

**Cell-free expression and molecular modeling  
of the  $\gamma$ -secretase complex and G-protein-coupled receptors**

**Dissertation  
Zur Erlangung des Doktorgrades  
der Naturwissenschaften**

**Vorlegt beim Fachbereich 14  
Biochemie, Chemie und Pharmazie  
der Johann Wolfgang Goethe Universität  
in Frankfurt am Main**

**von  
Umesh GHOSHASTIDER  
aus Kalkutta (Indien)  
Frankfurt 2012  
(D30)**

vom Fachbereich Biochemie, Chemie und Pharmazie (FB14)  
der Johann Wolfgang Goethe-Universität als Dissertation angenommen.

Dekan: Prof. Dr. Thomas Prisner

Gutachter: Prof. Dr. Volker Dötsch

Prof. Dr. Peter Güntert

Datum der disputation:

*Dedicated to my teachers and family*

## Table of Contents

Table of Contents.....	1
Abbreviations.....	7
Zusammenfassung.....	10
Summary.....	14
1 Introduction.....	18
1.1 Alzheimer's Disease.....	18
1.2 Regulated Intramembrane Proteolysis (RIP).....	19
1.2.1 Overview.....	19
1.2.2 Common Principles of RIP.....	20
1.2.3 I-CLiPs.....	20
1.2.4 RIP of APP.....	21
1.3 GxGD type protease $\gamma$ -Secretase (GS).....	22
1.3.1 Structure and interaction between the subunits.....	23
1.3.2 Processing and Maturation of GS.....	25
1.3.3 Endoproteolysis of PS.....	25
1.3.4 ER Retention Signals.....	25
1.3.5 Stoichiometry.....	26
1.3.6 Individual subunits of GS.....	26
1.3.7 GS regulating enzymes.....	27
1.3.8 Stepwise substrate processing in the water cavity.....	27
1.3.9 Substrate recognition .....	29
1.3.10 Effect of PS Mutations.....	29
1.3.11 Relationship with GPCRs and miRNAs.....	29
1.4 Structure of GxGD type protease.....	30
1.5 Amyloids and Amylome.....	34
1.5.1 Molecular Structures of Amyloids .....	36
1.5.2 Kinetics of the Growth of Amyloid Fibrils .....	41
1.5.3 Specific Mechanisms of Fibrillation .....	42
1.5.4 Conformationally Distinct Amyloid States .....	43
1.5.5 Molecular Simulations of Amyloids .....	44
1.5.6 Amyloid Can Be Beneficial for Cells and Also Convenient for Engineers .....	46
2 Cell-free Expression of Membrane Proteins .....	50
2.1 Overview.....	50

---

2.2 Cell-free expressed membrane proteins.....	51
2.3 Solved membrane protein structures of CF expressed proteins.....	51
2.4 Expression of Membrane Proteins (MPs):.....	52
2.5 Cell-free expression of MPs.....	53
2.6 CF extract sources:.....	53
2.7 E. coli based CF expression system.....	55
2.8 Various CF expression modes for MPs.....	57
2.8.1 P-CF Mode .....	58
2.8.2 D-CF Mode.....	58
2.8.3 L-CF Mode.....	58
3 Materials .....	60
3.1 Laboratory equipments.....	60
3.2 Chemicals.....	61
3.3 Software .....	61
3.4 Buffers and Media for S30 extract and T7 polymerase preparation: .....	62
3.5 Reagents for CF reaction: .....	62
3.6 SDS-gel buffers: .....	63
3.7 Gel .....	63
3.8 Sequencing primers for pET-Vectors .....	63
3.9 Materials for Nanodisc Preparation.....	63
3.10 Materials for Tag Variation Screen and Template Preparation .....	63
3.11 Microbial Strains.....	64
3.12 Oligonucleotides.....	64
4 Methods.....	65
4.1 Cell-free Expression of GS Components .....	65
4.1.1 S30 extract preparation .....	65
4.1.2 Production of T7RNAP.....	67
4.1.3 CECF reaction preparation for 1 ml RM and 16 ml FM.....	68
4.1.4 Template Production and Yield Optimization by Tag Variation Screen.....	69
4.1.5 General Template Design.....	69
4.1.6 DNA Template Preparation.....	70
4.1.7 Plasmid DNA.....	70
4.1.8 Linear DNA.....	70
4.1.9 Tag variation screen for improved expression.....	70
4.1.10 Primer Design.....	71
4.1.11 Linear DNA Template Preparation for Tag Variation Screen.....	71

---

4.1.12 General Setup of CF Expression Reactions.....	72
4.1.13 Analytical Scale CECF Reactions.....	72
4.1.14 Preparative Scale CECF Reactions.....	72
4.1.15 Quantification of Target Production by <sup>35</sup> S-Met Incorporation.....	73
4.1.16 Preparation of NDs as Supplement for CF Reactions.....	73
MSP1 Expression.....	73
MSP1 Purification.....	73
ND Assembly.....	74
ND Concentration, Storage and Stability.....	74
Co-translational Formation of MP/ND Complexes.....	75
Quantifying sGFP.....	75
CF Production of MP/ND Complexes.....	75
Purification of MP/ND Complexes.....	76
4.2 Protein Purification.....	76
4.3 SDS-PAGE and Western Blot:.....	77
4.4 Laser-induced liquid bead ion desorption-Mass spectrometry (LILBID).....	78
4.5 Blue Native PAGE.....	78
4.5.1 Silver Staining.....	78
5 Results.....	80
5.1 Cell-free Expression of GS Operon.....	80
5.2 Structural Investigation of Pen-2.....	84
5.3 Aph-1 tag variation.....	88
6 Introduction.....	89
6.1 Molecular Dynamics (MD) Simulations.....	89
6.1.1 Born-Oppenheimer approximation .....	89
6.1.2 Classical Newtonian Mechanics of Particles.....	89
6.1.3 Force Fields (FF).....	90
6.2 Reduced Representation of the system.....	92
6.2.1 Coarse-graining.....	92
Martini.....	92
6.2.2 Implicit Solvation.....	94
Effective Energy Function for Proteins in Solution (EEF1) .....	94
Implicit Membrane Model 1 (IMM1) .....	94
6.3 Analysis of MD Trajectories.....	95
6.3.1 Root Mean Square Deviation (RMSD).....	95
6.3.2 Secondary Structure .....	96

Table of Contents	4
6.3.3 Root Mean Square Fluctuations (RMSF).....	97
6.3.4 Solvent Accessible Surface Area .....	97
6.3.5 Principal Component Analysis (PCA).....	97
6.3.6 Change in radius of gyration (Rg) .....	99
6.3.7 Interaction energies plotted over time.....	99
6.3.8 Hydrogen bonding pattern .....	99
6.3.9 k-means or hierarchical clustering of trajectories .....	100
6.3.10 Free energy .....	100
6.4 Modeling Membrane Proteins.....	101
6.4.1 Membrane proteins.....	101
6.5 Protein-Protein and Protein-ligand Docking.....	102
7 Methods .....	102
7.1 Structure building and simulations of GS components.....	102
7.1.1 Modeling of Pen-2.....	102
Rosetta ab-initio membrane .....	102
HHSearch and Modeller.....	103
I-Tasser Server.....	104
7.2 Modeling of Bilyer Formation with Pen-2.....	105
7.3 MD Simulations of PS1-CTF .....	106
7.3.1 Coarse-grain (CG) simulations .....	106
7.3.2 Implicit Membrane Simulations.....	106
7.4 Simulations in Amber FF.....	106
7.5 Modeling of GPCRs (CB1,CB2 and $\beta$ 1AR) .....	107
7.6 Molecular modeling of FPR1.....	112
7.6.1 Homology modeling and refinement of FPR1.....	112
7.6.2 Receptor model equilibration in explicit membrane.....	112
7.6.3 Ligand preparation and docking.....	113
7.6.4 Molecular Dynamics.....	113
7.7 Homology Modeling of Human CXCR4 and Dopamine D3 Bound to Ligands.....	113
7.7.1 Homology Modeling.....	113
7.7.2 Criteria for prediction analysis, scoring and ranking: .....	114
7.7.3 PS1 CTF: APP simulations in implicit membrane.....	114
8 Results and Discussion .....	115
8.1 MD Simulation of NMR Structure of PS1-CTF .....	115
8.1.1 PS1-CTF MD Simulations in Detergent.....	115
8.1.2 MD Simulations in Lipid Bilayers.....	116

---

8.1.3 MD Simulations in Implicit Lipid Bilayer.....	118
8.1.4 Topology of PS1-CTF.....	121
8.1.5 Modeling of Pen-2.....	122
8.2 Modeling of PS1 CTF-APP Interface and CTF L383 mutations.....	125
8.2.1 SVM Predictions.....	129
8.2.2 Modeling APP, PS1 CTF and NTF Interface:.....	131
Modeling based on density functional theory (DFT):.....	131
9 Modeling of PS1-NTF.....	132
10 Activation mechanism of GPCRs.....	133
10.1 Introduction .....	133
10.1.1 Superfamily of GPCRs .....	137
10.2 Switches in rhodopsin-like receptors .....	142
10.2.1 The Ionic Lock Switch .....	142
10.2.2 The 3-7 Lock Switch .....	143
10.2.3 Transmission Switch (Former Trp Rotamer Toggle Switch) .....	143
10.2.4 Tyrosine Toggle Switch (nPxxy Motif) .....	145
10.2.5 The Elusive “Global Toggle Switch” .....	146
10.2.6 Role of Conserved Residues .....	147
10.2.7 Role of Extracellular Loops in Ligand Binding and Switching .....	148
10.3 Activation Schemes .....	149
10.3.1 Two Types of Activation Paths .....	150
10.4 Theoretical Studies on the Action of Molecular Switches.....	151
10.4.1 Single TM Studies.....	151
10.4.2 Studies on a Complete GPCR Model.....	152
Investigations of the Ionic Lock Switch .....	152
Breaking of the 3-7 Lock Switch .....	153
Beyond Classical MD Techniques .....	154
Consequences of the Ionic Lock Instability and TM Movements .....	155
10.5 Drug Design .....	157
10.6 Conclusions .....	159
10.7 Modeling of ligand binding to G protein coupled receptors: cannabinoid CB1, CB2 and adrenergic $\beta$ 2AR .....	162
10.7.1 Introduction.....	163
10.7.2 Results.....	165
10.7.3 Discussion.....	172
10.7.4 Conclusions.....	177



---

10.8 Water mediated activation mechanism of Formyl peptide receptor 1 (FPR1).....	178
10.8.1 Introduction.....	178
10.8.2 Results.....	179
FPR1 structure and the binding pocket.....	179
Interactions of ligands with binding site of FPR1.....	181
Models of FPR2 and FPR3 receptors.....	183
10.8.3 Discussion.....	184
The choice of CXCR4 as template structure.....	184
Comparison of structures based on rhodopsin and CXCR4 templates .....	184
Binding of tripeptide and tetrapeptide ligands.....	185
Role of water molecules in ligand binding.....	186
10.9 GPCRDock 2010 Results.....	189
10.10 References.....	190
11 Appendix.....	207
12 Acknowledgements.....	209
13. Curriculum Vitae.....	211

---

## Abbreviations

Å	Ångström
A $\beta$	amyloid $\beta$ -peptide
A $\beta$ 40/42	amyloid $\beta$ -peptide 40/42 aa
AD	Alzheimer disease
AICD	APP intracellular domain
Amp	ampicilline
Aph-1	Anteriorpharynx defective 1
APP	amyloid precursor protein
APS	ammoniumperoxodisulfate
ATP	adenosine triphosphate
$\beta$ -OG	n- $\beta$ -octyl-D-glucopyranoside
CD	circular dichroism
CECF	continuous exchange cell-free
CF	cell-free
CHAPS	3-[(3-Cholamidopropyl)dimethylammonio]propanesulfonic acid
CHAPSO	3-[(3-Cholamidopropyl)dimethylammonio]-2-hydroxy-1-propanesulfonate
CMC	critical micellar concentration
CTF	C-terminal fragment of Presenilin 1
CV	column volume
Da	Dalton
D-CF	detergent cell-free
DDM	n-dodecyl- $\beta$ -D-maltoside
DHPC	1,2-diheptanoyl-sn-glycero-3-phosphocholine
DNA	desoxyribonucleinacid
DPC	Fos- $\beta$ -choline® 12
dNTP	desoxyribonucleitriphosphate
DTT	dithiothreitol
E. coli	Escherischia coli
EDTA	ethylenediaminetetraacetic acid
ER	endoplasmic reticulum
et al.	and others
FAD	familial Alzheimer disease
FM	feeding mix
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HRP	horseradish peroxidase

---

HSQC	heteronuclear single quantum coherence
I-CliP	intramembrane-cleaving protease
IMAC	immobilised metal affinity chromatography
IPTG	isopropyl- $\beta$ -D-thiogalactopyranosid
kDa	kilodalton
LB	Luria-Bertani
L-CF	liposome cell-free
LMPG	1-myristoyl-2-hydroxy-sn-glycero-3-[phospho-RAC-(1-glycerol)]
LPPG	1-palmitoyl-2-hydroxy-sn-glycero-3-[phospho-RAC-(1-glycerol)]
M	Molar (mol/l)
mA	milliampère
mdeg	millidegree
mg	milligram
min	minute
ml	milliliter
mM	millimolar
MP	membrane protein
MW	molecular mass
NaP	sodiumphosphate
NG	N-nonyl- $\beta$ -D-glucoside
Nct	nicastrin
Ni-NTA	nickel-nitrilotriacetic acid
nm	nanometer
NMR	nuclear magnetic resonance
NTF	N-terminal fragment of presenilin 1
OD <sub>x</sub>	optical density at x nm
PAGE	polyacrylamide gel electrophoresis
P-CF	precipitate cell-free
PCR	polymerase chain reaction
Pen-2	presenilin enhancer 2
PS1	presenilin 1
PS2	presenilin 2
RM	reaction mix
ppm	parts per million
rpm	round per minute
S2P	site 2 protease
sAPP	soluble APP ectodomain
SEC	size exclusion chromatography
SDS	sodium dodecyl sulphate

---

SPP	signal peptide protease
CV	column volume
TA	Annealing-temperature
TEMED	N,N,N',N'-tetramethylethyldiamin
T <sub>M</sub>	melting temperature of the oligonucleotide
TMS	transmembrane segment
TRIS	Tris(hydroxymethyl)aminomethane
TX100	α-[4-(1,1,3,3-Tetramethylbutyl)phenyl]- -hydroxy-poly(oxy-1,2-ethanediyl)
UV	ultraviolet
V	volt
(v/v)	volume per volume
(w/v)	weight per volume

## Zusammenfassung

Die Alzheimer-Krankheit (AK), die erstmals vor mehr als einem Jahrhundert von Alzheimer beschrieben wurde, ist eine der häufigsten Formen der Demenz, von der über 30 Millionen Menschen weltweit betroffen sind (über 8 Millionen in Europa). Die Entstehung und Pathogenese von AK ist wenig verstanden und bis zum heutigen Zeitpunkt gibt es keine Heilung für diese Krankheit. AK ist durch die Akkumulation von senilen Plaques charakterisiert, die aus Amyloid-Beta-Peptiden (A $\beta$  37-43) nach Spaltung des Amyloid- Precursor-Proteins durch den Gamma-Sekretase-(GS)-Komplex entstehen. Deshalb kann GS ein attraktives Ziel für Medikamente sein. Da GS auch andere Substrate wie Notch, CD44 und Cadherine hat, führt unspezifische Hemmung von GS zu vielen Nebenwirkungen. Auf Grund des Fehlens einer Kristallstruktur von GS, was den extremen Schwierigkeiten bei der Aufreinigung zugeschrieben wird, kann molekulares Modelling sinnvoll sein, um die Architektur dieses Enzyms zu verstehen. Bisher wurden nur niedrig aufgelöste Cryo-EM Strukturen des Komplexes gelöst, die nur eine ungenaues Bild bei einer Auflösung zwischen 12-15 Å liefern. Weiterhin kann die Aktivität von GS in vitro mittels zellfreier (CF)- Expression hergestellt werden.

