature safeguards the user against typos that can lead to wasted computational time. This availability check of MD engines can be skipped by the user. Before proceeding with any write operations, MDBenchmark will prompt the user to confirm the requested action.

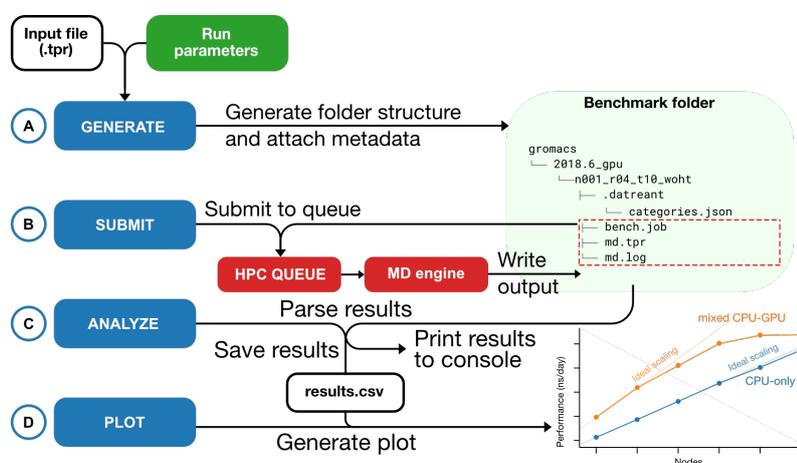


Figure 5.1: Workflow and implementation of MDBenchmark. (A) The `GENERATE` command is provided with run input files. An own folder is generated for every benchmark with its own metadata. (B) The benchmark job is submitted to the supercomputer queue using the `SUBMIT` command. (C) The engine writes a log file to the benchmark folder. Its contents can be parsed with the `ANALYZE` command. The results of each benchmark can be printed into the console or saved as a CSV file. (D) Using this CSV file the user can generate a plot with the `PLOT` command to visualize the results. The benchmark results are plotted as scaling plot with the performance (ns/day) as a function of the number of nodes, with every benchmark setting as a separate line. Adapted from Gecht *et al.* (2020) [291].

MDBenchmark generates a folder structure with a nested hierarchy that allows the user to browse the folders and explore files themselves (Figure 5.1A). At the top of this structure, every MD engine gets its own folder, with the exact version as a subfolder. Here, the usage of mixed CPU-GPU nodes are denoted with the “_gpu” suffix. If the suffix is missing, then CPU-only nodes are used. At the last layer the separate benchmark folders can be found. The folder name defines the number of nodes that will be used in this benchmark, as well as the number of MPI ranks, OpenMP threads and whether hyperthreading will be enabled or disabled. The run input file that was defined by the user will be copied into this folder, as well as a job submission script “bench.job” that contains all settings for the queuing system. A hidden folder generated by the datreant Python package contains all metadata in JSON format [292].

Benchmarks can be submitted at once using the `SUBMIT` command (Figure 5.1B).

MDBenchmark will gather information on all eligible benchmarks and show a confirmation prompt to the user. Only after confirming the prompt, will MDBenchmark submit all the benchmarks to the queuing system. All benchmarks that were already submitted are omitted from the submission, but can be manually added by the user. At the time of writing MDBenchmark can submit benchmarks to the queuing systems SGE, SLURM and IBM's LoadLeveler.

The status of currently submitted benchmarks can be checked with the `ANALYZE` command (Figure 5.1C). Upon calling, MDBenchmark will gather the latest data and print it to the console. Finished benchmarks will be annotated with their performance. At this stage the user is able to save the benchmark results to a CSV file for future analysis. Finally, the user can generate a plot from this CSV file using the `PLOT` command (Figure 5.1D). The resulting scaling plot shows the performance (ns/day) as a function of the number of nodes, with each line representing a different combination of parameters: CPU-only/mixed CPU-GPU nodes, the numbers of MPI ranks and OpenMP threads as well as the status of hyperthreading. A linear fit is plotted through the origin and the first data point to symbolize the ideal scaling. Multiple CSV files can be supplied to the `PLOT` to compare benchmarks from different systems, settings or supercomputers. The MDBenchmark software package is accompanied by an extensive documentation found at <https://mdbenchmark.readthedocs.io/en/>. The documentation contains an explanation on all available options and a guide to install the package and how to generate, submit and analyze benchmarks.

5.3.2 Biomolecular systems are mostly composed of solvent

MDBenchmark was used to run scaling benchmarks of 23 atomistic biomolecular systems with system sizes ranging from 4×10^4 to 4×10^6 atoms (Figure 5.2A). 65% of tested systems (15/23) have less than 300 k atoms. As is the case for most biomolecular simulations, the herein studied 23 simulation systems are for the most part composed of solvent: water molecules make up 55% to 99% of the total atoms (Figure 5.2B). Lipid bilayers, if present, account for 8% to 45% of the remaining atoms. All other solutes, for example proteins, ions, nucleic acids and small molecules, account for only 0.1% to 27%. The 23 systems use a broad range of force fields, compositions, box geometries, water models and sizes as listed in Table 5.1.

5.3.3 Defining a minimal run time of benchmarks

An ideal benchmark run should provide an accurate estimate of performance (ns/day) in the least amount of simulation time. A high throughput can be achieved with shorter benchmarks, allowing to sample more parameters and use the limited computer resources more efficiently. Benchmarks can be run with a fixed number of steps or a fixed amount of time. It is easier to define the latter, because this allows the queuing system to prioritize the benchmark jobs without user intervention. When using a fixed amount of steps, the user must still specify an estimated run time for the queuing system. Here, the user would be interested in specifying the shortest time possible, because that would give the benchmark

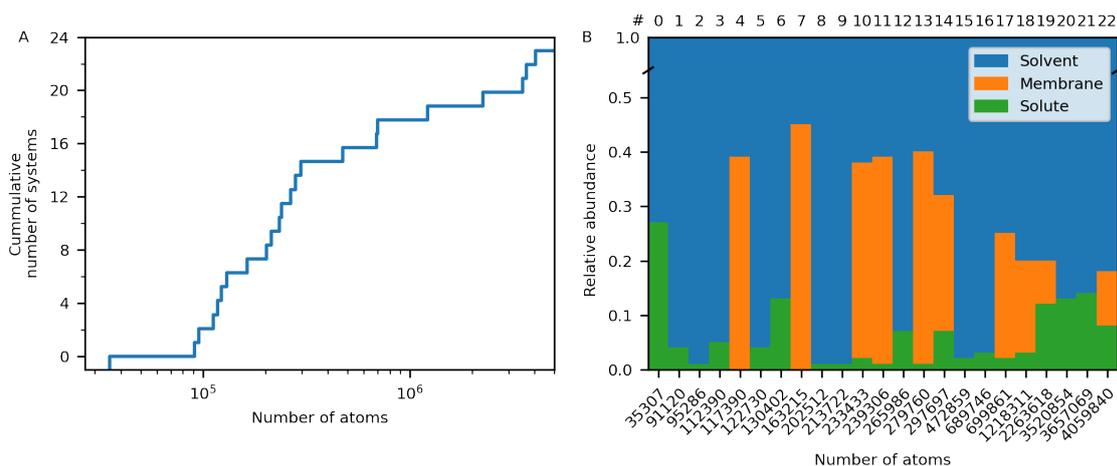


Figure 5.2: Size and composition of MD simulation systems. (A) Cumulative number of systems as a function of the number of atoms. Most MD systems (65 %) contain less than 300 k atoms. (B) Relative abundance of system constituents as a function of the atom number. Systems are composed of 55 % to 99 % solvent (blue). Lipid bilayers, if present, make up 8 % to 45 % (orange) of the systems total number of atoms. Other solutes make up at most 27 % but generally less than 10 % (green). Adapted from Gecht *et al.* (2020) [291].

run higher priority in the queue. To test the minimum time for a benchmark job, two differently sized systems (#10 and #21) were scaled from 1–10 nodes and run for 5, 10 and 15 minutes each (Figure 5.3). The shortest run times of 5 minutes provide a fair estimate of the possible performance, but it tends to overestimate the performance on higher node numbers with increased uncertainty. In contrast, run times of 15 minutes provide a more certain estimate for the two systems independent of their size. In the following sections all benchmarks were run for 15 minutes each independent of their system size.

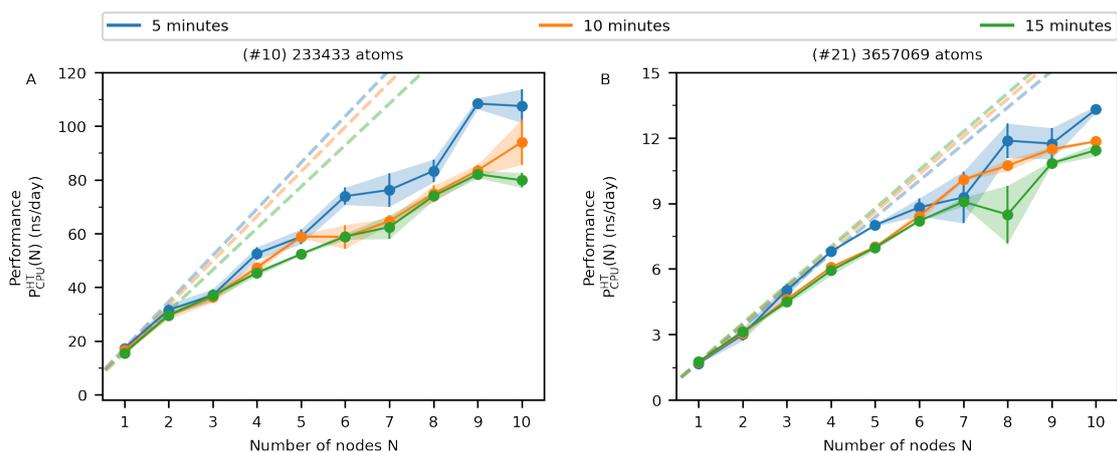


Figure 5.3: Performance estimated for different run times. MD simulations with (A) 233 k and (B) 3.6 M atoms were scaled from 1 to 10 nodes and run for 5 (blue), 10 (orange) and 15 (green) minutes. The average of 5 independent runs with error bars and filled-in area showing the standard deviation is shown for every data point. Transparent dashed lines show ideal scaling according to Equation (5.4). Adapted from Gecht *et al.* (2020) [291].

5.3.4 Optimizing performance of two systems

The number of optimal number of MPI ranks and OpenMP threads depend on the system size and hardware architecture. To better understand the scaling behavior of MD simulations across different node numbers and cluster settings, the two systems from the previous section (#10 and #21) were scaled from 1–10 nodes. Simultaneously different numbers of MPI ranks and OpenMP threads were scanned, while enabling or disabling hyperthreading to check its influence on simulation performance. The effect of hyperthreading was quantified for both CPU-only and mixed CPU-GPU nodes using a ratio of the benchmark performance with ($P_x(N)$) and without ($P_x^{\text{HT}}(N)$) hyperthreading:

$$\Delta P_x(N) = P_x(N)/P_x^{\text{HT}}(N) \quad (5.6)$$

In the following, the subscripts $x = \text{CPU}$ and $x = \text{GPU}$ denote CPU-only and mixed CPU-GPU nodes, respectively. The superscript HT symbolizes that hyperthreading is enabled. If the superscript HT is omitted, hyperthreading is disabled. Simulations run with CPU-only nodes achieve the best performance with $n_{\text{ranks}} = 40$ independent of the system size (Figure 5.4).

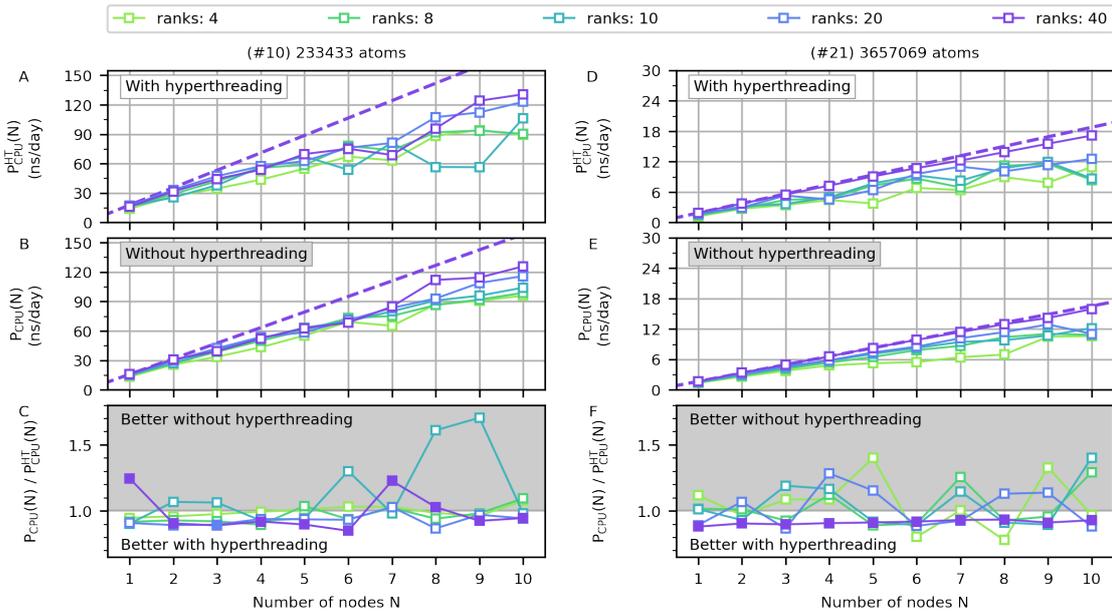


Figure 5.4: Scaling of two systems on CPU-only nodes. Performance $P_{\text{CPU}}(N)$ for systems scaled on N nodes using CPU-only nodes and different numbers of MPI ranks n_{ranks} (colors) with (A and D) and without (B and E) hyperthreading. The ideal scaling was estimated with Equation (5.4) (dashed lines; strongest ideal scaling as thick dashed line). (A and B) 233 k atoms membrane protein system. (D and E) 3.6 M atoms dense protein solution. Both systems scale the best with $n_{\text{ranks}} = 40$. (C and F) Ratios between performances without and with hyperthreading as per Equation (5.6). Both systems generally benefit from running simulations with hyperthreading for the best performing rank settings $n_{\text{ranks}} = 40$ (filled squares). Adapted from Gecht *et al.* (2020) [291].

Choosing less MPI ranks can degrade the performance 2-fold for certain number of nodes. Enabling hyperthreading leads to a slight increase of simulation performance with optimal settings, whereas with other numbers of MPI ranks the benefits are less pronounced

(Figure 5.4C and F). Independent of the settings and the system size, simulations on CPU-only nodes scale well and perform almost as good as the linear optimum, estimated through $N = 0$ and $N = 1$ (Figure 5.4).

The optimal settings on mixed CPU-GPU nodes are different than on CPU-only nodes. The same benchmarks were repeated on mixed CPU-GPU nodes, scaling the two systems from 1–10 nodes and scanning different MPI rank and OpenMP thread combinations. $n_{\text{ranks}} = 40$ remains to be the optimal choice for the 3.6 M atom system on mixed CPU-GPU nodes (Figure 5.5D and E). In contrast, the smaller system with 233k atoms shows a different dependence on the MPI ranks. Here, the highest performance is achieved for $n_{\text{ranks}} = 10$, whereas the best setting obtained for CPU-only nodes $n_{\text{ranks}} = 40$ results in a highly degraded performance (Figure 5.5A and B). These results point to the conclusion that smaller systems require less MPI ranks when run on mixed CPU-GPU nodes. In addition, hyperthreading does not increase the performance on mixed CPU-GPU nodes and should thus not be used. When comparing the benchmarks run on mixed CPU-GPU nodes with those on CPU-only nodes it becomes apparent that the former scale worse. As noted earlier, simulations on CPU-only nodes scale almost linearly up to $N = 10$, whereas on mixed CPU-GPU nodes the performance starts to degrade considerably for the 233k atom system on $N > 3$, but less so on the 3.6 M atoms system.

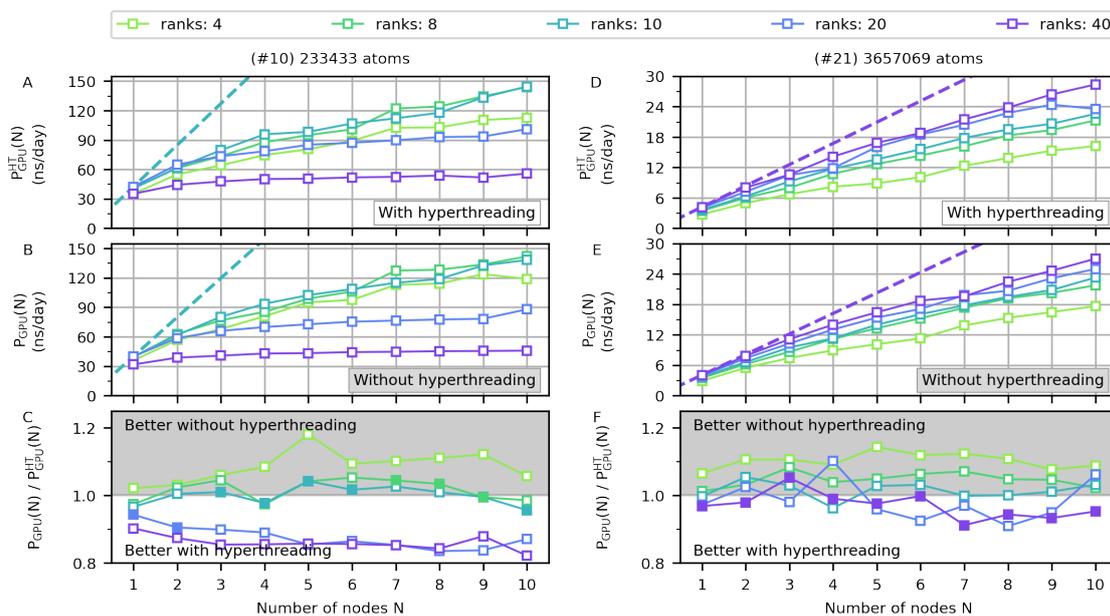


Figure 5.5: Scaling of two systems on mixed CPU-GPU nodes. Performance $P_{\text{GPU}}(N)$ for systems scaled on N nodes using mixed CPU-GPU nodes and different numbers of MPI ranks n_{ranks} (colors) with (A and D) and without (B and E) hyperthreading. The ideal scaling was estimated with Equation (5.4) (dashed lines; strongest ideal scaling as thick dashed line). (A and B) 233k atoms membrane protein system. (D and E) 3.6 M atoms dense protein solution. $n_{\text{ranks}} = 40$ is the optimal setting for the 3.6 M atom system. (C and F) Ratios between performances without and with hyperthreading as per Equation (5.6). The smaller 233k atoms system does not benefit from hyperthreading. Adapted from Gecht *et al.* (2020) [291].