GS enthält katalytische Untereinheiten, nämlich Preseniline, und die Stützelemente Pen-2, Aph-1 und Nicastrin. Die Entstehung von AK liegt in der regulierten Intramembran-Proteolyse (RIP) begründet, die in verschiedenen physiologischen Prozessen und auch bei Leukämie eine Rolle spielt. Bisher wurde RIP für Wachstumsfaktoren, Cytokine, Rezeptoren, virale Proteine, Zelladhäsionsproteine, Signalpeptide und GS gezeigt. Während der RIP durchlaufen die Substrate extrazellulären Verdau und Intramembran Proteolyse.

Diese These beruht auf Molecular Modelling, Molecular Dynamics (MD) Simulationen, zell-freie (CF)-Expression, Massenspektrometrie, NMR, Kristallisation sowie Aktivitätstests der Komponenten des GS und von G-Protein-gekoppelter Rezeptoren (GPCRs).

Zuerst wurde die NMR-Struktur von PS1 CTF in Detergensmicellen und Lipiddoppelschichten mit Coarse-grained MD-Simulationen mit MARTINI Kraftfeld in Gromacs validiert. CTF wurde in DPC-Mizellen, DPPC- und DLPC- Lipiddoppelschichten simuliert. Ausgehend von zufälligen Konfigurationen von Detergens und Lipiden wurden Mizellen und Lipiddoppelschichten jeweils in Gegenwart von CTF ausgebildet, das während der Simulation in der Mizelle und in der Lipiddoppelschicht ausgerichtet wurde. DPC-Moleküle haben Mizellen um CTF geformt, in Übereinstimmung der experimentellen Ergebnisse, in denen 80 bis 85 DPC-Moleküle erforderlich

sind, um Mizellen zu bilden. Die Struktur in DPC ist der NMR-Struktur ähnlich, unterscheidet sich aber in Simulationen Lipiddoppelschichten bezüglich der Möglichkeit des Substrat-Docking im konservierten PAL-Motiv. Simulationen von CTF in impliziter Membran (IMM1) in CHAMM ergab eine ähnliche Struktur wie die aus Coarse-grained MD.

Die zell-freie Expression wurde optimiert gefolgt von Kristallisation und NMR-Spektroskopie von Pen-2 in verschiedenen Detergens-Micellen. Zusätzlich wurde Pen-2 durch eine Kombination der Rosetta Membran ab-initio-Methode, HHPred Homologiemodellierung unter Einbeziehung von NMR-Constraints modelliert. Die Modelle wurden von All-Atom- und Coarse-Grained-MD-Simulationen in Detergens-Mizellen und POPC / DPPC Lipid-Doppelschichten mit MARTINI Kraftfeld validiert.

GS-Operon bestehend aus allen vier Untereinheiten wurde mittels CF coexprimiert und aufgereinigt. Das Vorhandensein von GS-Untereinheiten nach Pull-Down mit Aph-1 wurde mittels Western Blot (Pen-2) und Massenspektrometrie (Presenilin-1 und Aph-1) bestimmt. Darüber hinaus wurden zusätzlich Interaktionen von PS1 CTF, APP und NTF mittels Docking und MD untersucht.

Modelle und Kontaktflächen von Pen-2 und PS1 NTF wurden geprüft und ihre Stabilität aus MD-Simulationen mit experimentellen Ergebnissen verglichen. Das Ziel war die Kontaktflächen zwischen GS-Untereinheiten durch Molecular Modelling mit den verfügbaren experimentellen Daten aus Cross-linking, Mutationsstudien und mittels NMR-Struktur des C-terminalen Fragments von PS1 und der Transmembran-Domäne APP zu modellieren. Die erhaltenen Kontaktflächen von GS-Untereinheiten können helfen, den Katalyse-Mechanismus aufzuklären, der für ein neues Lead-Design genutzt werden kann. Auf Grund des Fehlens einer Kristall-/ NMR-Struktur der GS-Untereinheiten mit Ausnahme des PS1 CTF, ist es nicht möglich, die Wirkung von Mutationen in Bezug auf APP-Spaltung vorherzusagen. Daher wurde zusätzlich ein Sequenz-basierter Ansatz auf maschinellem Lernen mit Support-Vektor-Maschine entwickelt, um die Wirkung von PS1 CTF L383 Mutationen in Bezug auf das A $\beta$ 40/A $\beta$ 42-Verhältnis mit 88% iger Genauigkeit vorherzusagen. Mutations-Daten aus der MOLGEN Datenbank, von PS1-Mutationen abgeleitet, wurden zum Training verwendet.

GPCRs (auch genannt 7TM-Rezeptoren) bilden eine große Superfamilie von Membranproteinen, die durch kleine Moleküle, Lipide, Hormone, Peptide, Licht, Schmerz, Geschmack und Geruch aktiviert werden können. Obwohl 50% der Medikamente auf dem Markt GPCRs als Ziel haben, werden nur wenige therapeutisch angegangen. Eine solche Vielzahl von Zielen kommt durch Einbeziehung von GPCRs in Signalwegen im Zusammenhang mit vielen Krankheiten zu Stande, welche Demenz (wie

Alzheimer-Krankheit), Stoffwechselerkrankungen (wie Diabetes) einschließlich endokrinologischer Störungen, viraler Infektionen, kardiovaskulärer, entzündlicher, Sinnes-Störungen, Schmerzen und Krebs mit einbezieht.

Cannabinoid und adrenerge Rezeptoren gehören zu der Klasse A (ähnlich Rhodopsin) GPCRs. Docking von Agonisten und Antagonisten an CB1- und CB2-Cannabinoid-Rezeptoren zeigte die Bedeutung eines zentral gelegenen Rotamer Kippschalters und seine mögliche Rolle in dem Mechanismus der Agonist/Antagonist-Erkennung. Der Schalter wird von zwei Resten gebildet, F3.36 und W6.48, die sich auf den gegenüberliegenden Transmembranhelices TM3 und TM6 im zentralen Teil der Transmembran-Domäne von Cannabinoid-Rezeptoren befinden. Die CB1- und CB2-Rezeptor-Modelle wurden basierend auf dem Adenosin-A<sub>2A</sub>-Rezeptor als Template konstruiert. Die zwei am genauesten beschriebenen Konformationen jedes Rezeptors wurden für die Docking-Versuche verwendet. In allen Posen (Ligand-Rezeptor-Konformationen), die durch die niedrigste intermolekulare Ligand-Rezeptor Energie und freie Bindungsenergie charakterisiert sind, entsprach der Liganden-Typ dem Zustand des Rotamer Kippschalters: Antagonisten fixierten einen inaktiven Zustand des Schalters, wohingegen Agonisten ihn veränderten. Unter Beibehaltung der gleichen mittleren Position des Liganden an der Bindungsstelle, haben die Molecular Dynamics-Simulationen im Falle der  $\beta$ 2AR-Agonisten, (R, R)- und (S, S)-Stereoisomere von Fenoterol verschiedene Bindungsarten nachgewiesen. Das (S, S)-Isomer war viel labiler in der Bindungsstelle und nur eine stabile Wasserstoffbrücke wurde ausgebildet. Solche dynamischen Bindungsmodi können vielleicht auch für Liganden von Cannabinoid-Rezeptoren gültig sein und dies auf Grund der hydrophoben Natur ihrer Ligand-Rezeptor-Wechselwirkungen. Allerdings können nur sehr lange Molekulardynamik-Simulationen die Gültigkeit solcher Bindungsmodi und wie sie sich auf den Prozess der Aktivierung auswirken verifizieren.

Humane N-Formyl Peptidrezeptoren (FPR) sind GPCRs, die an vielen physiologischen Vorgängen beteiligt sind, einschließlich Wirtsverteidigung gegen bakterielle Infektion und das Auflösen von Entzündungen. Die drei humanen FPRs (FPR1, FPR2 und FPR3) zeigen signifikante Sequenzhomologie und erfüllen ihre Wirkung über die Kopplung an das G<sub>i</sub>-Protein. Aktivierung von FPRs induziert eine Vielzahl von Antworten, welche von Agonist, Zelltyp, Rezeptorsubtyp, sowie Spezies abhängig sind. FPRs werden hauptsächlich durch phagozytische Leukozyten exprimiert. Gemeinsam binden diese Rezeptoren eine große Anzahl strukturell verschiedener Gruppen von agonistischen Liganden, einschließlich N-Formyl und nicht-Formyl Peptiden unterschiedlicher Zusammensetzung, die über Chemotaxis Phagozyten anziehen und aktivieren. Beispielsweise aktiviert N-Formyl-Met-Leu-Phe (fMLF), ein FPR1 Agonist, humane phagozytische entzündliche Antworten, wie intrazelluläre Calciummobilisierung, die Produktion von Zytokinen, Erzeugung von

reaktiven Sauerstoffspezies und Chemotaxis. Dieser Ligand kann effizient die wichtigsten bakterizide Neutrophilen-Funktionen aktivieren und es war eines der ersten charakterisierten bakteriell chemotaktischen Peptide. Während fMLF das bei weitem am häufigsten verwendete chemotaktische Peptid in Studien von Neutrophilen-Funktionen ist, sind atomistische Beschreibungen für den fMLF-FPR1 Bindungsmodus noch mangelhaft, vor allem wegen des Fehlens einer Kristallstruktur dieses Rezeptors. Aufklärung der Bindungsmodi kann zur Gestaltung neuartiger und effizienter Nicht-Peptid-FPR1 Medikamente beitragen. Molecular Modelling von FPR1, auf der anderen Seite, kann eine effiziente Möglichkeit sein, Details der Ligandenbindung und Aktivierung des Rezeptors zu offenbaren. Allerdings wurden kürzlich durchgeführte Modellierungen von FPRs nur auf Rinderrhodopsin als Vorlage beschränkt.

Um spezifische Liganden-Rezeptor-Wechselwirkungen auf einer besser geeigneten Vorlage als Rhodopsin zu lokalisieren, wurde ein Homologie-Modell von FPR1 mit Hilfe der Kristallstruktur des Chemokinrezeptor CXCR4 generiert, das mehr als 30% Sequenzidentität mit FPR1 zeigt und in dem gleichen  $\gamma$ -Strang des Stammbaums der GPCRs zugeordnet ist (Rhodopsin gehört zum  $\alpha$ -Strang). Danach wurden Docking und Verfahren zur Modellverfeinerung verfolgt. Schließlich wurden 40 ns Full-Atom MD Simulationen für die Apo-Form sowie für Komplexe aus fMLF (Agonist) und tBocMLF (Antagonist) mit FPR1 in der Membran durchgeführt. Basierend auf der Lokalisierung der N- und C-Termini des Liganden konnte die extrazelluläre FPR1-Bindetasche in zwei Zonen unterteilt werden, nämlich den Anker- und die Aktivierungs-Regionen. Der formylierte M1-Rest von fMLF gebunden an den Aktivierungsbereich führte zu einer Reihe von Konformationsänderungen von konservierten Resten. Interne Wassermoleküle, die in erweiterte Wasserstoffbrücken-Netzwerke beteiligt sind, spielen eine entscheidende Rolle beim Übertragen der Agonist-Rezeptor-Wechselwirkungen. Ein Mechanismus der ersten Schritte der Aktivierung bei gleichzeitiger Ligandenbindung wird vorgeschlagen.

Die Struktur und Ligandenbindungs-Pose des Dopamin-Rezeptor-3 (RMSD mit der Kristallstruktur: 2,13 Å) und Chemokinrezeptor 4 (CXCR4, RMSD mit der Kristallstruktur 3,21 Å) wurde mit hoher Genauigkeit im GPCR-Dock 2010 Wettbewerb vorhergesagt. Dabei erzielte das hier beschriebene Homologie-Modell des Dopamin-Rezeptor-3 die achtbeste Gesamtleistung.



---

## Summary

Alzheimer's disease (AD), which was first reported more than a century ago by Alhzeimer, is one of the commonest forms of dementia which affects >30 million people globally (>8 million in Europe). The origin and pathogenesis of AD is poorly understood and there is no cure available for the disease. AD is characterized by the accumulation of senile plaques composed of amyloid beta peptides (Ab 37-43) which is formed by the gamma secretase (GS) complex by cleaving amyloid precursor protein. Therefore GS can be an attractive drug target. Since GS processes several other substrates like Notch, CD44 and Cadherins, nonspecific inhibition of GS has many side effects. Due to the lack of crystal structure of GS, which is attributed to the extreme difficulties in purifying it, molecular modeling can be useful to understand its architecture. So far only low resolution cryoEM structures of the complex has been solved which only provides a rough structure of the complex at low 12-15 A resolution. Furthermore the activity of GS in vitro can be achieved by means of cell-free (CF) expression.

GS comprises catalytic subunits namely presenilins and supporting elements containing Pen-2, Aph-1 and Nicastrin. The origin of AD is hidden in the regulated intramembrane proteolysis (RIP) which is involved in various physiological processes and also in leukemia. So far growth factors, cytokines, receptors, viral proteins, cell adhesion proteins, signal peptides and GS has been shown to undergo RIP. During RIP, the target proteins undergo extracellular shredding and intramembrane proteolysis.

This thesis is based on molecular modeling, molecular dynamics (MD) simulations, cell-free (CF) expression, mass spectrometry, NMR, crystallization, activity assay etc of the components of GS complex and G-protein coupled receptors (GPCRs).

First I validated the NMR structure of PS1 CTF in detergent micelles and lipid bilayers using coarse-grained MD simulations using MARTINI forcefield implemented in Gromacs. CTF was simulated in DPC micelles, DPPC and DLPC lipid bilayer. Starting from random configuration of detergent and lipids, micelle and lipid bilyer were formed respectively in presence of CTF and it was oriented properly to the micelle and bilyer during the simulation. Around DPC molecules formed micelle around CTF in agreement of the experimental results in which 80-85 DPC molecules are required to form micelles. The structure obtained in DPC was similar to that of NMR structure but differed in bilayer simulations showed the possibility of substrate docking in the conserved PAL motif. Simulations of CTF in implicit membrane (IMM1) in CHAMM yielded similar structure to that from coarse grained MD.

I performed cell-free expression optimization, crystallization and NMR spectroscopy of Pen-2 in various detergent micelles. Additionally Pen-2 was modeled by a combination of rosetta membrane ab-initio method, HHPred distant homology modeling and incorporating NMR constraints. The models were validated by all atom and coarse grained MD simulations both in detergent micelles and POPC/DPPC lipid bilayers using MARTINI forcefield.

GS operon consisting of all four subunits was co-expressed in CF and purified. The presence of GS subunits after pull-down with Aph-1 was determined by western blotting (Pen-2) and mass spectrometry (Presenilin-1 and Aph-1). I also studied interactions of especially PS1 CTF, APP and NTF by docking and MD.

I also made models and interfaces of Pen-2 with PS1 NTF and checked their stability by MD simulations and compared with experimental results. The goal is to model the interfaces between GS subunits using molecular modeling approaches based on available experimental data like cross-linking, mutations and NMR structure of C-terminal fragment of PS1 and transmembrane part of APP. The obtained interfaces of GS subunits may explain its catalysis mechanism which can be exploited for novel lead design. Due to lack of crystal/NMR structure of the GS subunits except the PS1 CTF, it is not possible to predict the effect of mutations in terms of APP cleavage. So I also developed a sequence based approach based on machine learning using support vector machine to predict the effect of PS1 CTF L383 mutations in terms of A $\beta$ 40/A $\beta$ 42 ratio with 88% accuracy. Mutational data derived from the Molgen database of Presenilin 1 mutations was using for training.

GPCRs (also called 7TM receptors) form a large superfamily of membrane proteins, which can be activated by small molecules, lipids, hormones, peptides, light, pain, taste and smell etc. Although 50% of the drugs in market target GPCRs, only few are targeted therapeutically. Such wide range of targets is due to involvement of GPCRs in signaling pathways related to many diseases i.e. dementia (like Alzheimer's disease), metabolic (like diabetes) including endocrinological disorders, immunological including viral infections, cardiovascular, inflammatory, senses disorders, pain and cancer.

Cannabinoid and adrenergic receptors belong to the class A (similar to rhodopsin) GPCRs. Docking of agonists and antagonists to CB1 and CB2 cannabinoid receptors revealed the importance of a centrally located rotamer toggle switch, and its possible role in the mechanism of agonist/antagonist recognition. The switch is composed of two residues, F3.36 and W6.48, located on opposite transmembrane helices TM3 and TM6 in the central part of the membranous domain of cannabinoid receptors. The CB1 and CB2 receptor models were constructed based on the adenosine A2A receptor template. The two best scored conformations of each receptor were used for the docking procedure. In all poses (ligand-receptor conformations) characterized by the lowest ligand-receptor

intermolecular energy and free energy of binding the ligand type matched the state of the rotamer toggle switch: antagonists maintained an inactive state of the switch, whereas agonists changed it. In case of agonists of  $\beta$ 2AR, the (R,R) and (S,S) stereoisomers of fenoterol, the molecular dynamics simulations provided evidence of different binding modes while preserving the same average position of ligands in the binding site. The (S,S) isomer was much more labile in the binding site and only one stable hydrogen bond was created. Such dynamical binding modes may also be valid for ligands of cannabinoid receptors because of the hydrophobic nature of their ligand-receptor interactions. However, only very long molecular dynamics simulations could verify the validity of such binding modes and how they affect the process of activation.

Human N-formyl peptide receptors (FPRs) are G protein-coupled receptors (GPCRs) involved in many physiological processes, including host defense against bacterial infection and resolving inflammation. The three human FPRs (FPR1, FPR2 and FPR3) share significant sequence homology and perform their action via coupling to  $G_i$  protein. Activation of FPRs induces a variety of responses, which are dependent on the agonist, cell type, receptor subtype, and also species involved. FPRs are expressed mainly by phagocytic leukocytes. Together, these receptors bind a large number of structurally diverse groups of agonistic ligands, including *N*-formyl and nonformyl peptides of different composition, that chemoattract and activate phagocytes. For example, *N*-formyl-Met-Leu-Phe (fMLF), an FPR1 agonist, activates human phagocyte inflammatory responses, such as intracellular calcium mobilization, production of cytokines, generation of reactive oxygen species, and chemotaxis. This ligand can efficiently activate the major bactericidal neutrophil functions and it was one of the first characterized bacterial chemotactic peptides. Whereas fMLF is by far the most frequently used chemotactic peptide in studies of neutrophil functions, atomistic descriptions for fMLF-FPR1 binding mode are still scarce mainly because of the absence of a crystal structure of this receptor. Elucidating the binding modes may contribute to designing novel and more efficient non-peptide FPR1 drug candidates. Molecular modeling of FPR1, on the other hand, can provide an efficient way to reveal details of ligand binding and activation of the receptor. However, recent modelings of FPRs were confined only to bovine rhodopsin as a template.

To locate specific ligand-receptor interactions based on a more appropriate template than rhodopsin we generated the homology models of FPR1 using the crystal structure of the chemokine receptor CXCR4, which shares over 30% sequence identity with FPR1 and is located in the same  $\gamma$  branch of phylogenetic tree of GPCRs (rhodopsin is located in  $\alpha$  branch). Docking and model refinement procedures were pursued afterward. Finally, 40 ns full-atom MD simulations were conducted for the Apo form as well as for complexes of fMLF (agonist) and tBocMLF (antagonist) with FPR1 in the

membrane. Based on locations of the N- and C-termini of the ligand the FPR1 extracellular pocket can be divided into two zones, namely, the anchor and activation regions. The formylated M1 residue of fMLF bound to the activation region led to a series of conformational changes of conserved residues. Internal water molecules participating in extended hydrogen bond networks were found to play a crucial role in transmitting the agonist-receptor interactions. A mechanism of initial steps of the activation concurrent with ligand binding is proposed.

I accurately predicted the structure and ligand binding pose of dopamine receptor 3 (RMSD to the crystal structure: 2.13 Å) and chemokine receptor 4 (CXCR4, RMSD to the crystal structure 3.21 Å) in GPCR-Dock 2010 competition. The homology model of the dopamine receptor 3 was 8 th best overall in the competition.

## Part I: Cell-free expression of the $\gamma$ -secretase complex

### 1 Introduction

#### 1.1 Alzheimer's Disease

Alzheimer's Disease (AD) first described by Alzheimer in 1907 is one of the most common forms of dementia affecting the elderly [1]. It affects >30 million people worldwide (>8 million in Europe), and is a leading cause of death among the elderly population. It is projected to affect >1% population globally by 2050 [2]. The disability caused by AD among people older than 60 years is higher than that of cancer, stroke and cardiovascular disease. As a result the economic cost of treating AD is very high. AD is characterized by the progressive decline in memory and cognitive abilities.

Currently there is no cure available to stop progression of AD, and therefore novel drugs are urgently required. Four drugs currently approved for AD e.g. tacrine, donepezil, rivastigmine, galantamine, and memantine only provides temporary relief. Memantine is an NMDA receptor antagonist while others are acetylcholinesterase inhibitors [3].

There are two types of AD - early onset or Familial Alzheimer's Disease (FAD) and late onset. Early onset AD affected patients (~5%) have mutations in the genes related to the processing of APP (i.e. Presenilin). The most prevalent late onset AD mainly affects people older than 65 years. The exact mechanism of development and progression of AD is controversial although Amyloid cascade hypothesis widely accepted. Review by Jakob-Roetne et al 2009 provides a good overview on it. The hypothesis states that the formation and aggregation of A $\beta$  oligomers extracellularly (known as senile plaques) and tau proteins intracellularly (known as neurofibrillary tangles) and on the walls of cerebral blood vessels [4] resulting in malfunction and loss of synapse and neurons leads to AD. The aggregation of tau protein is proposed to be due to the imbalance between A $\beta$  production and clearance. Loss of neurons takes place mainly in cortex and hippocampus (Fig 1.1).

The origin of AD is hidden in the regulated intramembrane proteolysis of APP which is involved in various physiological processes and also in leukemia [5].  $\gamma$ -secretase complex (GS) (Fig 1.3) is an intramembrane cleaving protease (iCLIP) [6] that cleaves the Amyloid Precursor Protein into Amyloid  $\beta$  peptides (A $\beta$  39-43), which aggregate in the brain of the Alzheimer's patients as senile plaques. So GS is a potential target for drugs and compounds that modulate its activity.

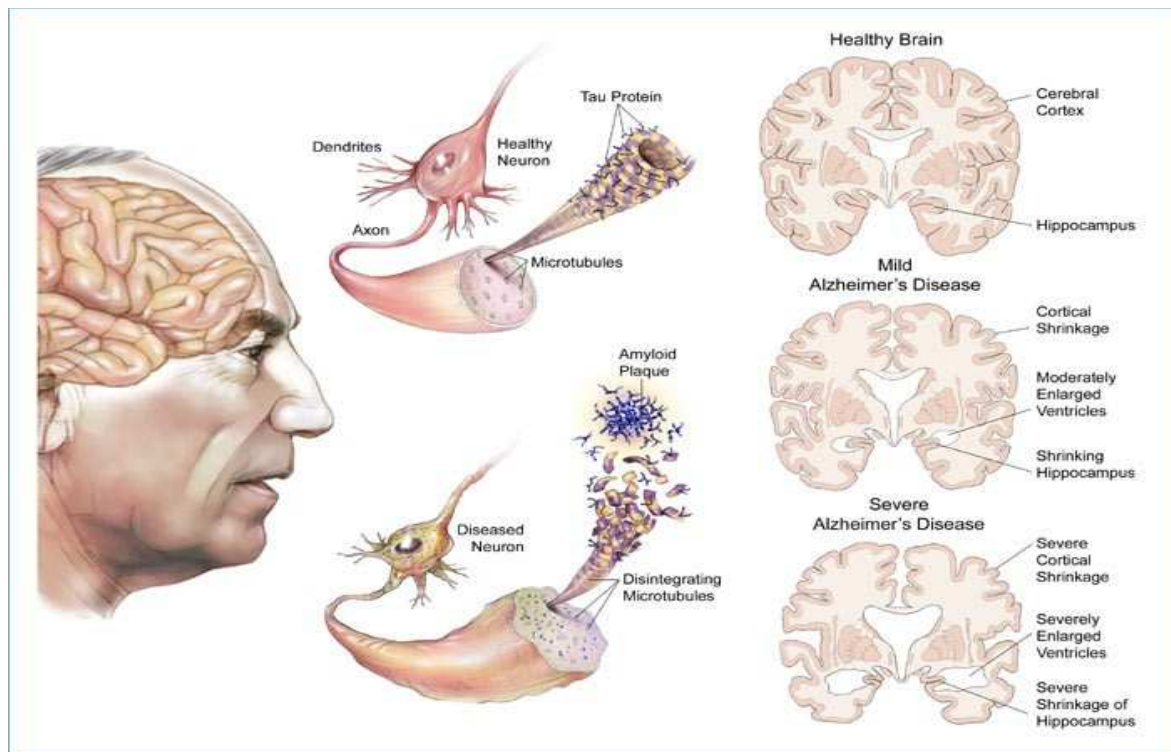


Fig 1.1 Diagram showing various characteristics of AD: deposition of amyloid plaques and tau proteins resulting in neuronal death and shrinking of brain regions. Source: <http://sierram.web.unc.edu/2011/04/22/caffeine-and-alzheimers-disease/>

## 1.2 Regulated Intramembrane Proteolysis (RIP)

### 1.2.1 Overview

The cell uses a variety of ways to communicate or respond to the environment. Regulated Intramembrane Proteolysis (RIP) is one of the many ways of doing that. For detailed review of RIP, refer to [5,7]. The term RIP was coined by Brown et al in 2000 when only a handful of proteins were found to undergo two step processing: extracellular shredding and intramembrane proteolysis [8]. Recent evidences suggest that RIP is not only involved in normal physiological processes but also in disease. More than 60 substrates of RIP have been identified so far [9,10] Substrates of RIP include growth factors, cytokines, receptors, viral proteins, cell adhesion proteins, signal peptides etc have been shown to undergo RIP [7,8]. (Table 1) Lack of proper RIP leads to diseases like AD and leukemia. RIP takes place not only in plasma membrane, but also in golgi apparatus and endoplasmic reticulum. RIP is involved in signal transduction during growth, development, immune response, cell differentiation, transcriptional regulation, cell adhesion, axon guidance, lipid metabolism etc [10]. Soluble intracellular protein products of RIP either act as signal transducer or transcription factor (i.e. Notch, Growth Factors, CD44, TNF alpha) or gets degraded [11]. Although several hundred proteins are subjected to shredding of the juxtamembrane domain, the following intramembrane cleavage is yet to be determined for many of them [12]. For instance matrix metalloproteinases also cleave membrane proteins without leading to transmembrane cleavage. However it is still not clear how the

substrate recognition takes place during RIP and what are the common features of various RIP.

Generally proteases are classified into Serine, Threonine, Cysteine, Aspartate, Metalloproteases and Glutamic proteases. However the initial shredding of the substrate ectodomain is carried out by so called shredders which include:

- ADAM (a disintegrin and metalloprotease) family
- aspartyl proteases  $\beta$ -site APP-cleaving enzymes (BACE1/2)

The following transmembrane cleavage is carried out by intramembrane cleaving proteases (I-CLiPs) [6] which consists of:

- GxGD type aspartyl protease (G- Gly, x – any amino acid, D – Asp): GS, SPP, SPPLs, bacterial type four prepilin peptidases
- S2P metalloproteases (zinc metalloprotease site 2 protease) : S2P is involved in the processing of sterol regulatory element binding protein. It has 4 TMDs and HEXXH motif.
- rhomboid serine proteases

### 1.2.2 Common Principles of RIP

- Substrate is a transmembrane protein
- Proteolysis is mediated by an I-CLiP
- Hydrolysis occurs within/close to the membrane
- Regulation by biological stimuli
- The cleaved intracellular domain of the substrate possesses a signaling function
- RIP results in a defined biological response [7]

### 1.2.3 I-CLiPs

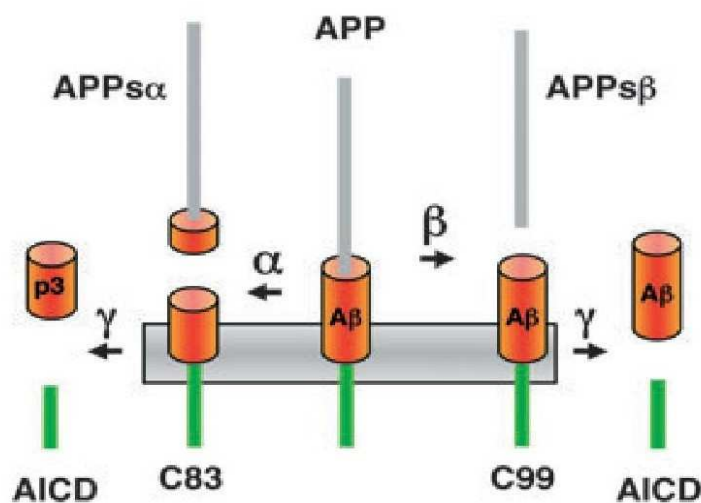
I-CLiPs are integral membrane proteins which carry their active sites in the hydrophobic helices buried inside the hydrophobic membrane environment as shown in the crystal structure of rhomboid protease GlpG from *E. coli* [6,13]. Also the catalysis takes place in presence of water inside the membrane cavity. In case of GlpG, the catalytic triad consists of His and Ser and is situated  $\sim 10$  Å below the membrane lumen interface in a water cavity formed by 6 transmembrane helices. The substrate enters into the hydrophobic cavity in a stepwise manner as revealed in case of GlpG where substrate enters between TM 3 and TM 5 in which TM 5 can be gating helix which modulate substrate accessibility to the active site [14,15]. Similar strategy for substrate catalysis is probably shared by other members of the family.