5.3.5 Performance estimates from Amdahl's law

The scaling behavior of simulations can be captured by Amdahl's law. As noted earlier, the performance of simulations first follows a linear increase for small node numbers, but starts to level off with higher node numbers. This is especially visible for simulations with a system size of 233 k atoms. Using Equation (5.2) it is possible to fit the observed data from benchmarks and estimate the maximal performance that can be achieved with a given system. This helps to better understand if it is feasible to add more nodes to a simulation, or whether this would not significantly increase the performance wasting more resources. As a proof of concept the two systems were scaled up to 64 nodes using the optimal settings as described in the previous section, using both CPU-only and mixed CPU-GPU nodes (Figure 5.6).

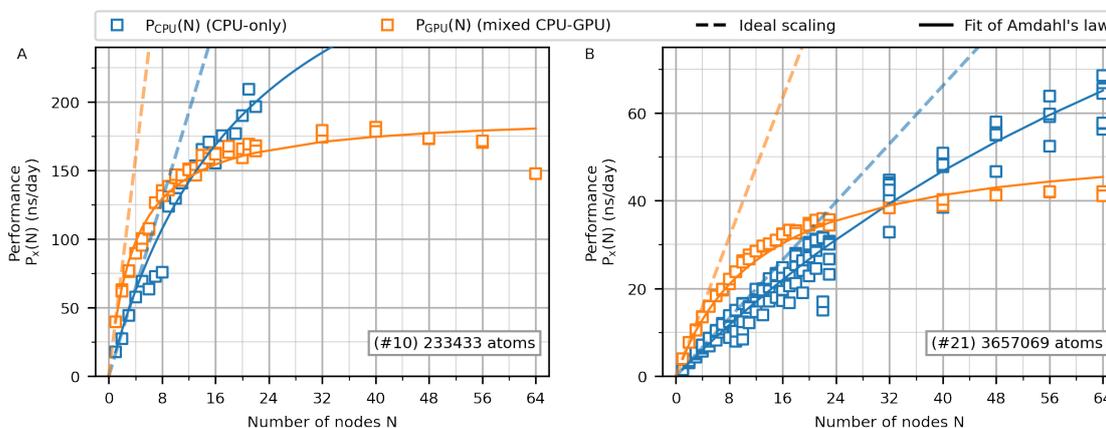


Figure 5.6: Amdahl's law captures the scaling behavior. Performance $P_x(N)$ of two MD simulations with varying sizes as a function of nodes N . Each panel shows the best performing n_{ranks} values on mixed CPU-GPU nodes without hyperthreading (orange) and CPU-only nodes with 40 n_{ranks} and hyperthreading (blue) from Figures 5.4 and 5.5. Each combination of nodes N was run in multiple independent simulations of 15 minutes each and all replicates are shown. Solid colored line shows the fits to the data using Equation (5.2). Transparent dashed lines show ideal scaling according to Equation (5.4). Adapted from Gecht *et al.* (2020) [291].

Bigger systems scale better than smaller ones. On a small number of nodes, $N < 14$, mixed CPU-GPU nodes perform considerably better than CPU-only nodes for the 233 k system (Figure 5.6A). At this low node count the system already deviates considerably from linear scaling. In contrast, for CPU-only nodes the system scales almost linearly and deviates much slower from this linear optimum. At bigger node numbers ($N > 14$) CPU-only nodes achieve a higher performance than is possible with mixed CPU-GPU nodes (Figure 5.6A). A similar observation can be made for the big system with 3.6 M atoms. Simulations running on mixed CPU-GPU nodes achieve a higher performance on low node numbers ($N < 32$), but start to level off earlier providing less net gain for every node added (Figure 5.6B). Again, simulations run on CPU-only nodes achieve higher absolute performances and scale closer to the linear optimum (Figure 5.6B)

5.3.6 Optimal parameters depend on system size

The optimal number of MPI ranks strictly depend on the system size for mixed CPU-GPU nodes. In the previous sections it has been shown that the optimal parameters for two systems of different sizes are the same on CPU-only nodes, but have different results for mixed CPU-GPU nodes. To further test this observation more benchmarks were performed. 21 additional systems of sizes ranging from 4×10^4 atoms to 4×10^6 atoms were scaled from 1–10 nodes, while simultaneously the number of MPI ranks were varied on mixed CPU-GPU nodes. The systems were scaled from 1–10 nodes on CPU-only using the optimal settings of $n_{\text{ranks}} = 40$. Similar to the previous results, smaller systems require less MPI ranks for optimal performance (Figure 5.7).

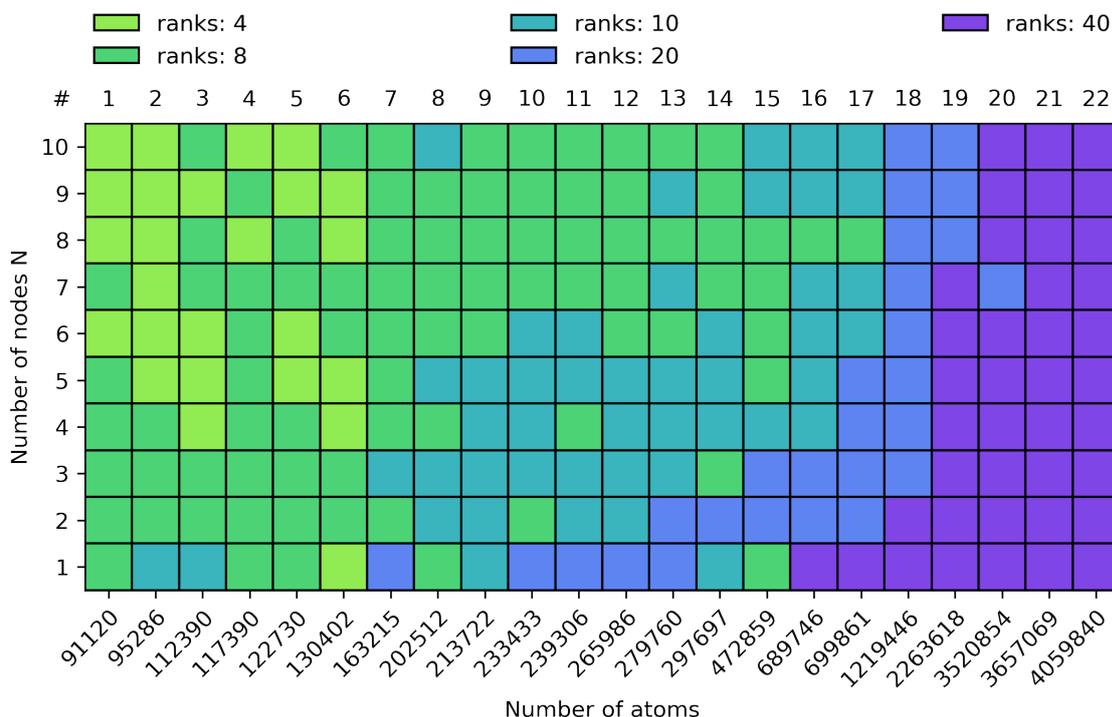


Figure 5.7: The same systems (columns) were scaled over different numbers of nodes (rows), using different n_{ranks} values. Optimal number of MPI ranks depends on system size. The optimal number of MPI ranks (colors) decreases with increasing node numbers N and increases with system size. The top axis indicates the system numbers according to Table 5.1. Adapted from Gecht *et al.* (2020) [291].

The number of MPI ranks decreases with the number of nodes while it grows with the system size. The data for all MPI rank combinations on mixed CPU-GPU nodes, as well as for $n_{\text{ranks}} = 40$ for CPU-only nodes are shown in Figure 5.10. For some combinations of nodes, for example $N = 7$ or $N = 9$, an unfavorable domain decomposition leads to a degraded performance. This behavior showcases the requirement to benchmark all systems, even though the herein presented plots provide a general guidance. Mixed CPU-GPU nodes perform better than CPU-only nodes on a low number of nodes. Comparing the absolute performances for the optimal number of MPI ranks using either mixed CPU-GPU nodes or CPU-only nodes, it becomes clear that almost all systems double their performance on a single node when using CPU-only nodes, with the only exceptions being systems

#6 and #12 (Figure 5.8). When running a simulation on two nodes all systems with more than 233 k atoms increase their performance by a factor of 2. Here, the two systems #12 and #16 are the exception. It is important to note that the cut-off distance in the real space interactions systems has been reduced for systems #6, #12, #16, #20 and #21, while the grid spacing for PME calculations was increased. These settings lead to an improved performance on CPU-only nodes, but do not change that of mixed CPU-GPU nodes. Doubling N on CPU-only nodes also doubles performance, but this does not affect mixed CPU-GPU nodes.

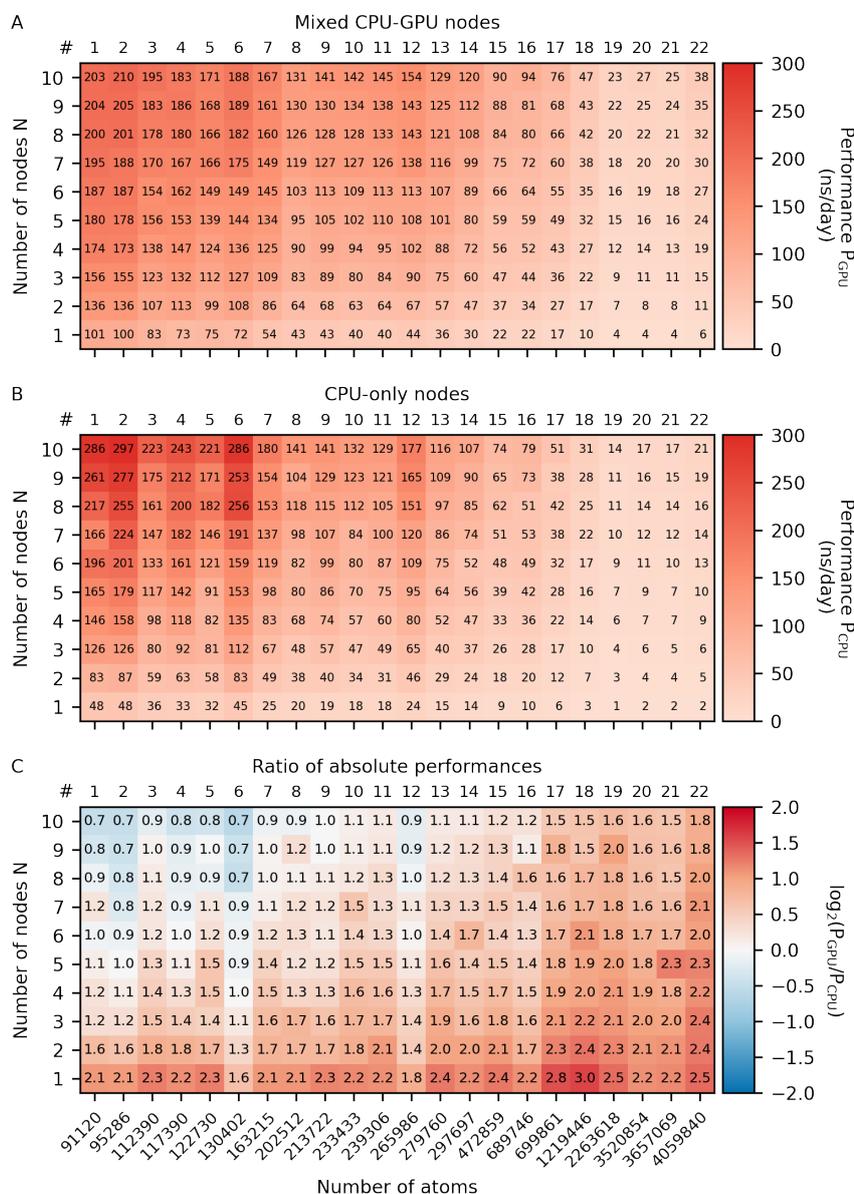


Figure 5.8: Performance comparison between different architectures. Absolute performances for 22 MD simulations scaled from 1–10 nodes using (A) mixed CPU-GPU nodes and (B) CPU-only nodes. Numbers in the fields represent the absolute performance values in ns/day. (C) The performance ratio $P_{\text{GPU}}(N)/P_{\text{CPU}}(N)$ is shown on a logarithmic (base 2) color scale. The color scale is centered at equal performance ($P_{\text{GPU}}(N)/P_{\text{CPU}}(N) = 1$, white). The numbers in the cells are the actual values of the ratios. Performance data was taken for the best performing n_{ranks} values on mixed CPU-GPU nodes without hyperthreading (Figure 5.7) for each node individually. For CPU-only nodes, $n_{\text{ranks}} = 40$ and hyperthreading was activated. The top axis indicates the system numbers according to Table 5.1. Adapted from Gecht *et al.* (2020) [291].

5.3.7 Running multiple simulations on a single node

Multiple simulations on a single node result in a higher accumulated performance. The observation that simulations increase their performance only by a little when being run on two instead of one mixed CPU-GPU nodes indicates that the GPUs are not fully utilized. To investigate this, new benchmarks were performed where the same system was run on a single node multiple times. The system was started using GROMACS “multidir” feature with $n = 1, 2, 4$ and 8 identical simulations at the same time. The highest total performance, the sum of the performances of each repetition, can be achieved when running $n = 4$ simulations on a single node (Figure 5.9).

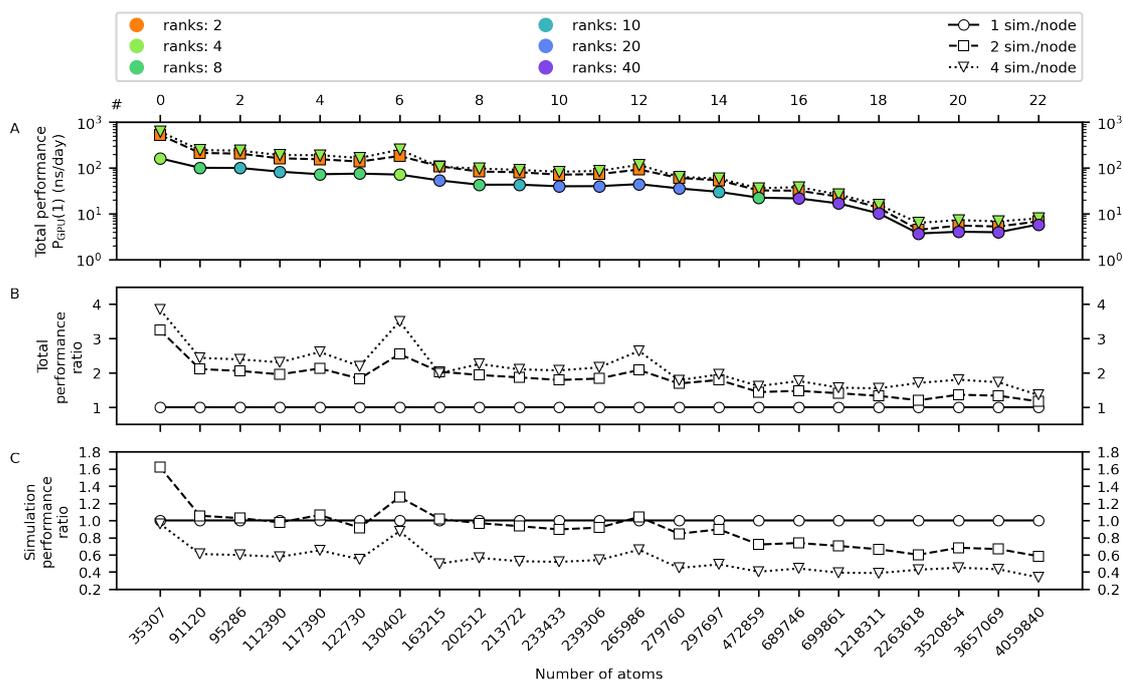


Figure 5.9: Multiple simulations per node increase performance. n identical simulations run on a single node increase the total performance per node when compared to a single simulation on a single node ($n = 1$), independent of system size. (A) The total performance for a single simulation per node ($n = 1$, circles, solid lines), two simulations per node ($n = 2$, squares, dashed lines), and four simulations per node ($n = 4$, pentagons, dotted lines) for the optimal rank/thread combinations (color). (B) The ratio of the total performance per node when running n and $n = 1$ simulations per node ($n = 1, 2, 4$). (C) The ratio of the simulation performance with n and $n = 1$ equivalent simulations running on a single node. The summed performance of all n simulations divided by the performance of a single ($n = 1$) simulation. Lines serve as guide to the eye. Adapted from Gecht *et al.* (2020) [291].

Running more than four repetitions on a single node did not increase the performance and the data was omitted. When simulations are run in this mode, the optimal number of MPI ranks is not the same as was observed in the previous sections. The number of MPI ranks is the same independent of the system size when multiple simulations are run on the same node (Figure 5.9A). The smallest system of 35 k atoms increases its performance by 4-fold, whereas systems with less than 100 k atoms increase their performance by 2-fold. Bigger systems are able to increase their performance by 1.5-fold. It is important to note that running multiple simulations on a single node will in most cases decrease the length

of a single simulation (Figure 5.9C). Nevertheless the total simulation time increases.

5.4 Discussion

The choice of run settings for a MD simulation is crucial for its achieved performance in ns per day. For the single user the wrong choice can degrade performance, increase the amount of real-time required to obtain results and thus delay their scientific progress. In the bigger picture, wasteful resource usage amplifies this problem even further. Queued jobs might not start because resources are blocked by too many requested nodes of one job, thus limiting the supercomputers' job turnover. On a global scale, running a job on more computational resources than necessary directly increases energy consumption, through the hardware of the supercomputer itself (CPU, GPU, RAM) and its accompanying cooling. In today's awareness of global warming, it is important for everyone to think about their carbon footprint. Scientists conducting research through MD simulations should try and limit their personal CO₂ emission by optimizing the resource usage of their simulations.

The accessible command-line MDBenchmark tool was developed to help researchers benchmark their MD simulations and estimate the best settings to use. The choice of $n_{\text{ranks}} = 40$ are most favorable for CPU-only nodes independent of system size, as well as for large systems with more than 2 M atoms when using mixed CPU-GPU nodes. These kinds of systems might dominate the MD field in the future, but nowadays smaller systems (<400 k atoms) are more common and require different settings, especially when run with the addition of GPUs. The results presented in this chapter show that few MPI ranks, i.e., $n_{\text{ranks}} = 8$, are the best choice for systems of this size on mixed CPU-GPU nodes without hyperthreading. The number of required MPI ranks increases with the system size, as well as with the number of nodes a system is run on. This is correlated with the domain decomposition that is performed by GROMACS using MPI ranks. A big simulation box can be decomposed into smaller boxes that can be calculated separately. In contrast, a small system is not able to be decomposed into many smaller boxes, because it is just too small. Although the results in this chapter are a good guidance, it is recommended to benchmark every system and make sure that the chosen settings are indeed most efficient by personal measures. The MDBenchmark software provides an easy-to-use interface to set up, run and analyze such benchmarks. The most surprising results stem from the benchmarks of running multiple simulations on a single node. As the GPU resources of a single node are not fully used by one simulation they can be shared with more instances of the same. When running multiple simulations on a single node, each instance produces a shorter trajectory per day. At the same time the sum of performance for all these instances is greater than when a single simulation is run on one node. Therefore, if one is interested in running many short trajectories, for example when trying to observe rare events, it is beneficial to start multiple simulations on the same node. In contrast, if one wants to produce one single long trajectory, it remains advisable to run simulation on as many nodes as needed, but as few as possible. The number of nodes can easily be estimated with MDBenchmark.

The effects of offloading specific calculations, i.e., PME, to separate GPU ranks remains unexplored. Future work will hopefully automate the search of parameter space, to further

tune the efficiency of simulations. In the long run, as computational resources become more and more ubiquitous, researchers should not spend their time with optimization of simulation parameters and decisions of which is the correct supercomputer and hardware for their calculations. Instead, a software solution that automates the job submission to available supercomputers, data management and synchronization, as well as preliminary analysis and reporting would be suitable for this job.

5.5 Appendix

Table 5.1: Atomistic simulation systems used in this chapter to perform scaling benchmarks. The systems vary in composition, box geometry, and size, as well as in force fields and water models: CHARMM36 [189], CHARMM36m [179], AMBER99SB*-ILDN [293–295], AMBER99SB*-ILDN-Q [293–297], TIP3P [250], and TIP4P-D [298]. The box geometries are cuboids (C), rhombic dodecahedra (RD), and hexagonal prisms (H). Reproduced from Gecht *et al.* (2020) [291].

	# atoms	Box	System size (nm ³)	Force field	Water model	Reference
0	35,307	RD	8.0	AMBER99SB*-ILDN	TIP3P	Unpublished
1	91,120	C	9.9x9.9x9.9	CHARMM36	TIP3P	Unpublished
2	95,286	C	10.0x10.0x10.0	CHARMM36	TIP3P	Unpublished
3	112,390	C	10.0x10.0x10.0	AMBER99SB*-ILDN-Q	TIP3P	Unpublished
4	117,390	C	11.0x11.0x10.0	CHARMM36	TIP3P	Hofbauer <i>et al.</i> [157]
5	122,730	C	10.0x10.0x10.0	AMBER99SB*-ILDN-Q	TIP4P-D	Unpublished
6	130,402	C	9.9x9.9x9.9	AMBER99SB*-ILDN-Q	TIP4P-D	Bülow <i>et al.</i> [299]
7	163,215	C	13.1x13.1x9.1	CHARMM36m	TIP3P	Unpublished
8	202,512	RD	13.0	AMBER99SB*-ILDN-Q	TIP4P-D	Unpublished
9	213,722	RD	14.5x14.5x10.3	AMBER99SB*-ILDN	TIP3P	Unpublished
10	233,433	C	14.6x14.6x10.5	CHARMM36m	TIP3P	Wu <i>et al.</i> [300]
11	239,306	C	14.6x14.6x10.9	CHARMM36	TIP3P	Unpublished
12	265,986	C	12.6x12.6x12.6	AMBER99SB*-ILDN-Q	TIP4P-D	Bülow <i>et al.</i> [299]
13	279,760	C	15.9x15.9x10.8	CHARMM36	TIP3P	Unpublished
14	297,697	C	13.8x13.8x15.4	CHARMM36m	TIP3P	Hofmann <i>et al.</i> [301]
15	472,859	C	15.0x15.0x15.0	AMBER99SB*-ILDN-Q	TIP4P-D	Unpublished
16	689,746	C	17.3x17.3x17.3	AMBER99SB*-ILDN-Q	TIP4P-D	Bülow <i>et al.</i> [299]
17	699,861	C	19.5x19.5x18.1	CHARMM36m	TIP3P	Unpublished
18	1,219,446	C	23.0x23.0x23.0	CHARMM36m	TIP3P	Unpublished
19	2,263,618	C	38.2x21.6x31.5	CHARMM36	TIP3P	Unpublished
20	3,520,854	C	30.0x30.0x30.0	AMBER99SB*-ILDN-Q	TIP4P-D	Bülow <i>et al.</i> [299]
21	3,657,069	C	30.4x30.4x30.4	AMBER99SB*-ILDN-Q	TIP4P-D	Bülow <i>et al.</i> [299]
22	4,059,840	H	34.9x34.9x37.9	CHARMM36m	TIP3P	Turoňová <i>et al.</i> [288]

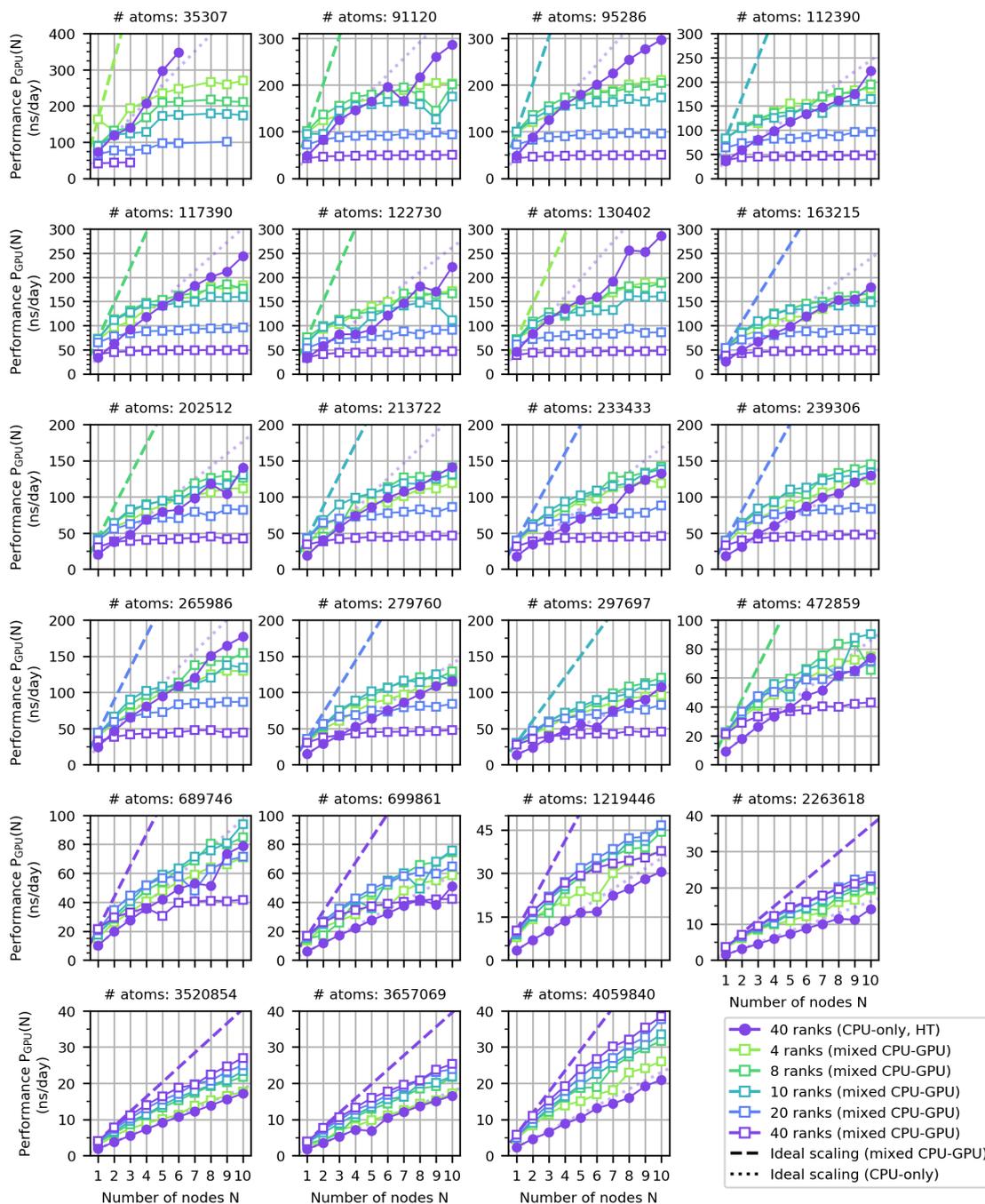


Figure 5.10: The optimal choice of n_{ranks} depends on the system size when using mixed CPU-GPU nodes. Performance $P_{\text{GPU}}(N)$ of 23 MD simulations with varying sizes as a function of nodes N for different numbers of MPI ranks n_{ranks} on different numbers of mixed CPU-GPU nodes without hyperthreading (colored squares). The best performing setting with CPU-only nodes is shown as reference with circles. Each data point in the performance plot shows one independent run of 15 minutes each. Transparent dotted and dashed lines show ideal scaling for CPU-only and mixed CPU-GPU benchmarks according to Equation (5.4). Adapted from Gecht *et al.* (2020) [291].

```
1 #!/bin/bash -l
2 #SBATCH -o ./benchjob.out.%j
3 #SBATCH -e ./benchjob.err.%j
4 #SBATCH -D ./
5 #SBATCH -J n001_r40_t02_wht
6 #
7 #SBATCH --nodes=1
8 # Set the number of tasks per node (=MPI ranks)
9 #SBATCH --ntasks-per-node=40
10 # Set the number of threads per rank (=OpenMP threads)
11 #SBATCH --cpus-per-task=2
12 # Enable hyperthreading
13 #SBATCH --ntasks-per-core=2
14 #SBATCH --time=00:17:00
15
16 module purge
17 module load gcc
18 module load impi
19 module load cuda
20 module load gromacs/2018.8
21
22 # Set number of OpenMP threads and proper core pinning with hyperthreading
23 export OMP_NUM_THREADS=$SLURM_CPUS_PER_TASK
24 export OMP_PLACES=threads
25 export SLURM_HINT=multithread
26
27 # Run gromacs/2018.8 for 15 minutes
28 srun gmx_mpi mdrun -v -ntomp $OMP_NUM_THREADS -maxh 0.25 -rethway -deffnm
    md -noconfout
```

Listing 5.1: SLURM submission script for a 15 minutes run on a CPU-only single node with 40 MPI ranks, 2 OpenMP threads and hyperthreading enabled. Reproduced from Gecht *et al.* (2020) [291].

```
1 #!/bin/bash -l
2 #SBATCH -o ./benchjob.out.%j
3 #SBATCH -e ./benchjob.err.%j
4 #SBATCH -D ./
5 #SBATCH -J n001_r20_t02_woht
6 #
7 #SBATCH --constraint="gpu"
8 #SBATCH --gres=gpu:rtx5000:2
9 #
10 #SBATCH --nodes=1
11 # Set the number of tasks per node (=MPI ranks)
12 #SBATCH --ntasks-per-node=20
13 # Set the number of threads per rank (=OpenMP threads)
14 #SBATCH --cpus-per-task=2
15 #SBATCH --time=00:17:00
16
17 module purge
18 module load gcc
19 module load impi
20 module load cuda
21 module load gromacs/2018.8
22
23 # Set number of OpenMP threads and proper core pinning without
    hyperthreading
24 export OMP_NUM_THREADS=$SLURM_CPUS_PER_TASK
25 export OMP_PLACES=cores
26
27 # Run gromacs/2018.8 for 15 minutes
28 srun gmx_mpi mdrun -v -ntomp $OMP_NUM_THREADS -maxh 0.25 -rethway -deffnm
    md -noconfout
```

Listing 5.2: SLURM submission script for a 15 minutes run on a single mixed CPU-GPU node with 20 MPI ranks, 2 OpenMP threads and hyperthreading disabled. Reproduced from Gecht *et al.* (2020) [291].

Conclusions

The work presented in this thesis contributes to further the understanding of protein-lipid interactions. The interactions of proteins with membranes are an important function of biological life, but remain poorly understood. The atomistic details of these interactions are hard to study with traditional experimental techniques and therefore MD simulations are a great tool to shine light onto these processes. The result chapters investigate different aspects of how proteins interact with membranes and its constituents. In the second chapter the lipid interactions of two AH, those of Opi1 and PIP5K, are studied. The third chapter shows how the Opi1 AH can associate with the membrane in its multi-step lifecycle. The fourth chapter considered NMR structures of HIV-1 gp41 and studied their stability in lipid bilayers of different compositions. In the fifth chapter, the software package MDBenchmark is presented that help researchers run their MD simulations more efficiently and use only the necessary resources to achieve the most optimal simulation performance. This chapter concludes this dissertation by briefly summarizing the results of each chapter.

Extensive MD simulations were performed to study the lipid interactions between membranes and the AHs of Opi1 and PIP5K. The AHs were first modeled as ideal α -helices and then embedded into lipid membranes of varying compositions. In the case of the Opi1 AH, simple model membranes highly abundant in PC were used. In addition to the uncharged PC headgroups, charged lipid species such as PA and PS were added. These two charged species do not differ by their total charge of -2 in the simulation, but by their headgroup size and the charge distribution on the headgroup. The simulations revealed that the AH of Opi1 is lined with two regions of positively charged residues. These were named “KRK motif” and “3K motif” based on amino acids that are contained in each motif. Both of these motifs are able to locally enrich charged lipids with either PA or PS headgroups around them. Although the membranes are highly abundant in uncharged lipids with PC headgroups, they are not enriched by the AH. This points towards specific charged interactions between the lipid headgroup and the positively charged amino acids found in the two motifs. Most strikingly, in membranes containing both PA and PS headgroups the AH of Opi1 was able to specifically recognize PA and enrich it at the “3K motif”, whereas PS was excluded. This observation lead to the discovery that indeed the “3K motif” is tuned to interact with PA headgroups. The small headgroup of PA fits nicely into the binding pocket formed by the “3K motif”, whereas PS is too bulky to do so. In addition, its distributed charges allow its headgroup to interact with the binding pocket in various ways. This leads to a shorter lifetime of PS in the binding pocket and the possibility to be outcompeted by PA. In contrast, the PIP5K AH was studied with a more complex

membrane containing the lipid species PS, PI4P, PC and cholesterol. PIP5K catalyzes the phosphorylation reaction from PI4P to PI(4,5)P₂, so it must specifically recognize this headgroup and differentiate it from other charged lipid species. Its AH is lined with a variety of positively charged residues, but does not contain any localized binding pockets as found on the Opi1 AH. Instead, the amino acids are distributed along the sides and top of the AH, as seen when it is embedded into the membrane. This distribution of charged residues allows the AH to intimately interact with the highly charged headgroup of PI4P. Different binding modes between the AH and PI4P were also characterized. It was shown that the lipid can start interacting with the AH from afar and be dragged closer to it by electrostatic interactions. Similarly, when it binds to the AH it can reside at its flanks at slightly different positions, when considering only the acyl chains. The large and highly flexible inositol ring on its headgroup allows the lipid to bend around the AH and reach towards positively charged residues that could not be accessed with a smaller headgroup. Binding from both sides of the AH has been observed, but it remains unclear if the full-length protein can catalyze its phosphorylation reaction from either side.