Often the I-CLiPs don't function until the length of the ectodomain of the substrate is reduced to  $< 30$  amino acids by shredding at the extracellular scissile peptide bond. Probably the longer substrates can not penetrate into the I-CLiP active site due to steric clashes. However rhomboid does not require prior cleavage of substrate unlike GS [6]. Just the opposite i.e. extracellular shredding

without intramembrane cleavage takes place in glycosyl phosphatidy inositol (GPI) anchored proteins like prion which is bereft of a transmembrane segment [12]. Failure of RIP regulation i.e. in case of Notch, higher and lower Notch signaling results in Leukemia and developmental defects respectively. In case of beta-secretase, higher RIP results in early onset AD. Unraveling the regulatory mechanisms of RIP can be of potential benefit to drug targeting against various diseases. The regulation mostly takes place during ectodomain shredding.

### 1.2.4 RIP of APP

One of the first discovered substrate of RIP is type I membrane protein APP, which was studied in detail in the past decade. APP undergoes initial juxtamembrane shredding either by  $\alpha$ -secretase (which was identified to be ADAM10) or by  $\beta$ -secretase (which is BACE1). The product of BACE1 is called APPs $\beta$  which is released in the lumen and the membrane bound c99, which is further cleaved by the GS resulting in Ab 37-43 peptides, which are released in lumen whereas the other product AICD (APP intracellular domain) goes to the cytosol triggering signal transduction pathways. Ab42 and Ab38 species were shown to be the causative agents of the senile plaques observed in AD patients. Therefore both BACE1 and GS were subjects of intense research since past decade.



*Fig 1.2 Sequential cleavage (RIP) of APP by  $\alpha$ -,  $\beta$ - and  $\gamma$ -secretase. APP is cleaved in two competing pathways: amyloidogenic and non-amyloidogenic. In amyloidogenic pathway, APP is first cleaved by  $\beta$ -secretase in the extracellular domain resulting in soluble APP ectodomain (APPs $\beta$ ) and membrane bound C-terminal 99 amino acid long fragment of APP called C99. C99 is further cleaved by I-CLiP GS giving A $\beta$  37-43 species which goes to the lumen and APP intracellular domain (AICD) secreted in the cytoplasm. On the contrary in non-amyloidogenic*

*pathway ADAM metalloprotease  $\alpha$ -secretase shreds APP giving rise to soluble APP ectodomain APPs $\alpha$  and C-terminal 83 amino acid long fragment of APP called C83. Then GS cleaves C83, producing secreted p3 peptide and AICD. Figure adapted from [5]*

In the alternative pathway, APP is cleaved to membrane bound c83 and soluble APPs $\alpha$  which is released in the lumen and is shown to have neurotropic effects. c83 is then cleaved by GS giving rise to p3 peptide released in the lumen is not pathogenic.  $\alpha$ -,  $\beta$ - secretase compete with each other to process the APP.

Recently the structure of APP was solved (PDB ID: ) by NMR in LMPG micelles [16]. TMD of APP



is a curved helix making it suitable for progressive cleavage by GS. The conserved GxxxG motif APP (700-704) which is involved in its dimerization also binds to cholesterol. Gly708 renders flexibility of APP. G<sub>700</sub>AIIG<sub>704</sub> segment of APP plays pivotal role in biogenesis of Ab<sub>39-43</sub> species.

Pairwise replacement of Gly with Leu in APP enhances homodimerization but leads to a drastic reduction of Abeta secretion. Replica exchange MD simulations reveal that dimerization of the WT APP is mediated by C(alpha)-H...O hydrogen bonds between two APP fragments contrary to the hydrophobic interactions responsible for the dimerization of the mutant. So in the tilted mutant the gamma cleavage site is shifted resulting in inhibition of Ab production [17,18].

### 1.3 GxGD type protease $\gamma$ -Secretase (GS)

$\gamma$ -Secretase (GS) (for review see: [19]) is an ICLiP (i.e. it processes its substrates inside the membrane in presence of water), which cleaves the amyloid precursor protein to amyloid beta peptides (A $\beta$  37-43), that accumulate in the brains of Alzheimer's patients as senile plaques. GS comprises four subunits aka, Presenilins (PS1 and PS2), Presenilin Enhancer 2 (Pen-2), Anterior pharynx defective phenotype 1 (Aph-1) and Nicastrin which together have in total 19 transmembrane spanning domains [20]. Presenilin 1 acts as the catalytic subunit of the complex harbouring two catalytic aspartate residues in N- and C-terminal fragments (NTF TMD 6- D257 and CTF TMD7 – D385) [21]. GS also cleaves Notch among many other substrates, and it was shown to be functional *in vivo* in presence of Pen-2 [22].

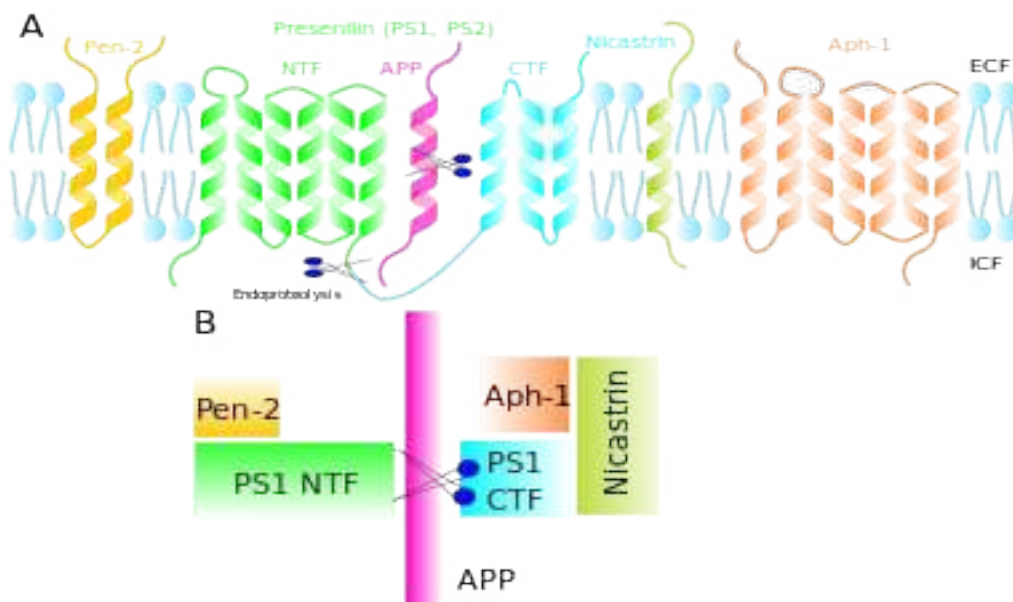


Fig 1.3 Schematic representation (A) and interactions (B) of the GS components Pen-2, Presenilin (NTF and CTF), Nicastrin and Aph-1. APP is processed by the catalytic aspartates located in the TMD 6 (NTF) and TMD 7 (CTF) of presenilin. NTF and CTF is formed by autoproteolysis by presenilin. B. Pen-2 was shown to interact with PS1 NTF whereas PS1 CTF, Aph-1 and Nicastrin interacts with each other [447].

Pen-2 plays crucial role in origin, maturation and functioning of the complex. Pen-2 is involved in the autoproteolysis of the PS1 into NTF and CTF, which is required for its functioning. Pen-2 associates with NTF and the complex of CTF, Aph-1 and Nicastrin to form an active GS complex. GS also plays important role in tumor development and cancer progression through APP and Notch [23].

The usefulness of GS as a drug target is limited by the fact that it has several other substrates including Notch, Cadherins, CD44 etc which are essential for viability. Consequently, nonspecific inhibition of GS have major side effects. Therefore the knowledge of the 3D architecture of the complex is required for rational drug design [5].

There are two isoforms of Presenilin : PS1 and PS2 and also Aph-1: Aph-1a and Aph-1b. Aph-1a can further have short (Aph-1aS) and long (Aph-1aL) splice variants [24]. However both of these PS isoforms are not associated to the complex simultaneously [25]. Thus there are six plausible GS complexes. However the role of these various GS variants in pathogenesis of AD has not been studied in detail. But specific inhibition of Aph1B GS reduced the phenotypes observed in mouse model of AD without affecting notch signalling [26]. However all the isoforms have been shown to form functional GS complex despite showing heterogeneity in substrate processing [27]. Although the activity of GS have been established in vitro only in presence of PS1 and Pen-2, nevertheless all four subunits are required for its functioning in vivo [20,22]. The most studied components of complex consist of PS1, Pen-2, Aph-1 and Nct.

### **1.3.1 Structure and interaction between the subunits**

GS subunits PS1, Pen-2, Aph-1 and Nct have 9, 2, 1, 7 TMDs respectively i.e. 19 TMDs in total. However crystallographic structure determination of the complex has not been possible so far due to technical difficulties in obtaining high amounts of the complex required for crystallization. So far only low resolution cryoEM structures of the complex has been solved which only provides a rough structure of the complex at low 12-18 Å resolution [28–30].

However the low resolution maps fail to deliver any information on molecular interactions between GS subunits. The cyroEM structure has several domains on the extracellular side, three solvent-accessible low-density cavities and a potential substrate-binding surface groove in the transmembrane region [29].

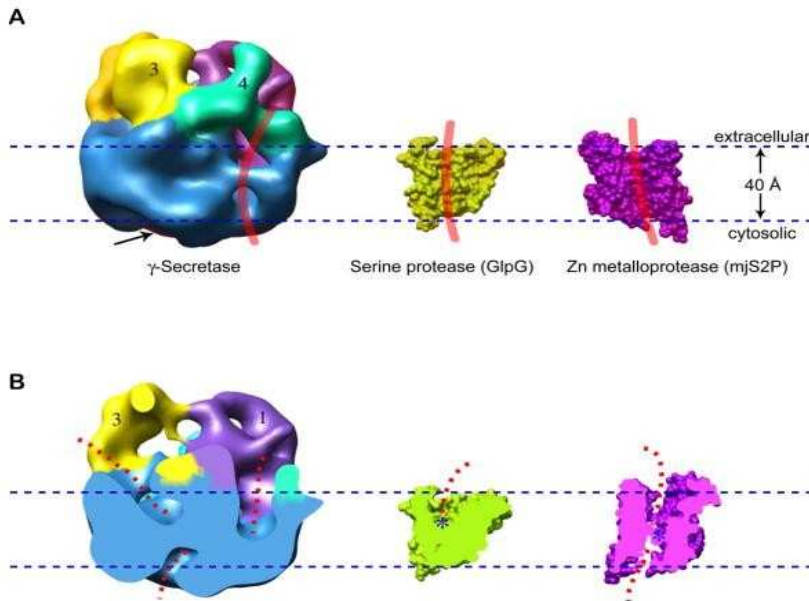


Fig 1.4 A comparison of the  $\gamma$ -secretase cryo-EM structure with the crystal structures of the bacterial and archaeal intramembrane proteases. Figure adapted from [29].

However biochemical chemical cross-linking and cysteine cross-linking experiments have revealed some conserved residues involved in the interface between GS subunits. For example, the conserved WNF motif of PS1 TMD4 is interacting with another N in Pen-2 TMD1. However it is controversial whether the interacting N is located in N-termini or C-termini of Pen-2 TMD1. Pen-2 and PS1 NTF were shown to form separate complex than that of PS1 CTF, Nct and Aph-1 complex [31–33]. WNF motif of PS1 TMD4 was also shown to be involved in ER retention and retrieval. Further NN motif of Pen-2 TMD1 was proposed to bind to PS1 TMD4 [33]. TMD 1 and 8 of PS are close to each other and might interact with the active site. TMD 8 is a distorted form of an ideal helix [34]. Gly 22 and Pro 27 of Pen-2 was found to be essential for GS complex formation. The TMD 9 is also in close proximity to the active site [35]. PAL motif and TMD 9 of PS are involved in the formation of the catalytic pore [36].

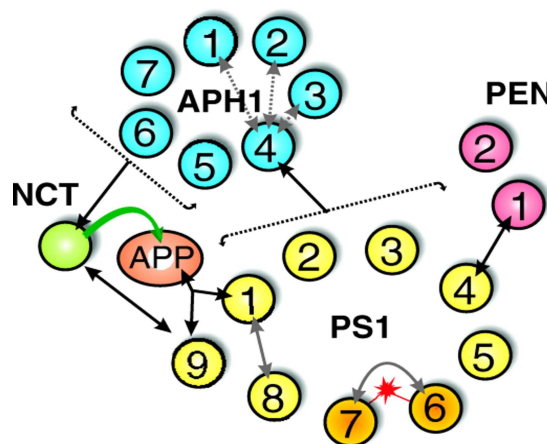


Fig 1.5: Intra- and intermolecular interactions in the  $\gamma$ -secretase complex. Given the hydrophobic nature of  $\gamma$ -secretase, most prominent and primary interactions are likely to be governed through or include their TMDs and hydrophobic domains. Here we present a bird's eye view of the TMDs of PS1 (yellow), NCT (green), APH1 (blue) and PEN2 (red), including the reported intra- (grey arrows) and intermolecular (black arrows) interactions. Suggested interactions are shown by dotted arrows, such as intramolecular interactions in APH1 (via GxxG motifs in TMD4) and the

as-yet-uncharacterised interaction domains for APH1 in PS1 and for NCT in APH1. Ectodomain interactions of NCT with APP-CTF are indicated with a green arrow (see text for details). The red sparkle denotes the catalytic aspartate dyad. Figure adapted from [38]

Mutations in TMD4 (G126) and TMD5 (H171) of Aph-1aS inhibits the formation of the Nct/Aph-1 subcomplex. Although mutations in TMD3 (Q83/E84/R85) and TMD6 (H197) of APH-1aS does not

affect subcomplex formation, they inhibit further association and autoproteolysis [37]. Two conserved His 171 and His 197 of Aph-1 has been shown to be important for GS activity.

### 1.3.2 Processing and Maturation of GS

After ribosomal translation, GS is assembled first in ER where Nct and Aph-1 forms initial complex (Reviewed in [38]) which then associates to PS. Next Pen-2 enters in the trimeric complex leading to the endoproteolysis of PS into NTF and CTF [39]. Thus Pen-2 is involved in the maturation and stabilizing the complex. Then the complex is transported to the plasma membrane through ER and golgi apparatus. Glycosylation of Nct takes place in golgi complex. ER retains the unassembled PS and Pen-2 [40,41]. Additional S-palmitoylation of Nct and Aph-1 is observed [42].

### 1.3.3 Endoproteolysis of PS

It is not yet fully understood, how the endoproteolysis of PS takes place inside the membrane. However  $\epsilon$ -,  $\zeta$ -, and  $\gamma$ -like sites of endoproteolysis have been identified at amino acids 292/3 (minor), 295/6, 298/9 (major) (Fig 1.6) [43]. The hydrophobic nature of the amino acids around the sites probably helps it to enter plasma membrane for a stepwise cleavage. It was shown that the cleavage occurs in successive interval of three amino acids each like in APP [44]. This helps in getting rid of the products from the catalytic pore which harbors the aspartates in TMD6 and 7 in NTF and CTF respectively. This stepwise cleavage was also found for APP which will be discussed in detail in the following paragraphs. However it is not clear how GS processes type-II membrane protein unlike the related SPP or SPPLs. Autoproteolysis of GS is necessary for its functionality. Mutation in one of the catalytic aspartates can block endoproteolysis [21]. It has been proposed that exon-9 encoded autoproteolysis site actually keeps the GS in inactive form to prevent it from non specific substrate cleavage [45].