The interaction of the Opi1 AH with different lipids species is just a part in the lifecycle of the AH. The protein Opi1 is soluble and localizes to the ER membrane with the help of its AH. The recruitment, association and insertion of the AH with lipid membranes was studied in the third chapter of this thesis. First, different extensive MD simulations were set up to study the association of the unfolded AH to lipid membranes. It was shown that the AH is quickly recruited to charged lipid surfaces, where it associates with the membrane by inserting either of two hydrophobic residues, L113 or I117, into the membrane. In the presence of an uncharged membrane, the AH is also able to transiently insert the two hydrophobic residues into the membrane. Nevertheless it fails to stay associated with it due to a lack of electrostatic interactions. Different simulations containing membrane compositions with varying amounts of PC, PE and PA revealed that the sole presence of the negatively charged PA headgroup is sufficient to permanently recruit the unfolded AH at the lipid-water interface. After recruitment and tethering to the membrane, AHs undergo a folding reaction where they form their α -helical secondary structure and embed into the membrane. An atomistic 18.8 μ s simulation captured the folding and insertion event of two turns of the AH. Unfortunately, no subsequent folding was observed in this simulation and the partially folded AH was ejected from the membrane after being inserted for a duration of 10 μ s. The folding of the AH stopped at the “KGK motif”, a short linear sequence of lysine-glycine-lysine. This short sequence turned out to be a switch that is able to destabilize the secondary structure of the AH and at the same time rate-limit the folding upon membrane association. In addition, Gly-120 is a highly conserved residue among *Saccharomyces*. These observations were tested in laboratory experiments by Dr. Harald Hofbauer, who was able to confirm that mutations of G120 lead to a defect Opi1 protein *in vivo*. To test if a mutated version of the AH is able to fold quicker, additional simulations of the unfolded Opi1 AH with G120A single and G120A/K121A double mutations were set up. Although transient folding of the AH was observed, the full α -helix was never formed in the simulations. These extensive simulations were unable to fully sample the folding

pathway of the AH. To overcome this sampling barrier, enhanced sampling techniques were used to fold the remainder of the α -helix. These attempts have managed to fully fold the AH outside of the membrane, but were not able to insert it. Nevertheless, the sequence of insertion and folding remains unclear. Does the full α -helix form outside the membrane and then insert at the same time, or does it form on a turn-by-turn basis?

The stability of the NMR structure of HIV-1 gp41 was studied in the fourth chapter of this thesis. While multiple structures of the HIV-1 Env proteins have been solved using both X-ray crystallography, they always lack the TMD of gp41. Multiple NMR structures have emerged over the last years that describe the TMD in different oligomeric states. In three structures, the TMD is trimeric, whereas it is monomeric in the other. To better understand the differences in structures, simulations of the gp41 TMD were performed. At first, a monomeric version of the TMD was modeled as ideal α -helix and embedded into a viral-like membrane. A highly conserved arginine, R696, is found at the center of the TMD. The monomeric simulations have shown that it preferentially snorkels towards the lipid-water interface, leading to a highly tilted conformation of the TMD. R696 can snorkel towards either of two leaflets, leading to two distinct conformations. The interaction of R696 with solvent locks the TMD in a certain orientation and forbids it to freely rotate along its principal axis. A second construct containing the MPER domain of gp41 was also set up as ideal α -helix and embedded into the same membrane. MPER is lined with hydrophobic tryptophan residues that have a preference to insert into lipid bilayers. The two possible orientations of R696 also have an effect on the interaction of MPER with the membrane, as in one R696 orientation the tryptophan residues are all unable to embed into it. The connecting region between MPER and TMD is found to be unstructured in one of the NMR structures, pointing to the fact that the observation in the monomeric MPER-TMD construct might not be of great importance. Additional simulations of trimeric gp41 TMD were set up to study the conformational dynamics of the construct in viral-like membranes. Similar to the monomer, R696 snorkeled towards the lipid-water interface, but was limited to the cytofacial leaflet only. This snorkeling led to the introduction of water molecules into the membrane leaflet and ultimately to the dissociation of the C-terminal ends of the gp41 TMD. The C_3 symmetric conformation was not recovered in the two 10 μ s simulations after its dissociation. To test if the instability of the trimeric structure is due to the lack of its N-terminally preceding MPER domain, new simulations with the trimeric MPER-TMD construct were set up. To reproduce the experimental conditions, a simple DMPC bilayer was set up to mimic a bicelle bilayer. The systems were equilibrated with additional caution to try and keep the structures stable. Multiple independent simulations of the same system revealed that the structure is inherently unstable and dissociates into a similar conformation as has been observed with the gp41 TMD construct in viral-like bilayers. Surprisingly, these atomistic MD simulations recover the epitopes of bnAbs whose crystal structures are available. This made it possible to superimpose different bnAbs Fab fragments and show how an antibody might bind to gp41. To complete the picture, the soluble part of the HIV-1 Env domain was prepared for simulations and a full-length construct of the envelope protein, without the CTD, was constructed. This full picture

of the Env protein showcases that it is possible albeit difficult for bnAbs to interact with their epitopes. To properly do this, the soluble part of Env must first tilt to allow for proper binding of the Fab fragment. Simulations of the Fab fragment bound to its epitope were ultimately not stable and always led to the dissociation of the fragment from its epitope. A future model must also be heavily glycosylated, because the glycans will limit the accessibility of the bnAb even further.

Throughout this thesis MD simulations were used to help answer research questions. These simulations use computational resources to calculate the trajectories of the biological systems of interest. As researchers are interested in a quick turnover of results, it is of utmost importance to run these simulations with a high performance. To do this, MD engines available at the time of writing parallelize their computations across multiple machines and their varying hardware. Doing this is a straightforward process, but it is easy to neglect the fact that adding more computing resources does not result in a linear performance increase. Instead, the performance saturates after the addition of a certain amount of resources. Researchers are supposed to benchmark their simulations to use as many resources as needed, as few as possible. At the start of the thesis no tools were available to automate this process and researchers often conducted such benchmark scaling studies in a labor intense manual process. This thesis presents the software package MDBenchmark that provides an easy to use CLI for users to set up, submit and analyze benchmarks with. The software allows the user to set up benchmarks that scale across a different number of nodes, toggle the usage of available GPU resources, as well as the number of MPI ranks and whether hyperthreading should be enabled. This enables the user to quickly generate different permutations of these settings and try them out before running their production simulation. The software was developed in the open on GitHub, where the code is available for others to verify and extend. All of the features of MDBenchmark are documented in its online documentation that can be found at <https://mdbenchmark.readthedocs.io/en/>. To this date the software is in use by hundreds of researchers all across the globe and helps to run their simulations more efficiently. As a proof of concept the fifth chapter of this thesis used MDBenchmark to perform a scaling study of 23 representative atomistic simulation systems. The results show that it is possible to achieve an increased simulation performance by only adjusting the number of MPI ranks. When using CPU-only nodes it is advisable to use as many MPI ranks as available physical cores, whereas the number of MPI ranks on mixed CPU-GPU nodes depends on the system size. Bigger systems require more MPI ranks, as their system size allows for a domain decomposition into lots of smaller domains. In contrast, smaller systems require less MPI ranks due to limitations in the possible domain decomposition. A surprising result was that running multiple simulations on the same mixed CPU-GPU node increases the overall simulation performance. This allows the more efficient usage of resources for certain use cases, for example when users are interested in running many repetitions of a simulation. It is desirable that MDBenchmark will be further extended by its users in the future and adapt to upcoming challenges. The open nature of the software allows everyone to help with the development.

This thesis has shown how MD simulations can be used to study direct protein-lipid

interactions and better understand the function of proteins. The simulations can also be used as a tool to propose new laboratory experiments, as observations from simulations can be carried over. In addition, the power of MD simulations to study the stability of NMR structures was showcased, as well as the possibility of using simulations as a “computational microscope” to understand the binding of antibodies to their epitope. The thesis ends with a practical application of how to enhance the performance of MD simulations and obtain results quicker.

Bibliography

- [1] R. Dawkins. *The Selfish Gene*. Oxford University Press, Oxford, 30. anniversary edition edition, 2006.
- [2] S. Grösch, S. Schiffmann and G. Geisslinger. Chain length-specific properties of ceramides. *Progress in Lipid Research*, 51(1):50–62, 2012.
- [3] A. Yamashita, Y. Hayashi, Y. Nemoto-Sasaki, M. Ito, S. Oka, T. Tanikawa, K. Waku and T. Sugiura. Acyltransferases and transacylases that determine the fatty acid composition of glycerolipids and the metabolism of bioactive lipid mediators in mammalian cells and model organisms. *Progress in Lipid Research*, 53:18–81, 2014.
- [4] T. Harayama, M. Eto, H. Shindou, Y. Kita, E. Otsubo, D. Hishikawa, S. Ishii, K. Sakimura, M. Mishina and T. Shimizu. Lysophospholipid Acyltransferases Mediate Phosphatidylcholine Diversification to Achieve the Physical Properties Required In Vivo. *Cell Metabolism*, 20(2):295–305, 2014.
- [5] B. Antonny, S. Vanni, H. Shindou and T. Ferreira. From zero to six double bonds: Phospholipid unsaturation and organelle function. *Trends in Cell Biology*, 25(7):427–436, 2015.
- [6] E. Sezgin, I. Levental, S. Mayor and C. Eggeling. The mystery of membrane organization: Composition, regulation and roles of lipid rafts. *Nature Reviews Molecular Cell Biology*, 18(6):361–374, 2017.
- [7] J. T. Hannich, D. Mellal, S. Feng, A. Zumbuehl and H. Riezman. Structure and conserved function of iso-branched sphingoid bases from the nematode *Caenorhabditis elegans*. *Chemical Science*, 8(5):3676–3686, 2017.
- [8] C. Tanford. The hydrophobic effect and the organization of living matter. *Science*, 200(4345):1012–1018, 1978.
- [9] J. C. M. Holthuis and A. K. Menon. Lipid landscapes and pipelines in membrane homeostasis. *Nature*, 510(7503):48–57, 2014.
- [10] E. Fahy, S. Subramaniam, R. C. Murphy, M. Nishijima, C. R. H. Raetz, T. Shimizu, F. Spener, G. van Meer, M. J. O. Wakelam and E. A. Dennis. Update of the LIPID MAPS comprehensive classification system for lipids. *Journal of Lipid Research*, 50(Supplement):S9–S14, 2009.
- [11] T. Harayama and H. Riezman. Understanding the diversity of membrane lipid composition. *Nature Reviews Molecular Cell Biology*, 2018.

- [12] D. Marsh. Lateral Pressure Profile, Spontaneous Curvature Frustration, and the Incorporation and Conformation of Proteins in Membranes. *Biophysical Journal*, 93(11):3884–3899, 2007.
- [13] G. van Meer, D. R. Voelker and G. W. Feigenson. Membrane lipids: Where they are and how they behave. *Nature Reviews Molecular Cell Biology*, 9(2):112–124, 2008.
- [14] K. Simons and E. Ikonen. Functional rafts in cell membranes. *Nature*, 387(6633):569–572, 1997.
- [15] J. H. Lorent, K. R. Levental, L. Ganesan, G. Rivera-Longworth, E. Sezgin, M. Doktorova, E. Lyman and I. Levental. Plasma membranes are asymmetric in lipid unsaturation, packing and protein shape. *Nature Chemical Biology*, 16(6):644–652, 2020.
- [16] J. Bigay and B. Antonny. Curvature, Lipid Packing, and Electrostatics of Membrane Organelles: Defining Cellular Territories in Determining Specificity. *Developmental Cell*, 23(5):886–895, 2012.
- [17] C. L. Jackson. Mechanisms of transport through the Golgi complex. *Journal of Cell Science*, 122(4):443–452, 2009.
- [18] G. H. Patterson, K. Hirschberg, R. S. Polishchuk, D. Gerlich, R. D. Phair and J. Lippincott-Schwartz. Transport through the Golgi Apparatus by Rapid Partitioning within a Two-Phase Membrane System. *Cell*, 133(6):1055–1067, 2008.
- [19] M.-O. Roy, R. Leventis and J. R. Silvius. Mutational and Biochemical Analysis of Plasma Membrane Targeting Mediated by the Farnesylated, Polybasic Carboxy Terminus of K-ras4B. *Biochemistry*, 39(28):8298–8307, 2000.
- [20] T. Yeung, G. E. Gilbert, J. Shi, J. Silvius, A. Kapus and S. Grinstein. Membrane Phosphatidylserine Regulates Surface Charge and Protein Localization. *Science*, 319(5860):210–213, 2008.
- [21] G. Fairn, N. Schieber, N. Ariotti, S. Murphy, L. Kuerschner, R. Webb, S. Grinstein and R. Parton. High-resolution mapping reveals topologically distinct cellular pools of phosphatidylserine. *The Journal of Cell Biology*, 194(2):257–275, 2011.
- [22] L. Vamparys, R. Gautier, S. Vanni, W. D. Bennett, D. P. Tieleman, B. Antonny, C. Etchebest and P. F. Fuchs. Conical Lipids in Flat Bilayers Induce Packing Defects Similar to that Induced by Positive Curvature. *Biophysical Journal*, 104(3):585–593, 2013.
- [23] S. Vanni, L. Vamparys, R. Gautier, G. Drin, C. Etchebest, P. F. J. Fuchs and B. Antonny. Amphipathic lipid packing sensor motifs: Probing bilayer defects with hydrophobic residues. *Biophysical Journal*, 104(3):575–584, 2013.
- [24] B. Antonny. Mechanisms of membrane curvature sensing. *Annual Review of Biochemistry*, 80:101–123, 2011.

- [25] B. Brügger, R. Sandhoff, S. Wegehingel, K. Gorgas, J. Malsam, J. B. Helms, W.-D. Lehmann, W. Nickel and F. T. Wieland. Evidence for Segregation of Sphingomyelin and Cholesterol during Formation of Copi-Coated Vesicles. *The Journal of Cell Biology*, 151(3):507–518, 2000.
- [26] A. Roux, D. Cuvelier, P. Nassoy, J. Prost, P. Bassereau and B. Goud. Role of curvature and phase transition in lipid sorting and fission of membrane tubules. *The EMBO Journal*, 24(8):1537–1545, 2005.
- [27] B. Sorre, A. Callan-Jones, J.-B. Manneville, P. Nassoy, J.-F. Joanny, J. Prost, B. Goud and P. Bassereau. Curvature-driven lipid sorting needs proximity to a demixing point and is aided by proteins. *Proceedings of the National Academy of Sciences of the United States of America*, 106(14):5622–5626, 2009.
- [28] J. Liu, Y. Sun, D. G. Drubin and G. F. Oster. The Mechanochemistry of Endocytosis. *PLOS Biology*, 7(9):e1000204, 2009.
- [29] H. Zhao, A. Michelot, E. V. Koskela, V. Tkach, D. Stamou, D. G. Drubin and P. Lapalainien. Membrane-Sculpting BAR Domains Generate Stable Lipid Microdomains. *Cell Reports*, 4(6):1213–1223, 2013.
- [30] A. Radhakrishnan, J. L. Goldstein, J. G. McDonald and M. S. Brown. Switch-like Control of SREBP-2 Transport Triggered by Small Changes in ER Cholesterol: A Delicate Balance. *Cell Metabolism*, 8(6):512–521, 2008.
- [31] R. B. Cornell and N. D. Ridgway. CTP:phosphocholine cytidyltransferase: Function, regulation, and structure of an amphitropic enzyme required for membrane biogenesis. *Progress in Lipid Research*, 59:147–171, 2015.
- [32] L. E. Cybulski, M. Martín, M. C. Mansilla, A. Fernández and D. de Mendoza. Membrane Thickness Cue for Cold Sensing in a Bacterium. *Current Biology*, 20(17):1539–1544, 2010.
- [33] M. E. Inda, M. Vandenbranden, A. Fernández, D. de Mendoza, J.-M. Ruysschaert and L. E. Cybulski. A lipid-mediated conformational switch modulates the thermosensing activity of DesK. *Proceedings of the National Academy of Sciences of the United States of America*, 111(9):3579–3584, 2014.
- [34] M. Jonikas, S. Collins, V. Denic and E. Oh. Comprehensive characterization of genes required for protein folding in the endoplasmic reticulum. *Science*, 323(5922):1693–1697, 2009.
- [35] M. A. Surma, C. Klose, D. Peng, M. Shales, C. Mrejen, A. Stefanko, H. Braberg, D. E. Gordon, D. Vorkel, C. S. Ejsing, R. Farese, K. Simons, N. J. Krogan and R. Ernst. A Lipid E-MAP Identifies Ubx2 as a Critical Regulator of Lipid Saturation and Lipid Bilayer Stress. *Molecular Cell*, 51(4):519–530, 2013.

- [36] S. Bernales, F. R. Papa and P. Walter. Intracellular Signaling by the Unfolded Protein Response. *Annual Review of Cell and Developmental Biology*, 22(1):487–508, 2006.
- [37] Y. Kimata, D. Oikawa, Y. Shimizu, Y. Ishiwata-Kimata and K. Kohno. A role for BiP as an adjustor for the endoplasmic reticulum stress-sensing protein Ire1. *The Journal of Cell Biology*, 167(3):445–456, 2004.
- [38] Y. Kimata, Y. Ishiwata-Kimata, T. Ito, A. Hirata, T. Suzuki, D. Oikawa, M. Takeuchi and K. Kohno. Two regulatory steps of ER-stress sensor Ire1 involving its cluster formation and interaction with unfolded proteins. *The Journal of Cell Biology*, 179(1):75–86, 2007.
- [39] C. Sidrauski and P. Walter. The Transmembrane Kinase Ire1p Is a Site-Specific Endonuclease That Initiates mRNA Splicing in the Unfolded Protein Response. *Cell*, 90(6):1031–1039, 1997.
- [40] T. Promlek, Y. Ishiwata-Kimata, M. Shido, M. Sakuramoto, K. Kohno and Y. Kimata. Membrane aberrancy and unfolded proteins activate the endoplasmic reticulum stress sensor Ire1 in different ways. *Molecular Biology of the Cell*, 22(18):3520–3532, 2011.
- [41] R. Volmer, K. van der Ploeg and D. Ron. Membrane lipid saturation activates endoplasmic reticulum unfolded protein response transducers through their transmembrane domains. *Proceedings of the National Academy of Sciences of the United States of America*, 110(12):4628–4633, 2013.
- [42] K. Halbleib, K. Pesek, R. Covino, H. F. Hofbauer, D. Wunnicke, I. Hänelt, G. Hummer and R. Ernst. Activation of the Unfolded Protein Response by Lipid Bilayer Stress. *Molecular Cell*, 67(4):673–684.e8, 2017.
- [43] J. Lee, S. G. Taneva, B. W. Holland, D. P. Tieleman and R. B. Cornell. Structural Basis for Autoinhibition of CTP:Phosphocholine Cytidylyltransferase (CCT), the Regulatory Enzyme in Phosphatidylcholine Synthesis, by Its Membrane-binding Amphipathic Helix. *The Journal of Biological Chemistry*, 289(3):1742–1755, 2014.
- [44] N. Krahmer, Y. Guo, F. Wilfling, M. Hilger, S. Lingrell, K. Heger, H. W. Newman, M. Schmidt-Supprian, D. E. Vance, M. Mann, R. V. Farese and T. C. Walther. Phosphatidylcholine Synthesis for Lipid Droplet Expansion Is Mediated by Localized Activation of CTP:Phosphocholine Cytidylyltransferase. *Cell Metabolism*, 14(4):504–515, 2011.
- [45] R. B. Cornell. Membrane lipid compositional sensing by the inducible amphipathic helix of CCT. *Biochimica et Biophysica Acta (BBA) – Molecular and Cell Biology of Lipids*, 1861(8, Part B):847–861, 2016.
- [46] T. Hoppe, K. Matuschewski, M. Rape, S. Schlenker, H. D. Ulrich and S. Jentsch. Activation of a Membrane-Bound Transcription Factor by Regulated Ubiquitin/Proteasome-Dependent Processing. *Cell*, 102(5):577–586, 2000.

- [47] R. Covino, S. Ballweg, C. Stordeur, J. B. Michaelis, K. Puth, F. Wernig, A. Bahrami, A. M. Ernst, G. Hummer and R. Ernst. A Eukaryotic Sensor for Membrane Lipid Saturation. *Molecular Cell*, 63(1):49–59, 2016.
- [48] R. Covino, G. Hummer and R. Ernst. Integrated Functions of Membrane Property Sensors and a Hidden Side of the Unfolded Protein Response. *Molecular Cell*, 71(3):458–467, 2018.
- [49] M. Giménez-Andrés, A. Čopič and B. Antony. The Many Faces of Amphipathic Helices. *Biomolecules*, 8(3):45, 2018.
- [50] R. Ernst, S. Ballweg and I. Levental. Cellular mechanisms of physicochemical membrane homeostasis. *Current Opinion in Cell Biology*, 53:44–51, 2018.
- [51] S. A. Henry, M. L. Gaspar and S. A. Jesch. The response to inositol: Regulation of glycerolipid metabolism and stress response signaling in yeast. *Chemistry and Physics of Lipids*, 180:23–43, 2014.
- [52] M. J. Kelley and G. M. Carman. Purification and characterization of CDP-diacylglycerol synthase from *Saccharomyces cerevisiae*. *The Journal of Biological Chemistry*, 262(30):14563–14570, 1987.
- [53] S. Chand, E. L. Messina, W. AlSalmi, N. Ananthaswamy, G. Gao, G. Uritskiy, V. Padilla-Sanchez, M. Mahalingam, K. K. Peachman, M. L. Robb, M. Rao and V. B. Rao. Glycosylation and oligomeric state of envelope protein might influence HIV-1 virion capture by A4 β 7 integrin. *Virology*, 508:199–212, 2017.
- [54] C. J. Clancey, S. C. Chang and W. Dowhan. Cloning of a gene (PSD1) encoding phosphatidylserine decarboxylase from *Saccharomyces cerevisiae* by complementation of an *Escherichia coli* mutant. *Journal of Biological Chemistry*, 268(33):24580–24590, 1993.
- [55] P. J. Trotter, J. Pedretti and D. R. Voelker. Phosphatidylserine decarboxylase from *Saccharomyces cerevisiae*. Isolation of mutants, cloning of the gene, and creation of a null allele. *The Journal of Biological Chemistry*, 268(28):21416–21424, 1993.
- [56] T. Kodaki and S. Yamashita. Characterization of the methyltransferases in the yeast phosphatidylethanolamine methylation pathway by selective gene disruption. *European Journal of Biochemistry*, 185(2):243–251, 1989.
- [57] J. Nikawa, Y. Tsukagoshi and S. Yamashita. Cloning of a gene encoding choline transport in *Saccharomyces cerevisiae*. *Journal of Bacteriology*, 166(1):328–330, 1986.
- [58] K. Kim, K. H. Kim, M. K. Storey, D. R. Voelker and G. M. Carman. Isolation and characterization of the *Saccharomyces cerevisiae* EKI1 gene encoding ethanolamine kinase. *The Journal of Biological Chemistry*, 274(21):14857–14866, 1999.

- [59] K. Hosaka, T. Kodaki and S. Yamashita. Cloning and characterization of the yeast CKI gene encoding choline kinase and its expression in *Escherichia coli*. *Journal of Biological Chemistry*, 264(4):2053–2059, 1989.
- [60] R. Min-Seek, Y. Kawamata, H. Nakamura, A. Ohta and M. Takagi. Isolation and Characterization of ECTI Gene Encoding CTP: Phosphoethanolamine Cytidylyltransferase of *Saccharomyces cerevisiae*. *The Journal of Biochemistry*, 120(5):1040–1047, 1996.
- [61] Y. Tsukagoshi, J.-i. Nikawa and S. Yamashita. Molecular cloning and characterization of the gene encoding cholinephosphate cytidylyltransferase in *Saccharomyces cerevisiae*. *European Journal of Biochemistry*, 169(3):477–486, 1987.
- [62] R. H. Hjelmstad and R. M. Bell. The sn-1,2-diacylglycerol ethanolaminephosphotransferase activity of *Saccharomyces cerevisiae*. Isolation of mutants and cloning of the EPT1 gene. *Journal of Biological Chemistry*, 263(36):19748–19757, 1988.
- [63] K. Natter, P. Leitner, A. Faschinger, H. Wolinski, S. McCraith, S. Fields and S. D. Kohlwein. The Spatial Organization of Lipid Synthesis in the Yeast *Saccharomyces cerevisiae* Derived from Large Scale Green Fluorescent Protein Tagging and High Resolution Microscopy. *Molecular & Cellular Proteomics*, 4(5):662–672, 2005.
- [64] S. A. Henry and J. L. Patton-Vogt. Genetic Regulation of Phospholipid Metabolism: Yeast as a Model Eukaryote. In *Progress in Nucleic Acid Research and Molecular Biology*, volume 61, pages 133–179. Academic Press, 1998.
- [65] A. S. Fischl and G. M. Carman. Phosphatidylinositol biosynthesis in *Saccharomyces cerevisiae*: Purification and properties of microsome-associated phosphatidylinositol synthase. *Journal of Bacteriology*, 154(1):304–311, 1983.
- [66] L. S. Klig and S. A. Henry. Isolation of the yeast INO1 gene: Located on an autonomously replicating plasmid, the gene is fully regulated. *Proceedings of the National Academy of Sciences of the United States of America*, 81(12):3816–3820, 1984.
- [67] T. Strahl and J. Thorner. Synthesis and Function of Membrane Phosphoinositides in Budding Yeast, *Saccharomyces cerevisiae*. *Biochimica et Biophysica Acta (BBA) – Molecular and Cell Biology of Lipids*, 1771(3):353–404, 2007.
- [68] L. A. Cowart and L. M. Obeid. Yeast Sphingolipids: Recent developments in understanding biosynthesis, regulation, and function. *Biochimica et Biophysica Acta (BBA) – Molecular and Cell Biology of Lipids*, 1771(3):421–431, 2007.
- [69] M. Pittet and A. Conzelmann. Biosynthesis and function of GPI proteins in the yeast *Saccharomyces cerevisiae*. *Biochimica et Biophysica Acta (BBA) – Molecular and Cell Biology of Lipids*, 1771(3):405–420, 2007.

- [70] C. J. R. Loewen, M. L. Gaspar, S. A. Jesch, C. Delon, N. T. Ktistakis, S. A. Henry and T. P. Levine. Phospholipid Metabolism Regulated by a Transcription Factor Sensing Phosphatidic Acid. *Science*, 304(5677):1644–1647, 2004.
- [71] G.-S. Han, W.-I. Wu and G. M. Carman. The *Saccharomyces cerevisiae* Lipin Homolog Is a Mg²⁺-dependent Phosphatidate Phosphatase Enzyme. *Journal of Biological Chemistry*, 281(14):9210–9218, 2006.
- [72] S. Rajakumari, R. Rajasekharan and G. Daum. Triacylglycerol lipolysis is linked to sphingolipid and phospholipid metabolism of the yeast *Saccharomyces cerevisiae*. *Biochimica et Biophysica Acta (BBA) – Molecular and Cell Biology of Lipids*, 1801(12):1314–1322, 2010.
- [73] C. Heier, U. Taschler, S. Rengachari, M. Oberer, H. Wolinski, K. Natter, S. D. Kohlwein, R. Leber and R. Zimmermann. Identification of Yju3p as functional orthologue of mammalian monoglyceride lipase in the yeast *Saccharomyces cerevisiae*. *Biochimica et Biophysica Acta (BBA) – Molecular and Cell Biology of Lipids*, 1801(9):1063–1071, 2010.
- [74] C. F. Kurat, K. Natter, J. Petschnigg, H. Wolinski, K. Scheuringer, H. Scholz, R. Zimmermann, R. Leber, R. Zechner and S. D. Kohlwein. Obese Yeast: Triglyceride Lipolysis Is Functionally Conserved from Mammals to Yeast. *Journal of Biological Chemistry*, 281(1):491–500, 2006.
- [75] N. Bachhawat, Q. Ouyang and S. A. Henry. Functional characterization of an inositol-sensitive upstream activation sequence in yeast. A cis-regulatory element responsible for inositol-choline mediated regulation of phospholipid biosynthesis. *The Journal of Biological Chemistry*, 270(42):25087–25095, 1995.
- [76] H. J. Schüller, R. Schorr, B. Hoffmann and E. Schweizer. Regulatory gene INO4 of yeast phospholipid biosynthesis is positively autoregulated and functions as a transactivator of fatty acid synthase genes FAS1 and FAS2 from *Saccharomyces cerevisiae*. *Nucleic Acids Research*, 20(22):5955–5961, 1992.
- [77] J. Ambroziak and S. A. Henry. INO2 and INO4 gene products, positive regulators of phospholipid biosynthesis in *Saccharomyces cerevisiae*, form a complex that binds to the INO1 promoter. *Journal of Biological Chemistry*, 269(21):15344–15349, 1994.
- [78] G. M. Carman and G.-S. Han. Regulation of phospholipid synthesis in *Saccharomyces cerevisiae* by zinc depletion. *Biochimica et Biophysica Acta (BBA) – Molecular and Cell Biology of Lipids*, 1771(3):322–330, 2007.
- [79] M. L. Greenberg and J. M. Lopes. Genetic regulation of phospholipid biosynthesis in *Saccharomyces cerevisiae*. *Microbiological Reviews*, 60(1):1–20, 1996.
- [80] L. S. Klig, M. J. Homann, G. M. Carman and S. A. Henry. Coordinate regulation of phospholipid biosynthesis in *Saccharomyces cerevisiae*: Pleiotropically constitutive opil mutant. *Journal of Bacteriology*, 162(3):1135–1141, 1985.

- [81] C. Wagner, M. Dietz, J. Wittmann, A. Albrecht and H.-J. Schüller. The negative regulator Opi1 of phospholipid biosynthesis in yeast contacts the pleiotropic repressor Sin3 and the transcriptional activator Ino2. *Molecular Microbiology*, 41(1):155–166, 2001.
- [82] G. Drin, J.-F. Casella, R. Gautier, T. Boehmer, T. U. Schwartz and B. Antonny. A general amphipathic α -helical motif for sensing membrane curvature. *Nature Structural & Molecular Biology*, 14(2):138–146, 2007.
- [83] M. Masuda, S. Takeda, M. Sone, T. Ohki, H. Mori, Y. Kamioka and N. Mochizuki. Endophilin BAR domain drives membrane curvature by two newly identified structure-based mechanisms. *The EMBO Journal*, 25(12):2889–2897, 2006.
- [84] S. Welker, B. Rudolph, E. Frenzel, F. Hagn, G. Liebisch, G. Schmitz, J. Scheuring, A. Kerth, A. Blume, S. Weinkauff, M. Haslbeck, H. Kessler and J. Buchner. Hsp12 is an intrinsically unstructured stress protein that folds upon membrane association and modulates membrane function. *Molecular Cell*, 39(4):507–520, 2010.
- [85] Y. Li, X. Gao and L. Chen. GroEL Recognizes an Amphipathic Helix and Binds to the Hydrophobic Side. *The Journal of Biological Chemistry*, 284(7):4324–4331, 2009.
- [86] G. Drin and B. Antonny. Amphipathic helices and membrane curvature. *FEBS Letters*, 584(9):1840–1847, 2010.
- [87] J. J. Smith and J. D. Aitchison. Peroxisomes take shape. *Nature Reviews Molecular Cell Biology*, 14(12):803–817, 2013.
- [88] Y. Liu, R. A. Kahn and J. H. Prestegard. Dynamic structure of membrane-anchored Arf•GTP. *Nature Structural & Molecular Biology*, 17(7):876–881, 2010.
- [89] B. Antonny, S. Beraud-Dufour, P. Chardin and M. Chabre. N-terminal hydrophobic residues of the G-protein ADP-ribosylation factor-1 insert into membrane phospholipids upon GDP to GTP exchange. *Biochemistry*, 36(15):4675–4684, 1997.
- [90] A. R. Thiam, R. V. Farese and T. C. Walther. The Biophysics and Cell Biology of Lipid Droplets. *Nature Reviews Molecular Cell Biology*, 14(12):775–786, 2013.
- [91] J. G. Donaldson and C. L. Jackson. ARF family G proteins and their regulators: Roles in membrane transport, development and disease. *Nature Reviews Molecular Cell Biology*, 12(6):362–375, 2011.
- [92] A. Godi, P. Pertile, R. Meyers, P. Marra, G. Di Tullio, C. Iurisci, A. Luini, D. Corda and M. A. De Matteis. ARF mediates recruitment of PtdIns-4-OH kinase-beta and stimulates synthesis of PtdIns(4,5)P2 on the Golgi complex. *Nature Cell Biology*, 1(5):280–287, 1999.

- [93] D. H. Jones, J. B. Morris, C. P. Morgan, H. Kondo, R. F. Irvine and S. Cockcroft. Type I phosphatidylinositol 4-phosphate 5-kinase directly interacts with ADP-ribosylation factor 1 and is responsible for phosphatidylinositol 4,5-bisphosphate synthesis in the golgi compartment. *The Journal of Biological Chemistry*, 275(18):13962–13966, 2000.
- [94] A. P. Herbert, M. Riesen, L. Bloxam, E. Kosmidou, B. M. Wareing, J. R. Johnson, M. M. Phelan, S. R. Pennington, L.-Y. Lian and A. Morgan. NMR Structure of Hsp12, a Protein Induced by and Required for Dietary Restriction-Induced Lifespan Extension in Yeast. *PLOS ONE*, 7(7):e41975, 2012.
- [95] J. Bigay, J.-F. Casella, G. Drin, B. Mesmin and B. Antonny. ArfGAP1 responds to membrane curvature through the folding of a lipid packing sensor motif. *EMBO Journal*, 24(13):2244–2253, 2005.
- [96] M. Magdeleine, R. Gautier, P. Gounon, H. Barelli, S. Vanni and B. Antonny. A filter at the entrance of the Golgi that selects vesicles according to size and bulk lipid composition. *eLife*, 5:e16988, 2016.
- [97] A. Čopič, S. Antoine-Bally, M. Giménez-Andrés, C. La Torre Garay, B. Antonny, M. M. Manni, S. Pagnotta, J. Guihot and C. L. Jackson. A giant amphipathic helix from a perilipin that is adapted for coating lipid droplets. *Nature Communications*, 9(1):1332, 2018.
- [98] M. Sharma, Z. Sasvari and P. D. Nagy. Inhibition of Sterol Biosynthesis Reduces Tombusvirus Replication in Yeast and Plants. *Journal of Virology*, 84(5):2270–2281, 2010.
- [99] S. Mañes, G. del Real, R. A. Lacalle, P. Lucas, C. Gómez-Moutón, S. Sánchez-Palomino, R. Delgado, J. Alcamí, E. Mira and C. Martínez-A. Membrane raft microdomains mediate lateral assemblies required for HIV-1 infection. *EMBO Reports*, 1(2):190–196, 2000.
- [100] B. M. Peterlin and D. Trono. Hide, shield and strike back: How HIV-infected cells avoid immune eradication. *Nature Reviews Immunology*, 3(2):97–107, 2003.
- [101] D. H. Nguyen and J. E. Hildreth. Evidence for budding of human immunodeficiency virus type 1 selectively from glycolipid-enriched membrane lipid rafts. *Journal of Virology*, 74(7):3264–3272, 2000.
- [102] Z. Liao, L. M. Cimasky, R. Hampton, D. H. Nguyen and J. E. Hildreth. Lipid rafts and HIV pathogenesis: Host membrane cholesterol is required for infection by HIV type 1. *AIDS Research and Human Retroviruses*, 17(11):1009–1019, 2001.
- [103] K. V. Khanna, K. J. Whaley, L. Zeitlin, T. R. Moench, K. Mehrazar, R. A. Cone, Z. Liao, J. E. Hildreth, T. E. Hoen, L. Shultz and R. B. Markham. Vaginal transmission of cell-associated HIV-1 in the mouse is blocked by a topical, membrane-modifying agent. *The Journal of Clinical Investigation*, 109(2):205–211, 2002.