### 1.3.4 ER Retention Signals

ER retention signals are important for studying subunit interactions of GS. The retention signals are masked during GS complex formation by its subunit interaction which results in the secretion of mature GS complex. Uncomplexed GS members are retained in the ER in this manner. Presenilin C-terminus is required for binding to Nicastrin, ER retention and GS activity [46]. Rer1p (endoplasmic retention factor 1p) competes with APH-1 for binding to the polar residues of nicastrin TMD and is involved in its ER retention [47]. These signals are different from RXR ER retention signals in ion channels [48].

Only fully assembled complexes are transported from ER which retains the unassembled subunits. Pen-2 and TMD 4 of PS carry ER retention/retrieval signals. When both of them interact, the ER

retention signal is masked, and it allows surface transport of GS complex [41]. Unassembled TMD1 of Pen-2 interacts with ER retention factor Rer1 to stay attached to the ER [40]. Over expression of Rer1 retains unassembled Pen-2 in ER. ER retention/retrieval signals like RXR are found in many ion channels. ER retention factors like Rer1 interacts with the polar residues of the membrane proteins [48].

Recent studies indicate that during AD, PS is found in high amounts in subcompartment of the endoplasmic reticulum (ER) that is physically and biochemically connected to mitochondria, called mitochondria-associated ER membranes (MAMs). These findings explain including altered lipid metabolism and calcium homeostasis during AD [49].

### 1.3.5 Stoichiometry

The stoichiometry of the complex was determined to be 1:1:1:1 which is in agreement with the molecular weight of each component [50]. However the MW of it varies from 250 kDa to 2000 kDa in literature depending on the method used for determining the MW, albeit agreeing the same stoichiometry [29]. But there are some evidences of dimeric GS with 2:2:2:2 stoichiometry [51].

### 1.3.6 Individual subunits of GS

All GS subunits are integral membrane proteins.

**Presenilin (PS):** Presenilin is an integral membrane protein in type II orientation. N and C-terminal fragment of PS spans the membrane 6 and 3 times respectively [52,53]. Recent evidences suggest that from TMD4, only N204 interacts with Pen-2, D194, T197 and N204 is involved in ER retention and D194 is required for complex stability. Most of the mutations during FAD are linked to PS.

**Pen-2:** Pen-2 is a 101 amino acid long double membrane spanning type 1 membrane protein which has both N and C-terminal facing the lumen [54]. It was identified in a genetic screen for modulators of PS activity in *C. elegans* [55]. In absence of Pen-2, APP and Notch can't be processed by PS. Pen-2 is involved in the endoproteolysis of GS as shown by RNA interference [39]. There are N-linked glycosylation sites only in N and C-terminal loop of Pen-2. Further glycosylated N-terminal of Pen-2 fails to bind to PS [54].

Biochemical experiments indicated that residues 18-38 and 58-80 form TMDs. DYSLF domain of Pen-2 (residues 90-94) at C-termini is responsible for binding to PS [56]. Incorporation of FLAG tag in C-termini of Pen-2 increases Ab42/40 ratio [57]. Furthermore, various GS modulators which lowers Ab42 were found to only bind to Pen-2. Cross-linking indicates that Pen-2 and PS1 CTF are in close proximity [58].

**Aph-1 and Nicastrin:** Glycoprotein Nicastrin is the largest subunit of the complex in type I

transmembrane orientation. However it has only one membrane spanning domain and it modulates presenilin mediated APP and notch processing [59]. Aph-1 interacts with Nct and PS CTF. Aph-1 TMD 5 (H171), 6 (H197A) were demonstrated to be vital for GS complex formation and stability [60]. Residues 245-630 are important for APP and Notch processing. This region consists of DAP domain (DYIGS and peptidase; residues 261-502), that is homologous to a tetratricopeptide repeat (TPR) domain commonly involved in peptide recognition. Leu571 in the TPR domain is involved in substrate binding [61,62].

### 1.3.7 GS regulating enzymes

Transmembrane protein 21 (TMP21, a member of the p24 cargo protein family) [63]  $\gamma$ -secretase activating protein (GASP) (He et al. 2010) has been shown to regulate GS activity. GASP increases A $\beta$  production by interacting with GS and APP. However it neither interacts with Notch nor affects its cleavage. Knockdown of GASP results in decreased A $\beta$  production in mouse. Anticancer drug imatinib inhibits GASP to reduce A $\beta$  formation in AD. However TMP21 and GASP are not associated (does not form complex) with GS.

### 1.3.8 Stepwise substrate processing in the water cavity

#### Residues in the catalytic water pore

PS TMD 6 (NTF) and 7 (CTF) harbouring the catalytic aspartate dyad are implicated to be present in the water containing cavity in GS [64]. Additionally the GxGD and the conserved PAL (pro, ala, leu) motif of CTF which is in close proximity TMD 6 is water accessible [36,65,66]. Further, TMD 1 and 9 of PS1 has been implicated to be present in the water cavity. TMD 1 is in proximity to GxGS and PAL motifs of PS CTF [67]. PAL motif is required for normal active site conformation but not for ER retention and GS complex formation [66].

The aspartate in TMD 7 (CTF) is a part of GxGD motif which is also found in other proteases [68]. There are initial substrate binding sites in PS TMD2 and 6 [69]. TMD1 of PS1 is a part of the catalytic pore [67].

APP is cleaved by GS in stepwise fashion in short intervals to get rid of hydrophobic APP from the membrane. First the cleavage occurs in  $\epsilon$ -site (Leu 49) which is very close to the membrane [70] (Fig 1.6) which releases AICD. It is followed by the cleavage of the intramembrane A $\beta$ 49 at the  $\zeta$ -site to generate A $\beta$ 46 [71]. Then GS cleaves at multiple  $\gamma$ -sites [72] giving rise to A $\beta$ 43, A $\beta$ 40, A $\beta$ 37 where A $\beta$ 40 is the major product. However these cleavages are heterogeneous giving rise to two product lines with 3 amino acid intervals major product line : A $\beta$ 49-37 (main product A $\beta$ 40) and minor product line A $\beta$ 48-39 (main product: A $\beta$ 42 and A $\beta$ 38 which are the causative agents of AD) [73]. Presence of the successive release of tri- and tetra-peptides from APP have been elucidated

[74]. It has been known that mutations at the GxxxG motif of APP decreases A $\beta$ 42 production. The GxxxG motif promotes APP dimerization, Of late, it has been found that due to steric hindrance, stepwise cleavage of dimeric APP by GS stops at  $\gamma$ -42 unlike going to  $\gamma$ -38 site. Consequently, higher amounts of A $\beta$ 42 is produced [75].

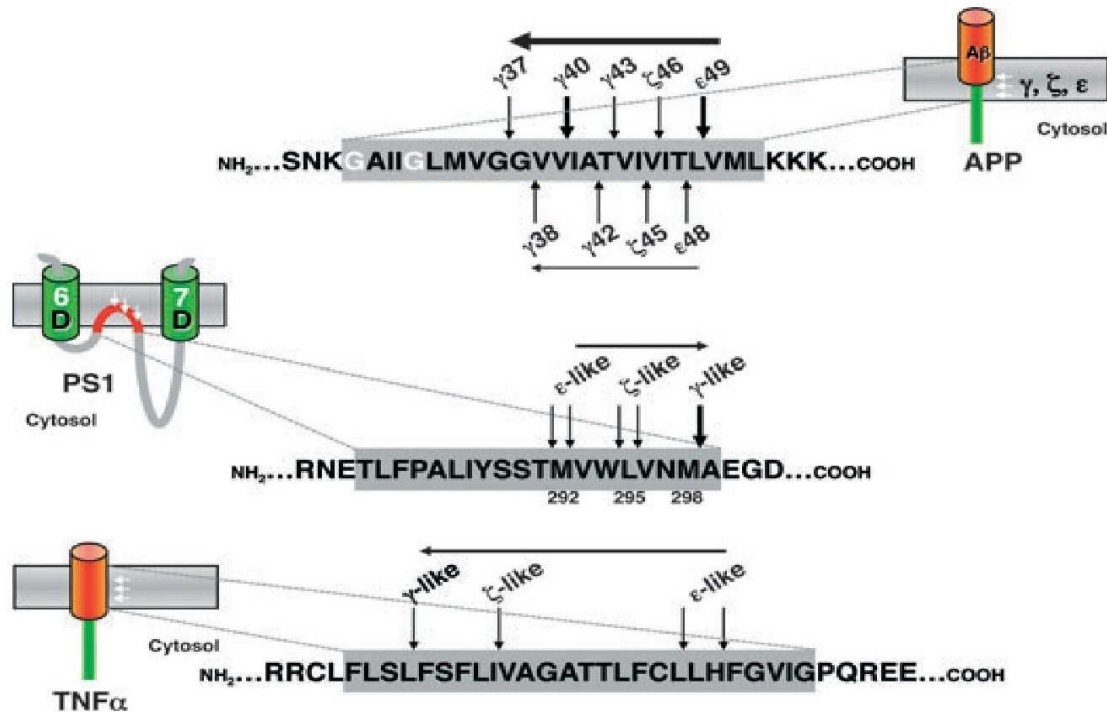


Fig 1.6 Stepwise cleavage of substrates by GxGD proteases. (Top) GS cleavage sites in the APP transmembrane domain (TMD) are shown in thick (major cleavage sites) and thin (minor cleavage sites) vertical arrows. The direction of the cleavages are given by horizontal arrows. The three step cleavage of APP from  $\epsilon$ 49 – $\gamma$ 37 gives rise to A $\beta$ 40 as a major product. In the alternative three step cleavage from  $\epsilon$ 48 – $\gamma$ 38, A $\beta$ 42 emerges as a minor product. White letters indicate the GxxxG dimerisation motif in the APP which regulate if A $\beta$ 40/A $\beta$ 42 will be main product. (middle) Similar  $\epsilon$ -,  $\zeta$ -, and  $\gamma$ -like cleavage sites during PS autoproteolysis. Numbers indicate amino acid number in PS [44].(bottom) GxGD protease SPPL2b mediated cleavage of TNF $\alpha$ . In case of PS autoproteolysis and TNF $\alpha$ , the direction of the cleavage is just the opposite [448]. Grey highlight indicates predicted TMD. Figure obtained from [5]

This stepwise cleavage of APP at  $\epsilon$ -,  $\zeta$ -, and  $\gamma$ -like sites separated by 3 amino acids has been also found during autoproteolysis of PS and cleavage of TNF-alpha by SPPL2b [44]. Similar type of step by step cleavage of Notch1 and APLP1 and CD44 has been detected [76]. Lately a coding mutation (A673T) in the APP gene was demonstrated to protect against AD. The mutation is to the aspartyl protease  $\beta$ -site of APP and reduces Ab production by 40% *in vitro* [77].

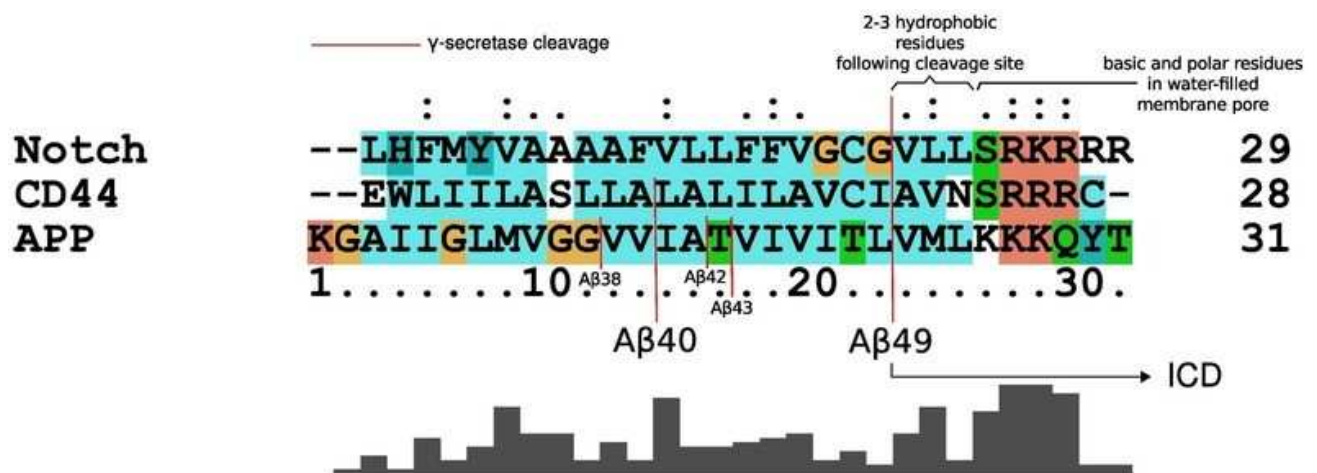


Fig 1.7: Sequence alignment of the GS substrates Notch, CD44 and APP showing cleavage sites

Kinetic studies show that FAD mutations affect the production of A $\beta$  species in three ways. FAD mutants don't show  $\epsilon$ -cleavage unlike GS inhibitors which also block Notch processing. GS modulators increases carboxypeptidase-like ( $\gamma$ ) activity of GS. These results could be useful in screening GS inhibitors [78].

### 1.3.9 Substrate recognition

The mechanism of substrate recognition has not been elucidated in detail yet. However evidences suggest that the substrate binding site is different from the active site [79]. Glu 333 of Nct has been shown to bind to substrate and participate in GS activity [80]. However there are contradictory evidences suggesting that GS is functional in absence of nicastrin [81]. Aph-1 has also been a candidate for substrate interaction prior to cleavage [82]. Another evidence suggest that the initial substrate binding site is located on presenilin near the active site [35]. Therefore, due to the close proximity of the docking site compared to the active site, any mutation near the active site i.e. in the GxGD motif has drastic effects on GS activity. Also mutations in the PAL motif results in hampering of substrate cleavage [66]. Further, the juxtamembrane, TMD and ICD can influence substrate processing by GS [83].

### 1.3.10 Effect of PS Mutations

The A $\beta$ 42/40 ratio is increased during FAD [84,85]. Even a negligible increase in the ratio can trigger AD by causing synaptic and cellular neurotoxicity. More than 180 mutations in PS1 and a few in PS2 has been related to FAD (<http://www.molgen.vib-ua.be/ADMutations>). Notch signaling is often severely affected by these mutations rather than increase in A $\beta$ 42 product line [86].