- [104] J. Mercer, M. Schelhaas and A. Helenius. Virus entry by endocytosis. *Annual Review of Biochemistry*, 79:803–833, 2010.
- [105] E. O. Freed. HIV-1 and the host cell: An intimate association. *Trends in Microbiology*, 12(4):170–177, 2004.
- [106] R. A. Weiss. Virulence and pathogenesis. *Trends in Microbiology*, 10(7):314–317, 2002.
- [107] D. Horst, M. E. Rensing and E. J. H. J. Wiertz. Exploiting human herpesvirus immune evasion for therapeutic gain: Potential and pitfalls. *Immunology and Cell Biology*, 89(3):359–366, 2011.
- [108] R. L. Willey, F. Maldarelli, M. A. Martin and K. Strebel. Human immunodeficiency virus type 1 Vpu protein induces rapid degradation of CD4. *Journal of Virology*, 66(12):7193–7200, 1992.
- [109] J. G. Magadán, F. J. Pérez-Victoria, R. Sougrat, Y. Ye, K. Strebel and J. S. Bonifacino. Multilayered mechanism of CD4 downregulation by HIV-1 Vpu involving distinct ER retention and ERAD targeting steps. *PLOS Pathogens*, 6(4):e1000869, 2010.
- [110] M. Nethe, B. Berkhout and A. C. van der Kuyl. Retroviral superinfection resistance. *Retrovirology*, 2(1):52, 2005.
- [111] F. Barré-Sinoussi, J. C. Chermann, F. Rey, M. T. Nugeyre, S. Chamaret, J. Gruest, C. Dautet, C. Axler-Blin, F. Vézinet-Brun, C. Rouzioux, W. Rozenbaum and L. Montagnier. Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science*, 220(4599):868–871, 1983.
- [112] A. D. Frankel and J. A. Young. HIV-1: Fifteen proteins and an RNA. *Annual Review of Biochemistry*, 67:1–25, 1998.
- [113] B. G. Turner and M. F. Summers. Structural biology of HIV. *Journal of Molecular Biology*, 285(1):1–32, 1999.
- [114] A. Seelamgari, A. Maddukuri, R. Berro, C. de la Fuente, K. Kehn, L. Deng, S. Dadgar, M. E. Bottazzi, E. Ghedin, A. Pumfery and F. Kashanchi. Role of viral regulatory and accessory proteins in HIV-1 replication. *Frontiers in Bioscience: A Journal and Virtual Library*, 9:2388–2413, 2004.
- [115] A. G. Dalgleish, P. C. Beverley, P. R. Clapham, D. H. Crawford, M. F. Greaves and R. A. Weiss. The CD4 (T4) antigen is an essential component of the receptor for the AIDS retrovirus. *Nature*, 312(5996):763–767, 1984.
- [116] G. Alkhatib and E. A. Berger. HIV coreceptors: From discovery and designation to new paradigms and promise. *European Journal of Medical Research*, 12(9):375–384, 2007.

- [117] T. Dragic, V. Litwin, G. P. Allaway, S. R. Martin, Y. Huang, K. A. Nagashima, C. Cayanan, P. J. Maddon, R. A. Koup, J. P. Moore and W. A. Paxton. HIV-1 entry into CD4+ cells is mediated by the chemokine receptor CC-CKR-5. *Nature*, 381(6584):667–673, 1996.
- [118] H. Deng, R. Liu, W. Ellmeier, S. Choe, D. Unutmaz, M. Burkhart, P. Di Marzio, S. Marmon, R. E. Sutton, C. M. Hill, C. B. Davis, S. C. Peiper, T. J. Schall, D. R. Littman and N. R. Landau. Identification of a major co-receptor for primary isolates of HIV-1. *Nature*, 381(6584):661–666, 1996.
- [119] Y. Wu and A. Yoder. Chemokine Coreceptor Signaling in HIV-1 Infection and Pathogenesis. *PLoS Pathogens*, 5(12):e1000520, 2009.
- [120] B. Chen, E. M. Vogan, H. Gong, J. J. Skehel, D. C. Wiley and S. C. Harrison. Structure of an unliganded simian immunodeficiency virus gp120 core. *Nature*, 433(7028):834–841, 2005.
- [121] O. K. Haffar, S. Popov, L. Dubrovsky, I. Agostini, H. Tang, T. Pushkarsky, S. G. Nadler and M. Bukrinsky. Two nuclear localization signals in the HIV-1 matrix protein regulate nuclear import of the HIV-1 pre-integration complex. *Journal of Molecular Biology*, 299(2):359–368, 2000.
- [122] Y. Jenkins, M. McEntee, K. Weis and W. C. Greene. Characterization of HIV-1 vpr nuclear import: Analysis of signals and pathways. *The Journal of Cell Biology*, 143(4):875–885, 1998.
- [123] C. L. Woodward, S. Prakobwanakit, S. Mosessian and S. A. Chow. Integrase Interacts with Nucleoporin NUP153 To Mediate the Nuclear Import of Human Immunodeficiency Virus Type 1. *Journal of Virology*, 83(13):6522–6533, 2009.
- [124] G. Neumann, M. R. Castrucci and Y. Kawaoka. Nuclear import and export of influenza virus nucleoprotein. *Journal of Virology*, 71(12):9690–9700, 1997.
- [125] Y. Suzuki and R. Craigie. The road to chromatin - nuclear entry of retroviruses. *Nature Reviews Microbiology*, 5(3):187–196, 2007.
- [126] E. O. Freed. HIV-1 assembly, release and maturation. *Nature Reviews Microbiology*, 13(8):484–496, 2015.
- [127] W. Zhou, L. J. Parent, J. W. Wills and M. D. Resh. Identification of a membrane-binding domain within the amino-terminal region of human immunodeficiency virus type 1 Gag protein which interacts with acidic phospholipids. *Journal of Virology*, 68(4):2556–2569, 1994.
- [128] A. Ono, S. D. Ablan, S. J. Lockett, K. Nagashima and E. O. Freed. Phosphatidylinositol (4,5) biphosphate regulates HIV-1 Gag targeting to the plasma membrane. *Proceedings of the National Academy of Sciences of the United States of America*, 101(41):14889–14894, 2004.

- [129] B. Brügger, B. Glass, P. Haberkant, I. Leibrecht, F. T. Wieland and H.-G. Krausslich. The HIV lipidome: A raft with an unusual composition. *Proceedings of the National Academy of Sciences of the United States of America*, 103(8):2641–2646, 2006.
- [130] M. A. Checkley, B. G. Luttge and E. O. Freed. HIV-1 Envelope Glycoprotein Biosynthesis, Trafficking, and Incorporation. *Journal of Molecular Biology*, 410(4):582–608, 2011.
- [131] O. K. Haffar, D. J. Dowbenko and P. W. Berman. Topogenic analysis of the human immunodeficiency virus type 1 envelope glycoprotein, gp160, in microsomal membranes. *The Journal of Cell Biology*, 107(5):1677–1687, 1988.
- [132] C. K. Leonard, M. W. Spellman, L. Riddle, R. J. Harris, J. N. Thomas and T. J. Gregory. Assignment of intrachain disulfide bonds and characterization of potential glycosylation sites of the type 1 recombinant human immunodeficiency virus envelope glycoprotein (gp120) expressed in Chinese hamster ovary cells. *The Journal of Biological Chemistry*, 265(18):10373–10382, 1990.
- [133] H. B. Bernstein, S. P. Tucker, E. Hunter, J. S. Schutzbach and R. W. Compans. Human immunodeficiency virus type 1 envelope glycoprotein is modified by O-linked oligosaccharides. *Journal of Virology*, 68(1):463–468, 1994.
- [134] J. S. Allan, J. E. Coligan, F. Barin, M. F. McLane, J. G. Sodroski, C. A. Rosen, W. A. Haseltine, T. H. Lee and M. Essex. Major glycoprotein antigens that induce antibodies in AIDS patients are encoded by HTLV-III. *Science*, 228(4703):1091–1094, 1985.
- [135] A. Pinter, W. J. Honnen, S. A. Tilley, C. Bona, H. Zaghouani, M. K. Gorny and S. Zolla-Pazner. Oligomeric structure of gp41, the transmembrane protein of human immunodeficiency virus type 1. *Journal of Virology*, 63(6):2674–2679, 1989.
- [136] P. L. Earl, S. Koenig and B. Moss. Biological and immunological properties of human immunodeficiency virus type 1 envelope glycoprotein: Analysis of proteins with truncations and deletions expressed by recombinant vaccinia viruses. *Journal of Virology*, 65(1):31–41, 1991.
- [137] J. M. McCune, L. B. Rabin, M. B. Feinberg, M. Lieberman, J. C. Kosek, G. R. Reyes and I. L. Weissman. Endoproteolytic cleavage of gp160 is required for the activation of human immunodeficiency virus. *Cell*, 53(1):55–67, 1988.
- [138] J. Perrin, A. Bary, A. Vernay and P. Cosson. Role of the HIV-1 envelope transmembrane domain in intracellular sorting. *BMC Cell Biology*, 19(1):3, 2018.
- [139] L. O. Arthur, J. W. Bess, R. C. Sowder, R. E. Benveniste, D. L. Mann, J. C. Chermann and L. E. Henderson. Cellular proteins bound to immunodeficiency viruses: Implications for pathogenesis and vaccines. *Science*, 258(5090):1935–1938, 1992.

- [140] E. O. Freed and M. A. Martin. Virion incorporation of envelope glycoproteins with long but not short cytoplasmic tails is blocked by specific, single amino acid substitutions in the human immunodeficiency virus type 1 matrix. *Journal of Virology*, 69(3):1984–1989, 1995.
- [141] C. P. Hill, D. Worthylake, D. P. Bancroft, A. M. Christensen and W. I. Sundquist. Crystal structures of the trimeric human immunodeficiency virus type 1 matrix protein: Implications for membrane association and assembly. *Proceedings of the National Academy of Sciences of the United States of America*, 93(7):3099–3104, 1996.
- [142] K. Holm, K. Weclawicz, R. Hewson and M. Suomalainen. Human Immunodeficiency Virus Type 1 Assembly and Lipid Rafts: Pr55^{gag} Associates with Membrane Domains That Are Largely Resistant to Brij98 but Sensitive to Triton X-100. *Journal of Virology*, 77(8):4805–4817, 2003.
- [143] C. Berlioz-Torrent, B. L. Shacklett, L. Erdtmann, L. Delamarre, I. Bouchaert, P. Sonigo, M. C. Dokhelar and R. Benarous. Interactions of the Cytoplasmic Domains of Human and Simian Retroviral Transmembrane Proteins with Components of the Clathrin Adaptor Complexes Modulate Intracellular and Cell Surface Expression of Envelope Glycoproteins. *Journal of Virology*, 73(2):1350–1361, 1999.
- [144] J. Chen, O. Nikolaitchik, J. Singh, A. Wright, C. E. Bencsics, J. M. Coffin, N. Ni, S. Lockett, V. K. Pathak and W.-S. Hu. High efficiency of HIV-1 genomic RNA packaging and heterozygote formation revealed by single virion analysis. *Proceedings of the National Academy of Sciences of the United States of America*, 106(32):13535–13540, 2009.
- [145] L. VerPlank, F. Bouamr, T. J. LaGrassa, B. Agresta, A. Kikonyogo, J. Leis and C. A. Carter. Tsg101, a homologue of ubiquitin-conjugating (E2) enzymes, binds the L domain in HIV type 1 Pr55^{Gag}. *Proceedings of the National Academy of Sciences of the United States of America*, 98(14):7724–7729, 2001.
- [146] G. A. Frank, K. Narayan, J. W. Bess, G. Q. Del Prete, X. Wu, A. Moran, L. M. Hartnell, L. A. Earl, J. D. Lifson and S. Subramaniam. Maturation of the HIV-1 core by a non-diffusional phase transition. *Nature Communications*, 6:5854, 2015.
- [147] K. Murata and M. Wolf. Cryo-electron microscopy for structural analysis of dynamic biological macromolecules. *Biochimica et Biophysica Acta (BBA) – General Subjects*, 1862(2):324–334, 2018.
- [148] E. H. Lee, J. Hsin, M. Sotomayor, G. Comellas and K. Schulten. Discovery Through the Computational Microscope. *Structure*, 17(10):1295–1306, 2009.
- [149] R. O. Dror, R. M. Dirks, J. Grossman, H. Xu and D. E. Shaw. Biomolecular Simulation: A Computational Microscope for Molecular Biology. *Annual Review of Biophysics*, 41(1):429–452, 2012.

- [150] S. W. de Leeuw, J. W. Perram, E. R. Smith and J. S. Rowlinson. Simulation of electrostatic systems in periodic boundary conditions. I. Lattice sums and dielectric constants. *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences*, 373(1752):27–56, 1980.
- [151] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola and J. R. Haak. Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics*, 81(8):3684–3690, 1984.
- [152] S. Nosé. A molecular dynamics method for simulations in the canonical ensemble. *Molecular Physics*, 52(2):255–268, 1984.
- [153] M. Parrinello and A. Rahman. Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied Physics*, 52(12):7182–7190, 1981.
- [154] J. Jacquemyn, A. Cascalho and R. E. Goodchild. The ins and outs of endoplasmic reticulum-controlled lipid biosynthesis. *EMBO Reports*, 18(11):1905–1921, 2017.
- [155] A. I. P. M. de Kroon, P. J. Rijken and C. H. De Smet. Checks and balances in membrane phospholipid class and acyl chain homeostasis, the yeast perspective. *Progress in Lipid Research*, 52(4):374–394, 2013.
- [156] R. M. Bell, L. M. Ballas and R. A. Coleman. Lipid topogenesis. *Journal of Lipid Research*, 22(3):391–403, 1981.
- [157] H. F. Hofbauer, M. Gecht, S. C. Fischer, A. Seybert, A. S. Frangakis, E. H. Stelzer, R. Covino, G. Hummer and R. Ernst. The molecular recognition of phosphatidic acid by an amphipathic helix in Opi1. *The Journal of Cell Biology*, 217(9):3109–3126, 2018.
- [158] J. Kunz, M. P. Wilson, M. Kisseleva, J. H. Hurley, P. W. Majerus and R. A. Anderson. The activation loop of phosphatidylinositol phosphate kinases determines signaling specificity. *Molecular Cell*, 5(1):1–11, 2000.
- [159] G. D. Fairn, K. Ogata, R. J. Botelho, P. D. Stahl, R. A. Anderson, P. De Camilli, T. Meyer, S. Wodak and S. Grinstein. An electrostatic switch displaces phosphatidylinositol phosphate kinases from the membrane during phagocytosis. *The Journal of Cell Biology*, 187(5):701–714, 2009.
- [160] J. Hu, Q. Yuan, X. Kang, Y. Qin, L. Li, Y. Ha and D. Wu. Resolution of structure of PIP5K1A reveals molecular mechanism for its regulation by dimerization and dishevelled. *Nature Communications*, 6:8205, 2015.
- [161] A. Liu, D. Sui, D. Wu and J. Hu. The activation loop of PIP5K functions as a membrane sensor essential for lipid substrate processing. *Science Advances*, 2(11):e1600925, 2016.

- [162] E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng and T. E. Ferrin. UCSF Chimera—A visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, 25(13):1605–1612, 2004.
- [163] D. H. de Jong, G. Singh, W. F. D. Bennett, C. Arnarez, T. A. Wassenaar, L. V. Schäfer, X. Periole, D. P. Tieleman and S. J. Marrink. Improved Parameters for the Martini Coarse-Grained Protein Force Field. *Journal of Chemical Theory and Computation*, 9(1):687–697, 2013.
- [164] T. A. Wassenaar, H. I. Ingólfsson, R. A. Böckmann, D. P. Tieleman and S. J. Marrink. Computational Lipidomics with insane: A Versatile Tool for Generating Custom Membranes for Molecular Simulations. *Journal of Chemical Theory and Computation*, 11(5):2144–2155, 2015.
- [165] M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess and E. Lindahl. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, 1-2:19–25, 2015.
- [166] S. J. Marrink, A. H. de Vries and A. E. Mark. Coarse Grained Model for Semiquantitative Lipid Simulations. *The Journal of Physical Chemistry B*, 108(2):750–760, 2004.
- [167] S. J. Marrink, H. J. Risselada, S. Yefimov, D. P. Tieleman and A. H. de Vries. The MARTINI Force Field: Coarse Grained Model for Biomolecular Simulations. *The Journal of Physical Chemistry B*, 111(27):7812–7824, 2007.
- [168] L. Monticelli, S. K. Kandasamy, X. Periole, R. G. Larson, D. P. Tieleman and S.-J. Marrink. The MARTINI Coarse-Grained Force Field: Extension to Proteins. *Journal of Chemical Theory and Computation*, 4(5):819–834, 2008.
- [169] F. A. Herzog, L. Braun, I. Schoen and V. Vogel. Improved side chain dynamics in MARTINI simulations of protein-lipid interfaces. *Journal of Chemical Theory and Computation*, 12(5):2246–2458, 2016.
- [170] G. Bussi, D. Donadio and M. Parrinello. Canonical sampling through velocity rescaling. *The Journal of Chemical Physics*, 126(1):014101, 2007.
- [171] M. Parrinello and A. Rahman. Crystal Structure and Pair Potentials: A Molecular-Dynamics Study. *Physical Review Letters*, 45(14):1196–1199, 1980.
- [172] S. Jo, T. Kim, V. G. Iyer and W. Im. CHARMM-GUI: A web-based graphical user interface for CHARMM. *Journal of Computational Chemistry*, 29(11):1859–1865, 2008.
- [173] T. A. Wassenaar, K. Pluhackova, R. A. Böckmann, S. J. Marrink and D. P. Tieleman. Going Backward: A Flexible Geometric Approach to Reverse Transformation from Coarse Grained to Atomistic Models. *Journal of Chemical Theory and Computation*, 10(2):676–690, 2014.

- [174] J. Huang and A. D. MacKerell. CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data. *Journal of Computational Chemistry*, 34(25):2135–2145, 2013.
- [175] S. Jo, T. Kim and W. Im. Automated Builder and Database of Protein/Membrane Complexes for Molecular Dynamics Simulations. *PLOS ONE*, 2(9):e880, 2007.
- [176] S. Jo, J. B. Lim, J. B. Klauda and W. Im. CHARMM-GUI Membrane Builder for Mixed Bilayers and Its Application to Yeast Membranes. *Biophysical Journal*, 97(1):50–58, 2009.
- [177] E. L. Wu, X. Cheng, S. Jo, H. Rui, K. C. Song, E. M. Dávila-Contreras, Y. Qi, J. Lee, V. Monje-Galvan, R. M. Venable, J. B. Klauda and W. Im. CHARMM-GUI Membrane Builder toward realistic biological membrane simulations. *Journal of Computational Chemistry*, 35(27):1997–2004, 2014.
- [178] J. Lee, X. Cheng, J. M. Swails, M. S. Yeom, P. K. Eastman, J. A. Lemkul, S. Wei, J. Buckner, J. C. Jeong, Y. Qi, S. Jo, V. S. Pande, D. A. Case, C. L. Brooks, A. D. MacKerell, J. B. Klauda and W. Im. CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM / OpenMM Simulations Using the CHARMM36 Additive Force Field. *Journal of Chemical Theory and Computation*, 12(1):405–413, 2016.
- [179] J. Huang, S. Rauscher, G. Nawrocki, T. Ran, M. Feig, B. L. de Groot, H. Grubmüller and A. D. MacKerell Jr. CHARMM36m: An improved force field for folded and intrinsically disordered proteins. *Nature Methods*, 14(1):71–73, 2017.
- [180] N. Michaud-Agrawal, E. J. Denning, T. B. Woolf and O. Beckstein. MDAnalysis: A toolkit for the analysis of molecular dynamics simulations. *Journal of Computational Chemistry*, 32(10):2319–2327, 2011.
- [181] R. J. Gowers, M. Linke, J. Barnoud, T. J. E. Reddy, M. N. Melo, S. L. Seyler, J. Domański, D. L. Dotson, S. Buchoux, I. M. Kenney and O. Beckstein. MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations. In S. Benthall and S. Rostrup, editors, *Proceedings of the 15th Python in Science Conference*, pages 98–105, 2016.
- [182] M. Gecht. *Biophysical Analysis of a Membrane Sensing Amphipathic Helix in Vitro and in Silico*. Master thesis, Fachbereich Biochemie, Chemie, Pharmazie, Johann Wolfgang Goethe-Universität Frankfurt, Frankfurt am Main, 2017.
- [183] T. Nishimura, M. Gecht, R. Covino, G. Hummer, M. A. Surma, C. Klose, H. Arai, N. Kono and C. J. Stefan. Osh Proteins Control Nanoscale Lipid Organization Necessary for PI(4,5)P₂ Synthesis. *Molecular Cell*, 2019.
- [184] G. M. Carman and G.-S. Han. Regulation of Phospholipid Synthesis in the Yeast *Saccharomyces cerevisiae*. *Annual Review of Biochemistry*, 80:859–883, 2011.

- [185] J. Kunz, A. Fuelling, L. Kolbe and R. A. Anderson. Stereo-specific substrate recognition by phosphatidylinositol phosphate kinases is swapped by changing a single amino acid residue. *The Journal of Biological Chemistry*, 277(7):5611–5619, 2002.
- [186] V. Corradi, E. Mendez-Villuendas, H. I. Ingólfsson, R.-X. Gu, I. Siuda, M. N. Melo, A. Moussatova, L. J. DeGagné, B. I. Sejdiu, G. Singh, T. A. Wassenaar, K. Delgado Magnero, S. J. Marrink and D. P. Tieleman. Lipid–Protein Interactions Are Unique Fingerprints for Membrane Proteins. *ACS Central Science*, 4(6):709–717, 2018.
- [187] V. Corradi, B. I. Sejdiu, H. Mesa-Galoso, H. Abdizadeh, S. Y. Noskov, S. J. Marrink and D. P. Tieleman. Emerging Diversity in Lipid–Protein Interactions. *Chemical Reviews*, 119(9):5775–5848, 2019.
- [188] R. B. Cornell and S. G. Taneva. Amphipathic helices as mediators of the membrane interaction of amphitropic proteins, and as modulators of bilayer physical properties. *Current Protein & Peptide Science*, 7(6):539–552, 2006.
- [189] R. B. Best, X. Zhu, J. Shim, P. E. M. Lopes, J. Mittal, M. Feig and A. D. MacKerell. Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone ϕ , ψ and Side-Chain χ_1 and χ_2 Dihedral Angles. *Journal of Chemical Theory and Computation*, 8(9):3257–3273, 2012.
- [190] D. Frishman and P. Argos. Knowledge-based protein secondary structure assignment. *Proteins*, 23(4):566–579, 1995.
- [191] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot and É. Duchesnay. Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, 12(85):2825–2830, 2011.
- [192] S. F. Altschul, T. L. Madden, A. A. Schäffer, J. Zhang, Z. Zhang, W. Miller and D. J. Lipman. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Research*, 25(17):3389–3402, 1997.
- [193] S. Henikoff and J. G. Henikoff. Amino acid substitution matrices from protein blocks. *Proceedings of the National Academy of Sciences of the United States of America*, 89(22):10915–10919, 1992.
- [194] J. S. Papadopoulos and R. Agarwala. COBALT: Constraint-based alignment tool for multiple protein sequences. *Bioinformatics*, 23(9):1073–1079, 2007.
- [195] G. E. Crooks, G. Hon, J.-M. Chandonia and S. E. Brenner. WebLogo: A Sequence Logo Generator. *Genome Research*, 14(6):1188–1190, 2004.
- [196] M. Marchi and P. Ballone. Adiabatic bias molecular dynamics: A method to navigate the conformational space of complex molecular systems. *The Journal of Chemical Physics*, 110(8):3697–3702, 1999.

- [197] G. A. Tribello, M. Bonomi, D. Branduardi, C. Camilloni and G. Bussi. PLUMED 2: New feathers for an old bird. *Computer Physics Communications*, 185(2):604–613, 2014.
- [198] H. Grubmüller, B. Heymann and P. Tavan. Ligand binding: Molecular mechanics calculation of the streptavidin-biotin rupture force. *Science*, 271(5251):997–999, 1996.
- [199] C. Klose, M. A. Surma and K. Simons. Organellar lipidomics — background and perspectives. *Current Opinion in Cell Biology*, 25(4):406–413, 2013.
- [200] C. Nick Pace and J. Martin Scholtz. A Helix Propensity Scale Based on Experimental Studies of Peptides and Proteins. *Biophysical Journal*, 75(1):422–427, 1998.
- [201] F. Pietrucci and A. Laio. A Collective Variable for the Efficient Exploration of Protein Beta-Sheet Structures: Application to SH3 and GB1. *Journal of Chemical Theory and Computation*, 5(9):2197–2201, 2009.
- [202] A. Acharyya, Y. Ge, H. Wu, W. F. DeGrado, V. A. Voelz and F. Gai. Exposing the Nucleation Site in α -Helix Folding: A Joint Experimental and Simulation Study. *The Journal of Physical Chemistry B*, 123(8):1797–1807, 2019.
- [203] N. Gopaldass, B. Fauvet, H. Lashuel, A. Roux and A. Mayer. Membrane scission driven by the PROPPIN Atg18. *The EMBO Journal*, 36(22):3274–3291, 2017.
- [204] H. Vogel. Incorporation of Melittin into phosphatidylcholine bilayers. *FEBS Letters*, 134(1):37–42, 1981.
- [205] E. Kuchinka and J. Seelig. Interaction of melittin with phosphatidylcholine membranes. Binding isotherm and lipid head-group conformation. *Biochemistry*, 28(10):4216–4221, 1989.
- [206] G. Beschiaschvili and H.-D. Baeuerle. Effective charge of melittin upon interaction with POPC vesicles. *Biochimica et Biophysica Acta (BBA) – Biomembranes*, 1068(2):195–200, 1991.
- [207] A. S. Ladokhin and S. H. White. Folding of amphipathic α -helices on membranes: Energetics of helix formation by melittin. *Journal of Molecular Biology*, 285(4):1363–1369, 1999.
- [208] L. Zhang, M. Rajendram, D. B. Weibel, A. Yethiraj and Q. Cui. Ionic Hydrogen Bonds and Lipid Packing Defects Determine the Binding Orientation and Insertion Depth of RecA on Multicomponent Lipid Bilayers. *The Journal of Physical Chemistry B*, 120(33):8424–8437, 2016.
- [209] C. L. Jackson, L. Walch and J.-M. Verbavatz. Lipids and Their Trafficking: An Integral Part of Cellular Organization. *Developmental Cell*, 39(2):139–153, 2016.
- [210] N. S. Hatzakis, V. K. Bhatia, J. Larsen, K. L. Madsen, P.-Y. Bolinger, A. H. Kundling, J. Castillo, U. Gether, P. Hedegård and D. Stamou. How curved membranes

- recruit amphipathic helices and protein anchoring motifs. *Nature Chemical Biology*, 5(11):835–841, 2009.
- [211] E. E. Kooijman, D. P. Tieleman, C. Testerink, T. Munnik, D. T. S. Rijkers, K. N. J. Burger and B. de Kruijff. An electrostatic/hydrogen bond switch as the basis for the specific interaction of phosphatidic acid with proteins. *The Journal of Biological Chemistry*, 282(15):11356–11364, 2007.
- [212] M. B. Ulmschneider, J. P. F. Doux, J. A. Killian, J. C. Smith and J. P. Ulmschneider. Mechanism and Kinetics of Peptide Partitioning into Membranes from All-Atom Simulations of Thermostable Peptides. *Journal of the American Chemical Society*, 132(10):3452–3460, 2010.
- [213] C. H. Chen, G. Wiedman, A. Khan and M. B. Ulmschneider. Absorption and folding of melittin onto lipid bilayer membranes via unbiased atomic detail microsecond molecular dynamics simulation. *Biochimica et Biophysica Acta (BBA) – Biomembranes*, 1838(9):2243–2249, 2014.
- [214] S. J. Irudayam and M. L. Berkowitz. Binding and reorientation of melittin in a POPC bilayer: Computer simulations. *Biochimica et Biophysica Acta (BBA) – Biomembranes*, 1818(12):2975–2981, 2012.
- [215] W. Im and C. L. Brooks. Interfacial folding and membrane insertion of designed peptides studied by molecular dynamics simulations. *Proceedings of the National Academy of Sciences of the United States of America*, 102(19):6771–6776, 2005.
- [216] UNAIDS Data 2020, 2020.
- [217] J. Lisziewicz, N. Bakare and F. Lori. Therapeutic vaccination for future management of HIV/AIDS. *Vaccine*, 21(7):620–623, 2003.
- [218] J. Lisziewicz, J. Trocio, L. Whitman, G. Varga, J. Xu, N. Bakare, P. Erbacher, C. Fox, R. Woodward, P. Markham, S. Arya, J.-P. Behr and F. Lori. DermaVir: A Novel Topical Vaccine for HIV/AIDS. *Journal of Investigative Dermatology*, 124(1):160–169, 2005.
- [219] Y. Z. Cohen and R. Dolin. Novel HIV vaccine strategies: Overview and perspective. *Therapeutic Advances in Vaccines*, 1(3):99–112, 2013.
- [220] T. W. Chun, L. Carruth, D. Finzi, X. Shen, J. A. DiGiuseppe, H. Taylor, M. Hermandkova, K. Chadwick, J. Margolick, T. C. Quinn, Y. H. Kuo, R. Brookmeyer, M. A. Zeiger, P. Barditch-Crovo and R. F. Siliciano. Quantification of latent tissue reservoirs and total body viral load in HIV-1 infection. *Nature*, 387(6629):183–188, 1997.
- [221] J. C. Grivel, M. L. Penn, D. A. Eckstein, B. Schramm, R. F. Speck, N. W. Abbey, B. Herndier, L. Margolis and M. A. Goldsmith. Human immunodeficiency virus type 1 coreceptor preferences determine target T-cell depletion and cellular tropism in human lymphoid tissue. *Journal of Virology*, 74(11):5347–5351, 2000.

- [222] J. M. Brenchley, T. W. Schacker, L. E. Ruff, D. A. Price, J. H. Taylor, G. J. Beilman, P. L. Nguyen, A. Khoruts, M. Larson, A. T. Haase and D. C. Douek. CD4⁺ T cell depletion during all stages of HIV disease occurs predominantly in the gastrointestinal tract. *The Journal of Experimental Medicine*, 200(6):749–759, 2004.
- [223] L. J. Picker, S. I. Hagen, R. Lum, E. F. Reed-Inderbitzin, L. M. Daly, A. W. Sylwester, J. M. Walker, D. C. Siess, M. Piatak, C. Wang, D. B. Allison, V. C. Maino, J. D. Lifson, T. Kodama and M. K. Axthelm. Insufficient production and tissue delivery of CD4⁺ memory T cells in rapidly progressive simian immunodeficiency virus infection. *The Journal of Experimental Medicine*, 200(10):1299–1314, 2004.
- [224] C. B. Wilen, J. C. Tilton and R. W. Doms. Molecular Mechanisms of HIV Entry. In *Viral Molecular Machines*, Advances in Experimental Medicine and Biology, pages 223–242. Springer US, Boston, MA, 2012.
- [225] S. C. Harrison. Viral membrane fusion. *Virology*, 479-480:498–507, 2015.
- [226] R. Wyatt and J. Sodroski. The HIV-1 Envelope Glycoproteins: Fusogens, Antigens, and Immunogens. *Science*, 280(5371):1884–1888, 1998.
- [227] M. Montero, N. E. van Houten, X. Wang and J. K. Scott. The Membrane-Proximal External Region of the Human Immunodeficiency Virus Type 1 Envelope: Dominant Site of Antibody Neutralization and Target for Vaccine Design. *Microbiology and Molecular Biology Reviews*, 72(1):54–84, 2008.
- [228] L. E. McCoy and D. R. Burton. Identification and specificity of broadly neutralizing antibodies against HIV. *Immunological Reviews*, 275(1):11–20, 2017.
- [229] N. Cerutti, J. L. Loredó-Varela, C. Caillat and W. Weissenhorn. Antip41 membrane proximal external region antibodies and the art of using the membrane for neutralization. *Current Opinion in HIV and AIDS*, 12(3):250–256, 2017.
- [230] J.-P. Julien, A. Cupo, D. Sok, R. L. Stanfield, D. Lyumkis, M. C. Deller, P.-J. Klasse, D. R. Burton, R. W. Sanders, J. P. Moore, A. B. Ward and I. A. Wilson. Crystal Structure of a Soluble Cleaved HIV-1 Envelope Trimer. *Science*, 342(6165):1477–1483, 2013.
- [231] M. Pancera, T. Zhou, A. Druz, I. S. Georgiev, C. Soto, J. Gorman, J. Huang, P. Acharya, G.-Y. Chuang, G. Ofek, G. B. E. Stewart-Jones, J. Stuckey, R. T. Bailer, M. G. Joyce, M. K. Louder, N. Tumba, Y. Yang, B. Zhang, M. S. Cohen, B. F. Haynes, J. R. Mascola, L. Morris, J. B. Munro, S. C. Blanchard, W. Mothes, M. Connors and P. D. Kwong. Structure and immune recognition of trimeric pre-fusion HIV-1 Env. *Nature*, 514(7523):455–461, 2014.
- [232] D. C. Chan, D. Fass, J. M. Berger and P. S. Kim. Core Structure of gp41 from the HIV Envelope Glycoprotein. *Cell*, 89(2):263–273, 1997.

- [233] W. Weissenhorn, A. Dessen, S. C. Harrison, J. J. Skehel and D. C. Wiley. Atomic structure of the ectodomain from HIV-1 gp41. *Nature*, 387(6631):426–430, 1997.
- [234] M. Caffrey. Three-dimensional solution structure of the 44 kDa ectodomain of SIV gp41. *EMBO Journal*, 17(16):4572–4584, 1998.
- [235] V. Buzon, G. Natrajan, D. Schibli, F. Campelo, M. M. Kozlov and W. Weissenhorn. Crystal Structure of HIV-1 gp41 Including Both Fusion Peptide and Membrane Proximal External Regions. *PLOS Pathogens*, 6(5):e1000880, 2010.
- [236] J. Dev, D. Park, Q. Fu, J. Chen, H. J. Ha, F. Ghantous, T. Herrmann, W. Chang, Z. Liu, G. Frey, M. S. Seaman, B. Chen and J. J. Chou. Structural basis for membrane anchoring of HIV-1 envelope spike. *Science*, 353(6295):172–175, 2016.
- [237] S. C. Chiliveri, J. M. Louis, R. Ghirlando, J. L. Baber and A. Bax. Tilted, Uninterrupted, Monomeric HIV-1 gp41 Transmembrane Helix from Residual Dipolar Couplings. *Journal of the American Chemical Society*, 140(1):34–37, 2018.
- [238] B. Kwon, M. Lee, A. J. Waring and M. Hong. Oligomeric Structure and Three-Dimensional Fold of the HIV gp41 Membrane-Proximal External Region and Transmembrane Domain in Phospholipid Bilayers. *Journal of the American Chemical Society*, 140(26):8246–8259, 2018.
- [239] Q. Fu, M. M. Shaik, Y. Cai, F. Ghantous, A. Piai, H. Peng, S. Rits-Volloch, Z. Liu, S. C. Harrison, M. S. Seaman, B. Chen and J. J. Chou. Structure of the membrane proximal external region of HIV-1 envelope glycoprotein. *Proceedings of the National Academy of Sciences of the United States of America*, 115(38):E8892–E8899, 2018.
- [240] L. Shang, L. Yue and E. Hunter. Role of the Membrane-Spanning Domain of Human Immunodeficiency Virus Type 1 Envelope Glycoprotein in Cell-Cell Fusion and Virus Infection. *Journal of Virology*, 82(11):5417–5428, 2008.
- [241] Y. Long, F. Meng, N. Kondo, A. Iwamoto and Z. Matsuda. Conserved arginine residue in the membrane-spanning domain of HIV-1 gp41 is required for efficient membrane fusion. *Protein & Cell*, 2(5):369–376, 2011.
- [242] V. Oakes, J. Torralba, E. Rujas, J. L. Nieva, C. Domene and B. Apellániz. Exposure of the HIV-1 broadly neutralizing antibody 10E8 MPER epitope on the membrane surface by gp41 transmembrane domain scaffolds. *Biochimica et Biophysica Acta (BBA) – Biomembranes*, 1860(6):1259–1271, 2018.
- [243] L. R. Hollingsworth, J. A. Lemkul, D. R. Bevan and A. M. Brown. HIV-1 Env gp41 Transmembrane Domain Dynamics Are Modulated by Lipid, Water, and Ion Interactions. *Biophysical Journal*, 115(1):84–94, 2018.
- [244] M. Choisy, C. H. Woelk, J.-F. Guegan and D. L. Robertson. Comparative Study of Adaptive Molecular Evolution in Different Human Immunodeficiency Virus Groups and Subtypes. *Journal of Virology*, 78(4):1962–1970, 2004.