### 1.3.11 Relationship with GPCRs and miRNAs

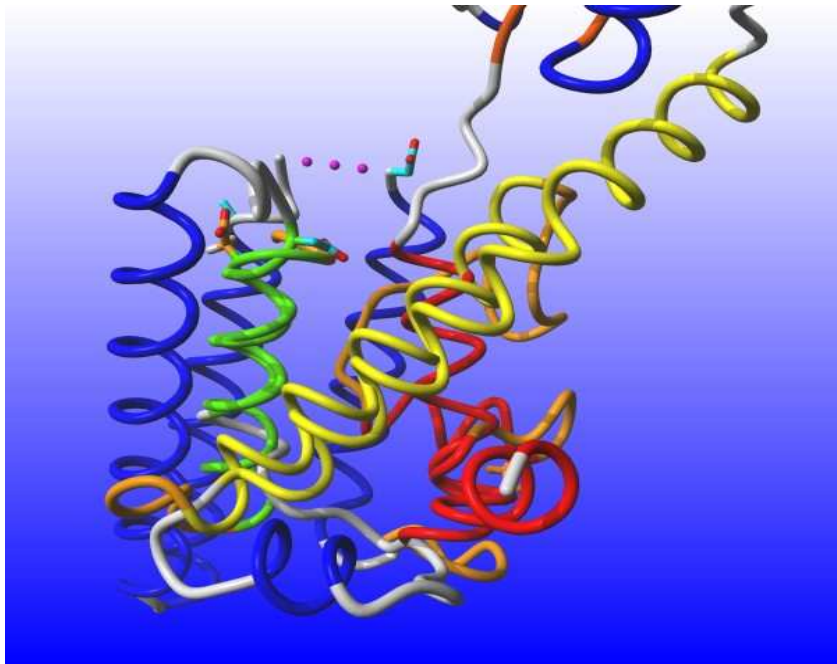
G protein-coupled receptors (GPCRs) mediate various signaling systems in neurons which are affected during AD. GPCRs can modulate  $\alpha$ -,  $\beta$ - and  $\gamma$ -secretases, proteolysis of the amyloid



precursor protein (APP) and regulation of amyloid- $\beta$  degradation. Moreover A $\beta$  has been demonstrated to rapture GPCR function. Therefore GPCRs can be potential targets for AD [87].

#### 1.4 Structure of GxGD type protease

Members of the GxGD family includes preflagellin peptidase, type 4 prepilin peptidase, presenilin and signal peptide peptidase (SPP) and signal peptide peptidase like (SPPL). Recently the 6TM structure of preflagellin peptidase Flak from *Methanococcus maripaludis* was solved which shows similarity to PS1 CTF structure (PDB: 3S0X) [88].



*Fig 1.8 Structural Alignment of PS1-CTF with Flak protease*

The archaeal site 2 protease (S2P) also has 6 TMDs. The active site containing Zn atom coordinated by two histidines is located in the middle of the lipid bilayer 14 Å up from the cytosolic surface [89]. (PDB ID: 3B4R)

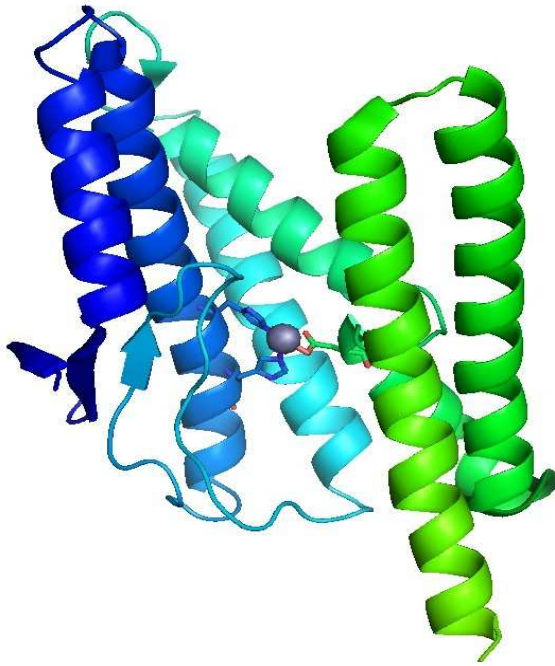
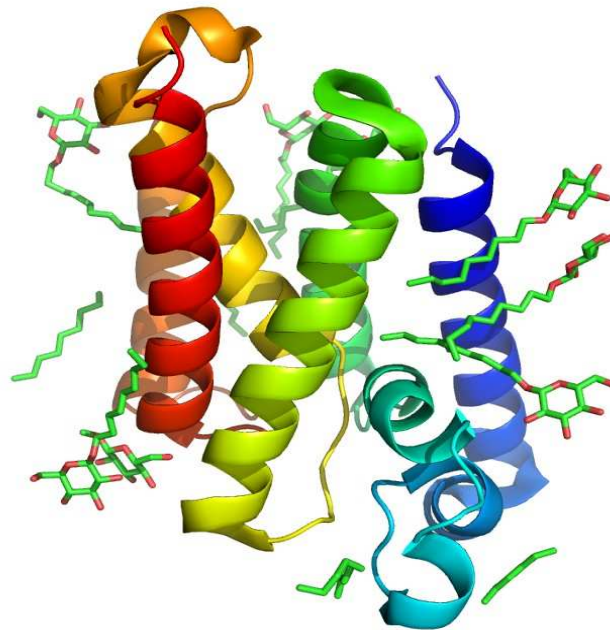


Fig 1.9 Structure of the S2P protease shown in cartoon form which has six transmembrane helices. The zinc atom (shown in van der Waal's sphere) is coordinating with His54 and His58 from helix  $\alpha 2$  and Asp148 from the N-terminal end of helix  $\alpha 4$ -C.

Fig 1.10 Structure of the GlpG rhomboid protease (PDB ID: 2IC8) [13]. Its structural analysis



reveals gating mechanism of substrate entry [15].

**Table 1.** List of selected RIP-mediated LPD signaling events (adapted from [7])

iCLiP	Substrate Protein	RIP Stimulus	LPD Signaling Function	Refs.
Presenilin/ $\gamma$ -Secretase	APP	TAG1	Nuclear signaling, suppression of neurogenesis	12
	$\beta 2$ Na channel	?, PMA	?, cell migration	22
	CD44	Loss of cell contact, PMA	Cell adhesion, nuclear signaling	45
	CD74	?	Nuclear signaling via NF- $\kappa$ B	3

<b>iCLiP</b>	<b>Substrate Protein</b>	<b>RIP Stimulus</b>	<b>LPD Signaling Function</b>	<b>Refs.</b>
			activation	
	CSF-1	CSF-1, PMA, TLR activation	?	17,6 7
	E-Cadherin	Calcium, apoptosis	?, $\beta$ -catenin signaling	38,4 0
	EpCAM	EpCAM ectodomain	Nuclear signaling, controls <i>c-myc</i> expression and cell proliferation	37
	EphrinB2	EphB	Src activation and sprouting of endothelial cell	16
	ErbB-4	EGF ligands (heregulin)	Nuclear signaling, neurogenesis	47,5 3,55
	IFN $\alpha$ R2	PMA, IFN-alpha	Nuclear signaling	54
	Fibrocystin/Polyductin	Calcium, PMA, Mechanosensation	?	19,2 1
	Il-1R2	?, PMA	?	25
	LRP1	?	?, association with adaptor proteins	42
	LRP1B	?	Tumor suppression, nuclear signaling?	33
	LRP6	Wnt3a, PMA	Wnt signaling	44
	Megalin	?, PMA	?, gene expression	31,7 3
	N-Cadherin	NMDA receptor agonists	Proteasomal-dependent degradation of CBP	39,5 2,72
	Nectin-1 $\alpha$	?, PMA	?, remodeling of cell-cell junctions	23
	Notch	Notch-Delta	Transcription factor	11
	Delta 1, Jagged		Nuclear signaling via AP-1	27,5 7
	p75 NTR	MAG	Rho activation, inhibition of neurite regeneration	13
	RPTP	Cell density	Dephosphorylation of nuclear $\beta$ -catenin, suppression of $\beta$ -catenin transcriptional activity	1
	Ryk	Constitutive. Wnt stimulates nuclear accumulation	Nuclear signaling, neurogenesis	35
	SorLA	?, PMA	Possible nuclear function	5
	Syndecan 3	bFGF, PMA, forskolin	Regulation of CASK nuclear translocation	56
SPP	HCV precore protein	?	HCV precore protein processing	43,6 5
	MHC	?	HLA-E signaling	29
	Pre-prolactin	?	CaM-dependent signaling	66
SPPL2a, SPPL2b	TNF- $\alpha$	LPS	Pro-inflammatory cytokine (IL-12) induction	14–1 5
	Bri2(Itm2b)	?	?	41

iCLiP	Substrate Protein	RIP Stimulus	LPD Signaling Function	Refs.
S2P	ATF6	ER stress, unfolded protein	Nuclear signaling, activation of the UPR	70
	CREB4	?	Nuclear signaling	58
	CREBH	ER stress, cytokines	Nuclear signaling, activation of the UPR and APR	72
	Luman/CREB3	ER stress	Nuclear signaling, activation of the UPR	32,50
	OASIS	ER stress	Nuclear signaling, activation of the UPR in astrocytes	24,46
	SREBP	Cholesterol levels	ER stress-induced bone formation	
Rhomboids	Spitz, Gurken, ?		Nuclear signaling, cholesterol biosynthesis	6,51
	Keren		EGF signaling	28,62,63
PARL	OPA1	?	Mitochondrial remodeling, anti-apoptotic	10
?	Polycystin-1	?, mechanosensation	Nuclear signaling, JNK, Wnt, STAT6 signaling	9,34

- APP, Alzheimer precursor protein; bFGF, basic fibroblast growth factor; CSF-1, colony-stimulating factor; EGF, epidermal growth factor; iCLiP, intramembrane-cleaving protease; HCV, hepatitis C virus; HLA-E, human lymphocyte antigen E; IFN $\alpha$ 2, type I interferon  $\alpha$  receptor 2; IL-1R2; interleukin-1 receptor II; LPD, liberated protein domain; LRP, low-density lipoprotein-related protein; MAG, myelin-associated glycoprotein; MHC, major histocompatibility complex; NTR, neurotrophin receptor; PARL, presenilin-associated rhomboid-like; PMA, phorbol 12-myristate 13-acetate; RPTP, receptor protein tyrosine phosphatase; S2P, site-2 protease; SPP, signal-peptide peptidase; SPPL, SPP-like; SREBP, sterol regulatory element-binding protein; STAT, signal transducers and activators of

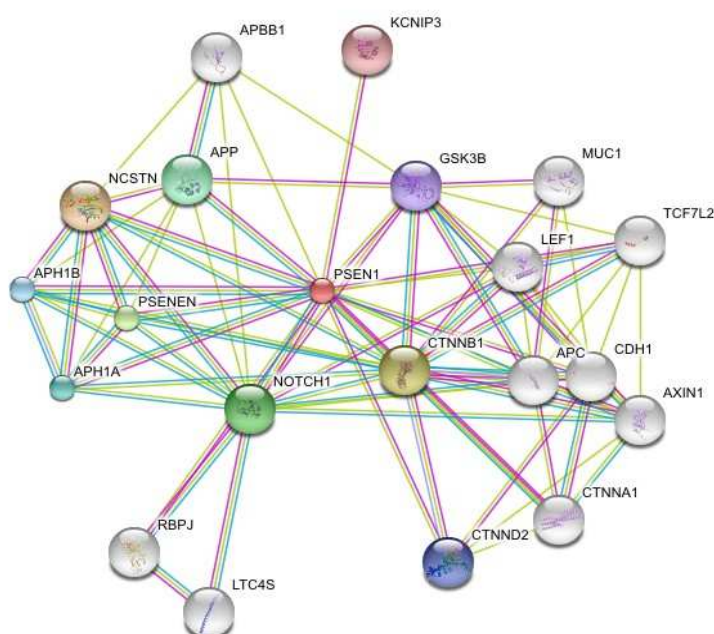


Fig 1.11 Interactions of presenilin 1 with various proteins as found by StringDB by text mining pubmed abstracts [449]

## 1.5 Amyloids and Amylome

Amyloids and prion proteins are thought to be culprits of various age related neurodegenerative diseases. However the strength and durability of specific forms of amyloids can be useful not only physiologically but also in nanotechnology. A $\beta$  peptides are formed by sequential cleavage of Amyloid precursor protein (APP) by  $\beta$  and  $\gamma$ -secretase. The function of APP is not clearly known but presumed to be involved in neuronal development. The length of A $\beta$  monomer varies from 39-43 amino acids. But A $\beta$ 40 and A $\beta$ 42 are most prevalent [90]. A $\beta$  contains hydrophobic C-terminal domain which adopts beta-strand structure and the N-terminal region can exist as an alpha-helical or beta-strand conformation depending on the environmental condition (pH and hydrophobicity surrounding the molecule) [91].

Many proteins can convert into amyloid fibrils either to comply with the physiological needs or as part of a pathological scenario. To fight against pathological amyloid states and to stop growth of particular amyloids, the prospective inhibitors of amyloid fibril formation may be helpful. Unfortunately, the structure-based drug design is hampered because amyloid proteins do not have defined structures. Nonetheless, in a recent paper [92], the Eisenberg and Baker groups described a structure-based design of such inhibitors. They demonstrated that a structure of a short segment directly engaged in fibril formation can be sufficient for the design of fibril formation inhibitors and that the computational methods may be successful in designing novel peptide-peptide interfaces. The inhibitory peptides were designed employing modeled structures of the so-called “steric zippers” which are dual  $\beta$ -sheets. One of the inhibitory peptides, consisting exclusively of d-amino acids, inhibited the formation of the tau protein tangles associated with Alzheimer’s disease [93]. Its target was a hexapeptide VQIVYK corresponding to tau protein residues 306–311. This fragment was shown to be important for fibril formation by the full-length tau protein [94,95], and fibrils formed by this fragment are similar to full-length tau fibrils. The researchers also designed a non-natural l-amino acid inhibitor of the amyloid fibril enhancing transmission of HIV. Its target was also a steric zipper structure of the GGVLVN peptide from a fragment of prostatic acid phosphatase [96]. The authors designed the specific and tight interface between the inhibiting peptide and the end of the steric zipper by maximizing the number of hydrogen bonds and hydrophobic interactions.

Eisenberg also introduced the concept of amyloids [97] defined as a large set of proteins capable of forming amyloid-like fibrils. It was suggested in this paper that the amyloid state is accessible to many more proteins than was originally thought—not only to those whose entire sequence is engaged in amyloid formation. In the classical view, in each disease of amyloid origin, one or two fibril-forming proteins were characterized, namely  $\beta$ -amyloid and tau proteins in Alzheimer’s disease,  $\alpha$ -synuclein in Parkinson’s disease, huntingtin polyglutamine stretch in Huntington’s disease, prion protein in Creutzfeldt-Jakob disease and amylin in type II diabetes [98]. Aggregates of these

proteins are toxic, highly stable, and are producing polymer-like amyloids by recruiting normal, soluble proteins [99].

Eisenberg and coworkers [97] investigated the factors that enable a protein to acquire an amyloid form. It turned out that the major factor responsible for amyloid formation is the presence of a segment in the protein that can form a tightly complementary interface with other mostly identical segments. Such interface between the segments was named “steric zipper.” It is usually created by self-complementary  $\beta$ -sheets that form the amyloid fibril. Another suggested factor is a sufficient conformational freedom of the self-complementary segment allowing for interaction with other identical segments. Eisenberg’s group examined more than 12,000 proteins whose folded, three-dimensional (3D) structures are already known. The predictions of an amyloid state were done by the modified 3D-Profile method [100] based on the crystal structure of the NNQQNY motif, known to form a steric zipper. They computationally examined proteins of three organisms: *Escherichia coli*, *Saccharomyces cerevisiae*, and *Homo sapiens*. The method identified protein segments with high tendency to form amyloid fibrils and demonstrated that a specific residue order is required for fiber formation. These segments were typically about six amino acids long and could be exposed for instance during thermal motion of the protein. It was found that 95% of the predicted amyloid-prone segments are buried within the protein, and those that are exposed are too twisted and inflexible to form a “steric zipper” with partner segments. Using bovine pancreatic ribonuclease A (RNase A) as a model system, they experimentally validated the accuracy of predictions and investigated the effect of sequence and residue composition. For instance, the FERQHM sequence was one of several segments predicted and experimentally confirmed not to form fibrils. However, when the residues of this segment were rearranged to QEMRHF, the energy of the rearranged segment fell below the formerly estimated threshold of  $-23$  kcal/mol; QEMRHF was thus predicted to form fibrils, which was subsequently confirmed by EM images. On the contrary, the fibril-forming segments QANKHI and STMSIT were rearranged to IHKAQN and ISMTTS, respectively. The rearranged sequences were predicted not to form fibrils, and it was also confirmed by experimental methods. Such shuffling experiments suggest that the tendency to form amyloid-like fibrils is strongly sequence-dependent and relatively insensitive to amino acid composition.

In earlier research it has also been shown [101,102] that many globular proteins can be converted to the amyloid state by a variety of denaturing processes, suggesting that conversion may generally be applicable to all proteins. The self-association of peptides and proteins into well-ordered supramolecular structures is of central importance in normal physiological processes such as the assembly of collagen fibrils [103,104], actin filaments [105] but also in pathophysiological cases [106]. Integration of old and new techniques and development of novel methods of nanoscience can provide powerful opportunities to increase our understanding of processes underlying amyloid-related disorders [107]. Until recently, it was commonly believed that amyloid formation is

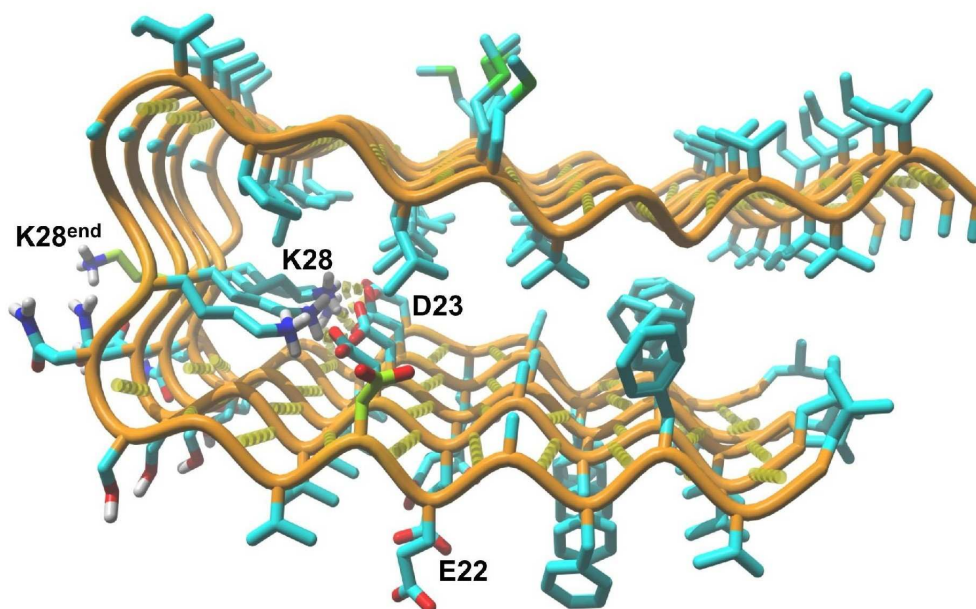
a feature of only a tiny fraction of proteins. Not all proteins, however, form amyloids because in most cases these potentially harmful segments are hidden deep inside the protein structure and are kept under control. Such behavior can be of evolutionary origin suggesting that evolution treats amyloids as a fundamental threat. The presence of different kinds of amyloids have been confirmed in some of the most common age-related diseases, so one can suppose that the accumulation of amyloid is unavoidable during aging. Sometimes, the presence of amyloid deposits does not give rise to neurodegenerative symptoms, indicating that amyloid fibrils do not cause the onset of disease. Therefore, one of the hypotheses suggested that the oligomeric intermediates are the toxic species while the fibrils are detoxification products [108]. Fibrils are not the only shape taken by amyloids especially during the nucleation process. For instance, spheroidal oligomeric species have been demonstrated for  $\alpha$ -synuclein—they are thought to be responsible for cytotoxicity towards the neuronal cells observed in Parkinson's disease [109,110].

On the basis of current research, it was proposed by Eisenberg [97] that the amyloid state is more like a default state of a protein especially in the absence of specific protective mechanisms such as chaperoning. Proteins that are not correctly folded and less protected (by chaperoning and/or disposal mechanisms) are predisposed to become amyloids. The amyloid-associated diseases that are known so far probably involve only the most vulnerable human proteins. Many research groups try to find ways to supplement or boost the protective mechanisms, in the hope of treating or preventing the original cause of amyloid-linked diseases. Even a subtle pharmacological interference in the process of amyloidogenesis might have a major effect on the disease and even on ageing in general. On the other hand, one can enhance the natural protective mechanisms that stabilize a protein. A review of potential strategies for tackling protein aggregation and the toxicity associated with it has been published by Bartolini and Andrisano [111]. However, the complexity of the aggregation processes and other related events account for the fact that no effective treatments for these disorders are currently available. Studies of the structures of amyloids and mechanisms of amyloid formation should unveil new molecular targets for potential anti-neurodegenerative drugs. Although the three characteristic stages of nucleation-dependent fibrillation—seed formation, accelerated fibrillar growth, and the stationary phase—have been examined separately, additional studies are required to unambiguously uncover the mechanism of amyloidogenesis.

### 1.5.1 Molecular Structures of Amyloids

Amyloid fibrils represent an energetically stable state of many proteins and peptides. Basically, amyloid fibers are a bundle of highly ordered filaments composed of ladders of  $\beta$ -strands that are placed perpendicular to the fiber axis and are arranged in hydrogen-bonded  $\beta$ -sheets [112]. Amyloid fibers have a diameter of about 7–10 nm and can be up to several micrometers long. In cross sections, amyloid assemblies appear as hollow cylinders or ribbons. The measurements of amyloid

fibers revealed that their strength is comparable to that of steel while their mechanical stiffness matches that of silk [113]. In general, amyloid structures attain their stability through non-covalent bonds, mainly hydrogen bonds stabilizing the  $\beta$ -sheets, but also through hydrophobic and  $\pi$ - $\pi$  stacking interactions of the side chains. The frequent occurrence of aromatic residues in short amyloid-related peptides suggests that  $\pi$  stacking may play a role in speeding-up the self-assembly process by providing geometrical constraints that promote directionality and orientation of the growing fibril. The importance of hydrogen bonds is especially seen in glutamine- and asparagine-rich proteins which form amyloids. Extended sequences of repeated glutamine (or asparagine) units are related to several amyloidoses such as Huntington's disease and spinocerebellar ataxia, and also to the aggregation of yeast proteins into prions.



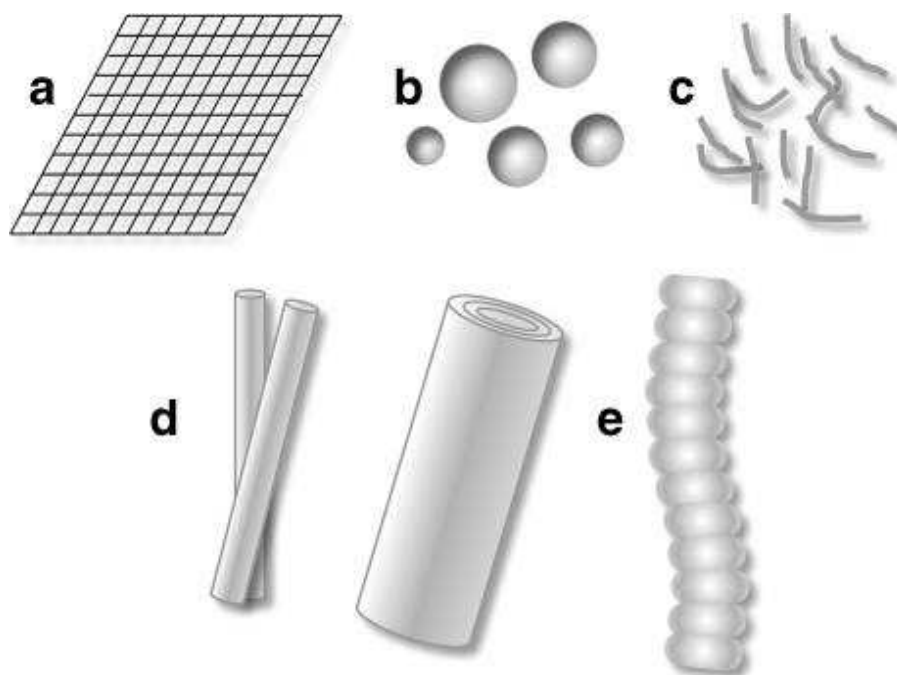
*Fig 1.12 The structure of fragment of  $\beta$ -amyloid ( $A\beta_{42}$ ) obtained by NMR methods (PDB code 2BEG). The salt bridge K28-D23 is linking adjacent  $\beta$ -sheets, therefore, the residues K28 and D23 from terminal strands are unpaired (in yellow-green). Hydrogen bonds shown as dashed yellow cylinders.*

The three-dimensional structure of the fibrils comprising  $A\beta_{42}$  (Protein Data Bank code 2BEG) was obtained using quenched hydrogen/deuterium exchange NMR in solution, while the  $\beta$ -sheet arrangement was taken from previous solid-state NMR studies of this structure. Residues 18–42 form a  $\beta$ -strand–turn– $\beta$ -strand motif while residues 1–17 are disordered and could not be detected. The parallel  $\beta$ -sheets are formed by residues 18–26 ( $\beta_1$  strand) and 31–42 ( $\beta_2$  strand). The repeating structure of a protofilament requires two monomers because of the salt bridge D23-K28 formed between adjacent monomers. This interaction pattern leads to the formation of partially unpaired  $\beta$ -strands at the ends of the  $A\beta_{42}$  fibrils (Fig. 1.12). Such unpaired ends explain the specific shape of these fibrils and could be a target for inhibitors of fibril growth [114]. The salt bridge and also the hydrophobic interactions of the side chains keep the structure rigid and compact despite the repulsion



between the charged residues E22, D23, and K28 from the adjacent  $\beta$ -strands.

Amyloid fibrils can be formed by different proteins and usually contain a common cross- $\beta$  spine. Glutamine repeats were first suggested to act as “polar zippers” joining monomeric units together and propagating the amyloid fibrils. The structure of the fibril-forming segment, GNNQQNY, of the yeast prion protein Sup35 has been recently revealed by crystallography [115]. It is formed by a pair of  $\beta$ -sheets, with the facing side chains of the two sheets locked together in an interdigitated way forming a so-called “dry steric zipper” (Fig. 1.14a). Eisenberg and coworkers [115] reported dozens of other segments from fibril-forming proteins that are able to form amyloid-like fibrils on their own. The segments from the  $\beta$ -amyloid and tau proteins, the PrP prion protein, insulin, islet amyloid polypeptide (IAPP), lysozyme, myoglobin,  $\alpha$ -synuclein, and  $\beta$ -2-microglobulin were analyzed. The obtained structures are characterized by structural features that are shared, at the molecular level, by all the proteins studied but some variations in the atomic architecture of the amyloid-like fibrils can provide some clues on their origin and the mode of growth. In the GNNQQNY amyloid, the peptide strands are parallel, and the Asn and Gln residues form regular rows connected by hydrogen bonds in addition to the hydrogen bonds in the  $\beta$ -sheet. The hydrophilic character of these residues and their length make the steric zipper interface highly interdigitated. In the other amyloid formed from the AILSST peptide (Fig. 1.14b), the strands are antiparallel, and the steric zipper interface is formed mostly by hydrophobic residues Ile and Leu. The hydrogen bonds between side chains of serine residues are bridged by water molecules.



*Fig 1.13: Different forms of amyloids: a squared plates, b nanospheres, c hydrogels, d tubular structures—single-walled and multi-walled tubes, e fibrils*

According to [115] there are eight types of the steric zipper interfaces classified according not only to the orientations of their strands (parallel or antiparallel) but also faces (face-to-face or face-to-back

arrangement) and the up or down orientations of the edges of the strands. So identical peptides can form different polymorphic structures characterized by distinctive phenotypes. New polymorphic crystal structures of segments of the prion and other amyloid proteins [116] proved to be useful for elucidating the structural mechanisms of different modes of fibrillation. Additionally,  $\beta$ -sheets formed by the same segment of a protein can reveal alternative packing arrangements (polymorphs). Such polymorphism can be responsible for enduring conformations capable of “encoding” prion strains. Such transfer of protein-encoded information into prion strains involves sequence specificity and recognition by means of noncovalent bonds.

Amyloid fibril formation is considered to be a signature of neurodegenerative processes. The exact processes leading to cellular degeneration remain unknown although several amyloid-involving mechanisms have been proposed [117]:

- amyloids occupy the extracellular space and destroy the structure of cells and tissues,
- amyloid fibrils destabilize cell membranes,
- heavy metals incorporate into amyloids and generate reactive oxygen compounds which affect cellular functions,
- some proteins essential for cell survival are trapped in protein aggregates.

In a recent review, Zerovnik et al. [118] classified the mechanisms by which proteins undergo ordered aggregation into amyloid fibrils:

- templating and nucleation;
- linear, colloid-like assembly of spherical oligomers,
- domain swapping.

The local environment and inter- and/or intra-molecular interactions may have a significant influence on the conformation of certain amino acid residues. Therefore, even small variations in pH, temperature, and ionic strength could induce changes in the conformational propensities of these residues (leading to a different secondary structure) including their ability to aggregate.

Some proteins forming amyloids, for instance  $\alpha$ -synuclein which contributes to the formation of intracellular Lewy bodies in Parkinson's disease [119], can exist without a defined structure. It was postulated that exogenous  $\alpha$ -synuclein fibrils induce the formation of Lewy body-like intracellular inclusions [120]. Other proteins with an unordered structure are the IAPP in type II diabetes [121] and  $\beta$ -amyloid in Alzheimer's disease [122]. Such an unfolded structure allows the protein to be rather easily self-assembled into fibrils. On the other hand, some amyloidogenic proteins preserve their 3D structure until the actual fibrillation [102]. This group of proteins includes  $\beta$ -2 microglobulin identified in dialysis-related amyloidosis [123], huntingtin in Huntington's disease [124], immunoglobulin VL domain in light-chain amyloidosis [125], lysozyme in hereditary systemic amyloidosis [126], prion protein in Creutzfeldt-Jakob disease [127], and transthyretin in

senile systemic amyloidosis [128]. However, regardless of the initial structure, the amyloid fibrils obtained from different amyloidogenic proteins and peptides are very similar and adopt a cross  $\beta$ -sheet conformation [101] even though these proteins and peptides share rather little amino acid sequence similarity. It was known that even all  $\alpha$  (protein composed of  $\alpha$ -helices only) or mixed  $\alpha/\beta$  protein types can form  $\beta$ -sheet fibrils. Therefore, it was tempting to suggest that when elucidated for a given protein in a particular disease, the molecular mechanism of amyloidogenesis will apply to other proteins and amyloid-related diseases. However, it became gradually recognized that amyloid fibrils exist in multiple fibrillar forms and exhibit so-called fibrillar polymorphisms. Even a single amyloidogenic protein can create multiple forms of amyloid fibrils depending on the conditions in which fibrillation occurred. This may indicate that amyloidogenesis can proceed via multiple mechanisms. Various types of possible amyloids structures are shown on Fig. 1.13.