- [245] N. Vincent, C. Genin and E. Malvoisin. Identification of a conserved domain of the HIV-1 transmembrane protein gp41 which interacts with cholesteryl groups. *Biochimica et Biophysica Acta (BBA) – Biomembranes*, 1567:157–164, 2002.
- [246] B. Apellániz, E. Rujas, P. Carravilla, J. Requejo-Isidro, N. Huarte, C. Domene and J. L. Nieva. Cholesterol-dependent membrane fusion induced by the gp41 membrane-proximal external region-transmembrane domain connection suggests a mechanism for broad HIV-1 neutralization. *Journal of Virology*, 88(22):13367–13377, 2014.
- [247] M. Lorizate, I. de la Arada, N. Huarte, S. Sánchez-Martínez, B. G. de la Torre, D. Andreu, J. L. R. Arrondo and J. L. Nieva. Structural Analysis and Assembly of the HIV-1 Gp41 Amino-Terminal Fusion Peptide and the Pretransmembrane Amphipathic-At-Interface Sequence. *Biochemistry*, 45(48):14337–14346, 2006.
- [248] Z.-Y. J. Sun, K. J. Oh, M. Kim, J. Yu, V. Brusica, L. Song, Z. Qiao, J.-h. Wang, G. Wagner and E. L. Reinherz. HIV-1 Broadly Neutralizing Antibody Extracts Its Epitope from a Kinked gp41 Ectodomain Region on the Viral Membrane. *Immunity*, 28(1):52–63, 2008.
- [249] B. T. Korber, B. T. Foley, C. L. Kuiken, S. K. Pillai and J. G. Sodroski. Numbering positions in HIV relative to HXB2CG. *AIDS Research and Human Retroviruses*, 3:102–111, 1998.
- [250] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics*, 79(2):926–935, 1983.
- [251] M. Lorizate, T. Sachsenheimer, B. Glass, A. Habermann, M. J. Gerl, H.-G. Kräuslich and B. Brügger. Comparative lipidomics analysis of HIV-1 particles and their producer cell membrane in different cell lines. *Cellular Microbiology*, 15(2):292–304, 2013.
- [252] M. V. Shapovalov and R. L. Dunbrack. A Smoothed Backbone-Dependent Rotamer Library for Proteins Derived from Adaptive Kernel Density Estimates and Regressions. *Structure*, 19(6):844–858, 2011.
- [253] F. Garces, J. H. Lee, N. de Val, A. Torrents de la Pena, L. Kong, C. Puchades, Y. Hua, R. L. Stanfield, D. R. Burton, J. P. Moore, R. W. Sanders, A. B. Ward and I. A. Wilson. Affinity Maturation of a Potent Family of HIV Antibodies Is Primarily Focused on Accommodating or Avoiding Glycans. *Immunity*, 43(6):1053–1063, 2015.
- [254] A. Fiser and A. Sali. Modeller: Generation and refinement of homology-based protein structure models. *Methods in Enzymology*, 374:461–491, 2003.
- [255] W. G. Hoover, A. J. C. Ladd and B. Moran. High-Strain-Rate Plastic Flow Studied via Nonequilibrium Molecular Dynamics. *Physical Review Letters*, 48(26):1818–1820, 1982.

- [256] D. J. Evans and B. L. Holian. The Nose–Hoover thermostat. *The Journal of Chemical Physics*, 83(8):4069–4074, 1985.
- [257] J. Kyte and R. F. Doolittle. A simple method for displaying the hydropathic character of a protein. *Journal of Molecular Biology*, 157(1):105–132, 1982.
- [258] R. Gautier, D. Douguet, B. Antony and G. Drin. HELIQUEST: A web server to screen sequences with specific α -helical properties. *Bioinformatics*, 24(18):2101–2102, 2008.
- [259] P. Liu, D. K. Agrafiotis and D. L. Theobald. Fast determination of the optimal rotational matrix for macromolecular superpositions. *Journal of Computational Chemistry*, 31(7):1561–1563, 2010.
- [260] A. Irimia, A. Sarkar, R. L. Stanfield and I. A. Wilson. Crystallographic Identification of Lipid as an Integral Component of the Epitope of HIV Broadly Neutralizing Antibody 4E10. *Immunity*, 44(1):21–31, 2016.
- [261] A. Irimia, A. M. Serra, A. Sarkar, R. Jacak, O. Kalyuzhnyi, D. Sok, K. L. Saye-Francisco, T. Schiffner, R. Tingle, M. Kubitz, Y. Adachi, R. L. Stanfield, M. C. Deller, D. R. Burton, W. R. Schief and I. A. Wilson. Lipid interactions and angle of approach to the HIV-1 viral membrane of broadly neutralizing antibody 10E8: Insights for vaccine and therapeutic design. *PLOS Pathogens*, 13(2):e1006212, 2017.
- [262] J. Huang, G. Ofek, L. Laub, M. K. Louder, N. A. Doria-Rose, N. S. Longo, H. Imamichi, R. T. Bailer, B. Chakrabarti, S. K. Sharma, S. M. Alam, T. Wang, Y. Yang, B. Zhang, S. A. Migueles, R. Wyatt, B. F. Haynes, P. D. Kwong, J. R. Mascola and M. Connors. Broad and potent neutralization of HIV-1 by a gp41-specific human antibody. *Nature*, 491(7424):406–412, 2012.
- [263] E. A. Kabat, T. T. Wu, H. M. Perry, C. Foeller and K. S. Gottesman. *Sequences of Proteins of Immunological Interest*. DIANE Publishing, 1992.
- [264] T. T. Pham and M. R. Shirts. Identifying low variance pathways for free energy calculations of molecular transformations in solution phase. *The Journal of Chemical Physics*, 135(3):034114, 2011.
- [265] T. T. Pham and M. R. Shirts. Optimal pairwise and non-pairwise alchemical pathways for free energy calculations of molecular transformation in solution phase. *The Journal of Chemical Physics*, 136(12):124120, 2012.
- [266] R. M. F. Cardoso, M. B. Zwick, R. L. Stanfield, R. Kunert, J. M. Binley, H. Katinger, D. R. Burton and I. A. Wilson. Broadly Neutralizing Anti-HIV Antibody 4E10 Recognizes a Helical Conformation of a Highly Conserved Fusion-Associated Motif in gp41. *Immunity*, 22(2):163–173, 2005.
- [267] A. B. Ward and I. A. Wilson. The HIV-1 envelope glycoprotein structure: Nailing down a moving target. *Immunological Reviews*, 275(1):21–32, 2017.

- [268] A. Piai, J. Dev, Q. Fu and J. J. Chou. Stability and Water Accessibility of the Trimeric Membrane Anchors of the HIV-1 Envelope Spikes. *Journal of the American Chemical Society*, 139(51):18432–18435, 2017.
- [269] P. N. Reardon, H. Sage, S. M. Dennison, J. W. Martin, B. R. Donald, S. M. Alam, B. F. Haynes and L. D. Spicer. Structure of an HIV-1-neutralizing antibody target, the lipid-bound gp41 envelope membrane proximal region trimer. *Proceedings of the National Academy of Sciences of the United States of America*, 111(4):1391–1396, 2014.
- [270] Z. Li, W. Li, M. Lu, J. Bess, C. W. Chao, J. Gorman, D. S. Terry, B. Zhang, T. Zhou, S. C. Blanchard, P. D. Kwong, J. D. Lifson, W. Mothes and J. Liu. Subnanometer structures of HIV-1 envelope trimers on aldrithiol-2-inactivated virus particles. *Nature Structural & Molecular Biology*, 27(8):726–734, 2020.
- [271] V. K. Gangupomu and C. F. Abrams. All-atom models of the membrane-spanning domain of HIV-1 gp41 from metadynamics. *Biophysical Journal*, 99(10):3438–3444, 2010.
- [272] M. K. Baker, V. K. Gangupomu and C. F. Abrams. Characterization of the water defect at the HIV-1 gp41 membrane spanning domain in bilayers with and without cholesterol using molecular simulations. *Biochimica et Biophysica Acta (BBA) – Biomembranes*, 1838(5):1396–1405, 2014.
- [273] X. Wei, J. M. Decker, S. Wang, H. Hui, J. C. Kappes, X. Wu, J. F. Salazar-Gonzalez, M. G. Salazar, J. M. Kilby, M. S. Saag, N. L. Komarova, M. A. Nowak, B. H. Hahn, P. D. Kwong and G. M. Shaw. Antibody neutralization and escape by HIV-1. *Nature*, 422(6929):307–312, 2003.
- [274] C. N. Scanlan, J. Offer, N. Zitzmann and R. A. Dwek. Exploiting the defensive sugars of HIV-1 for drug and vaccine design. *Nature*, 446(7139):1038–1045, 2007.
- [275] M. Crispin and K. J. Doores. Targeting host-derived glycans on enveloped viruses for antibody-based vaccine design. *Current Opinion in Virology*, 11:63–69, 2015.
- [276] K. J. Doores. The HIV glycan shield as a target for broadly neutralizing antibodies. *The FEBS Journal*, 282(24):4679–4691, 2015.
- [277] M. J. Harvey, G. Giupponi and G. D. Fabritiis. ACEMD: Accelerating Biomolecular Dynamics in the Microsecond Time Scale. *Journal of Chemical Theory and Computation*, 5(6):1632–1639, 2009.
- [278] R. Salomon-Ferrer, A. W. Götz, D. Poole, S. Le Grand and R. C. Walker. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. *Journal of Chemical Theory and Computation*, 9(9):3878–3888, 2013.

- [279] B. R. Brooks, C. L. Brooks, A. D. Mackerell, L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caffisch, L. Caves, Q. Cui, A. R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R. W. Pastor, C. B. Post, J. Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D. M. York and M. Karplus. CHARMM: The biomolecular simulation program. *Journal of Computational Chemistry*, 30(10):1545–1614, 2009.
- [280] K. J. Bowers, D. E. Chow, H. Xu, R. O. Dror, M. P. Eastwood, B. A. Gregersen, J. L. Klepeis, I. Kolossvary, M. A. Moraes, F. D. Sacerdoti, J. K. Salmon, Y. Shan and D. E. Shaw. Scalable Algorithms for Molecular Dynamics Simulations on Commodity Clusters. In *SC '06: Proceedings of the 2006 ACM/IEEE Conference on Supercomputing*, pages 43–43, 2006.
- [281] W. M. Brown, A. Kohlmeyer, S. J. Plimpton and A. N. Tharrington. Implementing molecular dynamics on hybrid high performance computers – Particle–particle particle–mesh. *Computer Physics Communications*, 183(3):449–459, 2012.
- [282] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kalé and K. Schulten. Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry*, 26(16):1781–1802, 2005.
- [283] P. Eastman, M. S. Friedrichs, J. D. Chodera, R. J. Radmer, C. M. Bruns, J. P. Ku, K. A. Beauchamp, T. J. Lane, L.-P. Wang, D. Shukla, T. Tye, M. Houston, T. Stich, C. Klein, M. R. Shirts and V. S. Pande. OpenMM 4: A Reusable, Extensible, Hardware Independent Library for High Performance Molecular Simulation. *Journal of Chemical Theory and Computation*, 9(1):461–469, 2013.
- [284] C. Kutzner, S. Páll, M. Fechner, A. Esztermann, B. L. de Groot and H. Grubmüller. Best bang for your buck: GPU nodes for GROMACS biomolecular simulations. *Journal of Computational Chemistry*, 36(26):1990–2008, 2015.
- [285] C. Kutzner, S. Páll, M. Fechner, A. Esztermann, B. L. de Groot and H. Grubmüller. More bang for your buck: Improved use of GPU nodes for GROMACS 2018. *Journal of Computational Chemistry*, 40(27):2418–2431, 2019.
- [286] D. T. Marr, F. Binns, D. L. Hill, G. Hinton, D. A. Koufaty, J. A. Miller and M. Upton. Hyper-threading technology architecture and microarchitecture. *Intel Technology Journal*, 6(1), 2002.
- [287] B. Hess, C. Kutzner, D. van der Spoel and E. Lindahl. GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *Journal of Chemical Theory and Computation*, 4(3):435–447, 2008.
- [288] B. Turoňová, M. Sikora, C. Schürmann, W. J. H. Hagen, S. Welsch, F. E. C. Blanc, S. von Bülow, M. Gecht, K. Bagola, C. Hörner, G. van Zandbergen, J. Landry, N. T. D. de Azevedo, S. Mosalaganti, A. Schwarz, R. Covino, M. D. Mühlebach, G. Hummer,

- J. Krijnse Locker and M. Beck. In situ structural analysis of SARS-CoV-2 spike reveals flexibility mediated by three hinges. *Science*, 370(6513):203–208, 2020.
- [289] G. M. Amdahl. Validity of the single processor approach to achieving large scale computing capabilities. *SSCS*, 12(3):19–20, 2007.
- [290] J. L. Furlani. Modules: Providing a flexible user environment. In *Proceedings of the Fifth Large Installation Systems Administration Conference (LISA V)*, pages 141–152, San Diego, CA, 1991.
- [291] M. Gecht, M. Siggel, M. Linke, G. Hummer and J. Köfinger. MDBenchmark: A toolkit to optimize the performance of molecular dynamics simulations. *The Journal of Chemical Physics*, 153(14):144105, 2020.
- [292] D. L. Dotson, S. L. Seyler, M. Linke, R. J. Gowers and Oliver Beckstein. Datreant: Persistent, Pythonic trees for heterogeneous data. In S. Benthall and Scott Rostrup, editors, *Proceedings of the 15th Python in Science Conference*, pages 51–56, Austin, TX, 2016.
- [293] J. Wang, P. Cieplak and P. A. Kollman. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *Journal of Computational Chemistry*, 21(12):1049–1074, 2000.
- [294] V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg and C. Simmerling. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins*, 65(3):712–725, 2006.
- [295] K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, J. L. Klepeis, R. O. Dror and D. E. Shaw. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins*, 78(8):1950–1958, 2010.
- [296] R. B. Best and G. Hummer. Optimized molecular dynamics force fields applied to the helix-coil transition of polypeptides. *The Journal of Physical Chemistry B*, 113(26):9004–9015, 2009.
- [297] R. B. Best, D. de Sancho and J. Mittal. Residue-specific α -helix propensities from molecular simulation. *Biophysical Journal*, 102(6):1462–1467, 2012.
- [298] S. Piana, A. G. Donchev, P. Robustelli and D. E. Shaw. Water Dispersion Interactions Strongly Influence Simulated Structural Properties of Disordered Protein States. *The Journal of Physical Chemistry B*, 119(16):5113–5123, 2015.
- [299] S. von Bülow, M. Siggel, M. Linke and G. Hummer. Dynamic cluster formation determines viscosity and diffusion in dense protein solutions. *Proceedings of the National Academy of Sciences of the United States of America*, 116(20):9843–9852, 2019.

- [300] X. Wu, M. Siggel, S. Ovchinnikov, W. Mi, V. Svetlov, E. Nudler, M. Liao, G. Hummer and T. A. Rapoport. Structural basis of ER-associated protein degradation mediated by the Hrd1 ubiquitin ligase complex. *Science*, 368(6489), 2020.
- [301] S. Hofmann, D. Januliene, A. R. Mehdipour, C. Thomas, E. Stefan, S. Brüchert, B. T. Kuhn, E. R. Geertsma, G. Hummer and A. Moeller. Conformation space of a heterodimeric ABC exporter under turnover conditions. *Nature*, 571(7766):580–583, 2019.

Declaration of contributions

Except where stated otherwise by reference or acknowledgment, the work presented was generated by myself under the supervision of my advisors during my doctoral studies. All contributions from colleagues and collaborators are explicitly referenced in the thesis. The material listed below was obtained in the context of collaborative research:

Figure 3.13: Dr. Harald F. Hofbauer (Institute of Molecular Biosciences, University of Graz, Graz), planned and performed the experiments and recorded the results. I, Michael Gecht, arranged the figure presented in this thesis.

Equations 5.2, 5.3, 5.4 and 5.5 were derived by Dr. Jürgen Köfnger (Department of Theoretical Biophysics, Max Planck Institute of Biophysics, Frankfurt).

Whenever a figure, table or text is identical to a previous publication, it is stated explicitly in the thesis that copyright permission and/or co-author agreement has been obtained.

The following parts of the thesis have been previously published:

Chapter 2

H.F. Hofbauer, M. Gecht, S.C. Fischer, A. Seybert, A.S. Frangakis, E.H.K. Stelzer, R. Covino, G. Hummer, R. Ernst. The molecular recognition of phosphatidic acid by an amphipathic helix in Opi1. *J Cell Biol*, 217(9):3109–3126, 2018.

- Figures 2.1, 2.2, 2.3, 2.4, 2.5, 2.6, 2.7, 2.8, 2.9 and 2.10
- Table 2.1

T. Nishimura, M. Gecht, R. Covino, G. Hummer, M.A. Surma, C. Klose, H. Arai, N. Kono, C.J. Stefan. Osh Proteins Control Nanoscale Lipid Organization Necessary for PI(4,5)P2 Synthesis. *Mol Cell*, 75(5):1043-1057.e8, 2019.

- Figures 2.11, 2.12 and 2.13
- Table 2.2

Chapter 5

M. Gecht, M. Siggel, M. Linke, G. Hummer, J. Köfnger. MDBenchmark: A toolkit to optimize the performance of molecular dynamics simulations. *J Chem Phys*, 153(14):144105, 2020.

- Figures 5.1, 5.2, 5.3, 5.4, 5.5, 5.6, 5.7, 5.8, 5.9 and 5.10
- Table 5.1
- Listings 5.1 and 5.2