The conversion and aggregation of proteins from their soluble states into well-organized fibrils is associated with a wide range of conditions, usually pathological, including neurodegenerative diseases and amyloidoses. Although a conformational change of the protein native state is generally necessary to initiate aggregation, it was shown that a transition across the large unfolding energy barrier is not essential and that the aggregation may be initiated from locally unfolded states that become accessible, for example, via thermal fluctuations occurring under physiological conditions [102]. Conformational states thermodynamically distinct from the native state, but structurally similar to it, can be easily accessed from the native state through thermal fluctuations. These states are separated from the native state by a relatively low energy barrier. They are therefore only transiently populated under physiological conditions, yet they can be sampled more frequently than the entirely unfolded state (global unfolding) or a partially folded state. The existence of such conformational states can be deduced from the observation that, under physiological conditions, the amide hydrogen atoms buried in the interior of a native protein can exchange with the solvent hydrogen atoms more rapidly than it could be expected from the rate of protein unfolding. The possibility of sampling of such partially unfolded states is also confirmed by long molecular dynamics simulations.

Amyloid self-polymerization is also the basis of the “protein-only” hypothesis for the mechanism of prion infectivity. The infectious prion conformation replicates itself in a host by pairing with the host protein and forcing it into the infectious, fibrillar conformation. It was found that amyloids, including  $\beta$ -amyloid, can also be infectious like the PrPSc prion protein. Data showed that  $\beta$ -amyloid, which is associated with Alzheimer’s disease, behaved like an infectious agent when injected into the brain of a mouse. The same mechanism was suggested in the case of other diseases in which amyloid forms of proteins were detected [129]. A self-complementary “steric zipper” structure identified in protein fibrils allows them to tangle very tightly with an identical segment exposed on another protein. Several of these segments are needed to seed, or nucleate, an amyloid. Segments attach to one

another and form fibrils. As they grow, fibrils are fringed by the remnants of the host protein segments (Fig. 1.15). Eventually, this developing fibril breaks to form two smaller fibrils, each of which starts to grow at both ends again. The nucleation events are rare but once the fibril is formed its spreading is fast. Ohhashi et al. [130], based on mutational and biophysical analyses, proposed that before fiber formation, the prion domain (Sup35<sup>NM</sup>, consisting of residues 1-254) of yeast prion Sup35 forms oligomers in a temperature-dependent reversible manner. Experiments revealed that “non-native” aromatic interactions outside the amyloid core drive oligomer formation by bringing together different monomers, which leads to the formation of new amyloid cores. In this way, the transient non-native interactions in the initial nucleus are responsible for the diversity of amyloid conformations.

### 1.5.2 Kinetics of the Growth of Amyloid Fibrils

Using quantitative measurements of protein aggregation rates, Buell et al. [131] developed a kinetic model of a conversion of a protein from a soluble to a fibrillar form which shows that there is a single free energy aggregation barrier controlling the addition of protein molecules into amyloid fibrils. Other characteristics of the aggregation process are natural consequences of finite diffusion times. These findings suggest that this process does not follow a simple chemical mechanism, but rather operates in a way analogous to the multitrajectory (landscape) models of protein folding defined by stochastic dynamics on the surface of the potential energy of the system. Another kinetic study [132] was based on quantitative quartz crystal microbalance measurements of the kinetics of the growth of amyloid fibrils in crowded environments. Such environments strongly modify the association of components, through attractive entropic interactions such as the depletion pressure that results from the entropically favorable overlap of the regions surrounding two aggregating particles. The complex effects of macromolecular crowding on the growth of amyloid fibrils can be described on the basis of established physical principles using a combination of osmotic effects and entropic interactions. Within this framework, it was possible to predict the aggregation susceptibility of many proteins with different structural properties. Campioni et al. [133] described two types of oligomers formed by the HypF-N protein (91-residue N-terminal domain of *E. coli* HypF) that are morphologically similar, as detected with atomic force microscopy and thioflavin T assays, though one is benign when added to cell cultures, whereas the other is toxic. They found that a lower degree of hydrophobic packing is correlated with a higher ability to penetrate the cell membrane and to cause an influx of calcium ions. It suggests that structural flexibility and hydrophobic exposure are primary determinants of the ability of oligomeric assemblies to cause cellular dysfunction and its consequences such as neurodegeneration. A broad review on aggregation kinetics and mechanisms of fibril formation was prepared by Morris et al. [134]. By employing an extensive mathematical framework, the authors revealed different aspects of nucleation, growth, and disintegration of various

amyloid intermediates.

### 1.5.3 Specific Mechanisms of Fibrillation

To explain the process of  $\alpha$ -synuclein amyloidogenesis, a specific mechanism named double-concerted fibrillation, corresponding to the prevailing nucleation-dependent fibrillation model, was introduced [135]. According to the double-concerted fibrillation, the amyloid fibril formation is achieved via two consecutive, concerted associations of monomers and the subsequently formed oligomeric granules. These newly formed oligomeric species act as units for fibril formation and subsequent growth in the absence of a template [117]. Template-dependent fibrillation requires a pre-existing fibril to which the incoming protein monomers or granules can attach if, due to a conformational change, they match the structure of the template. The fibril is extending, and the subsequent assembling step requires the exposure of the interactive domains of the protein to facilitate further molecular self-assembly. Template-dependent fibrillation is the most appropriate mechanism to study the infectivity of prion proteins. Prion protein (PrPC) is anchored to the cellular surface via the glycosylphosphatidylinositol moiety. Its conformational change into another structural entity (PrPSc) is associated with the occurrence of transmissible spongiform encephalopathies (a group of prion diseases). Exogenous PrPSc directs the conversion of PrPC into PrPSc conformation by acting as a template. In a template-independent fibrillation, the amyloidogenic conformations of building units are induced (by physical or chemical influences) before the main molecular assembly occurs. Polymorphism of amyloids, reflected by the existence of various types of amyloid fibrils, especially in the presence of specific ligands, is achieved via multiple pathways. The natively or partially unfolded amyloidogenic proteins are at a high-energy state, but increased conformational entropy could allow the self-interactive conformers to be stabilized. Initial stable seed formation is by no means privileged since the production of an oligomeric nucleus is an entropically expensive process which needs to be overcome by an enthalpic advantage [117].

Another mechanism, the 3D domain swapping, has been suggested to explain the development of protein oligomer assembly of cystatins and stefins [118]. These small globular proteins (11–13 kDa) are part of a large family of cysteine proteinase inhibitors which are also linked to amyloid diseases. The process of domain swapping is rate limiting for the initiation of amyloid fibril formation because of a high energetic barrier in this process. Nevertheless, it was suggested that, in principle, any protein is capable of oligomerization by 3D domain swapping [136]. Guo and Eisenberg [137] proposed the term “run-away domain-swapping” for a process of continuous domain swapping. Wahlbom et al. [138] used the term “propagated domain-swapping” to describe a similar process of continuous domain swapping in the formation of cystatin C prefibrillar oligomers and fibrils.

Apart from the oligomeric species formed on the route to mature fibrils the off-pathway oligomers are also formed. They are the dead ends of an alternative folding pathway because they are incapable

of converting directly to fibrils and substantially slow fibril formation. The OFF model for amyloid formation was first described by Pallitto and Murphy [139]. In this model, denatured monomers are refolded into either stable monomers or dimers or less stable dimeric intermediates which can form non-fibrillar oligomeric forms. These initial steps are followed by a cooperative assembly of the fibril-prone dimeric intermediates into a nucleus from which the protofibrils originate.

#### 1.5.4 Conformationally Distinct Amyloid States

Since it was known that the amino acid position specifically contributes to protein oligomerization, Maji et al. [140] performed amino acid substitution to determine the distribution frequency of the A $\beta$  oligomer. The substitutions were done at positions 1, 10, 20, 30, and 40 (for A $\beta$ 40) or 42 (for A $\beta$ 42). The effects of these mutations were probed using circular dichroism spectroscopy, thioflavin T binding, electron microscopy, and other techniques. All peptides displayed a transition from a random coil to  $\alpha/\beta$  and to all- $\beta$  structure, but substitution-dependent changes in the kinetics of assembly and the complexity of conformers were observed. The ability of a single substitution (Tyr in position 1) to alter the A $\beta$  assembly kinetics and the oligomer frequency distribution suggests that the N-terminus is also involved in the oligomerization process and that, most probably, there is a competition between the N- and C-termini to form a stable complex with the central hydrophobic cluster. Additionally, recent electron microscopy and AFM data for A $\beta$ 40 suggest that dimerization and subsequent monomer attachment are processes in which significant conformational changes occur in the monomer. It was also found that dimers were threefold more toxic than monomers, and tetramers were about 13-fold more toxic [141].

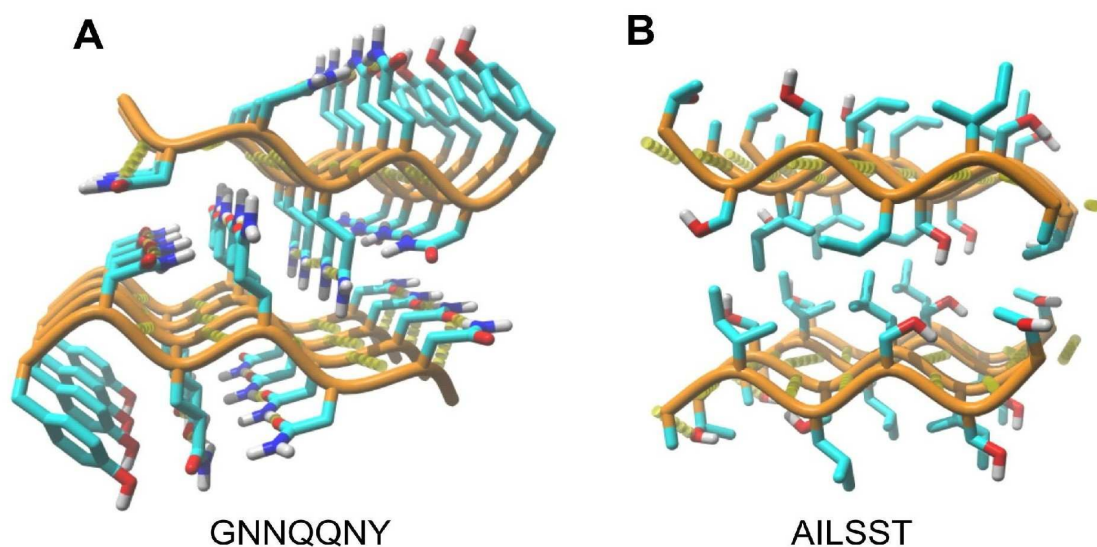


Fig 1.14 The crystal structures of amyloids (A) GNNQQNY (PDB ID: 2OMM) – adjacent  $\beta$ -strands are parallel, and (B) AILSST (PDB ID: 3FOD) – adjacent  $\beta$ -strands are antiparallel. Hydrogen bonds shown as dashed yellow cylinders.

Using mass spectrometry and ion mobility spectrometry, Bowers and coworkers [142] investigated a mixture of A $\beta$ 40 and A $\beta$ 42. A heterooligomer was formed composed of equal parts of both forms of A $\beta$ . These mixed species comprise an oligomer distribution extending to tetramers, similar to the structures created by A $\beta$ 40, whereas A $\beta$ 42 alone produced longer oligomers (dodecamers) indicating that A $\beta$ 40 inhibits oligomerization of A $\beta$ 42. In solution, A $\beta$ 40 and A $\beta$ 42 adopted similar random coil structures; however, A $\beta$ 42 was significantly more neurotoxic and formed amyloid fibrils much more rapidly than the shorter form of A $\beta$ . Although amyloid formation is triggered by a transient nucleus, the mechanism by which the initial nucleus is formed and allows the protein to acquire a specific amyloid conformation is still unclear. The observation that A $\beta$ 40 and A $\beta$ 42 self-assemble via different pathways put forward the A $\beta$ 42 dodecamers as candidate primary toxic species in Alzheimer's disease [143].

If mutations in sequence or changes in environmental conditions elicit partial unfolding of the native state of a protein, the protein will tend to aggregate, sometimes into fibrillar structures. The metastable, partially unfolded states that precede the aggregated states of proteins are of special interest because of their specific features and especially because of increased toxicity. It was found that protein aggregation is favored by conditions that promote stable intermolecular interactions, particularly the hydrogen bond formation. Calamai et al. [144] showed that human muscle acylphosphatase is able to form both fibrillar and non-fibrillar aggregates with a high  $\beta$ -sheet content from partially unfolded states with very different structural features due to the use of different destabilization factors: urea or increased temperature followed by incubation in the presence of different concentrations of 2,2,2-trifluoroethanol (solvent that has been found to promote aggregation of other polypeptides, including the natively unfolded A $\beta$  peptide). The same amino acid sequence can give rise to several conformationally distinct amyloid states. To address this puzzle, Ostapchenko et al. [145] studied two amyloid states of the prion protein (referred to as R- and S-fibrils). The obtained results suggested that the energy landscape for protein folding and assembly contains several close to global free-energy minima: one of which is occupied by the native state and the remaining ones by the amyloid states. The transmissible form of prion disease can be induced in wild-type animals by inoculation with R-fibrils while S-fibrils failed to induce the prion disease. Recently, an apparent generation of toxic prions (PrP<sup>Sc</sup>) in normal brain tissue in the presence of metal (steel wires) has been discovered. The metal catalyzed de novo formation of PrP<sup>Sc</sup> from a normal cellular prion protein [146]. Alternatively, metal surfaces might concentrate the already existing PrP<sup>Sc</sup> to the extent that it became quantifiable by the cell assay.

### 1.5.5 Molecular Simulations of Amyloids

The long time scale in which the aggregation takes place is prohibitive for molecular dynamics (MD) simulations. However, some structural and dynamic features of amyloids were investigated using

coarse grain protein models and specific MD or Monte Carlo procedures. Urbanc et al. [147] elucidated the structural characteristics of oligomers of A $\beta$ 40 and A $\beta$ 42 and of their mutants. They simulated oligomer formation using discrete MD with a four-bead protein model (the backbone is represented by three beads corresponding to the amide, alpha-carbon, and the carbonyl groups; the side chain, with the exception of glycine, is represented by only one bead). For the peptides under study, the characteristic oligomer size distributions were obtained, which were in agreement with experimental findings. A $\beta$ 42 had a high propensity to form pentameric and hexameric structures that could self-associate into higher-order oligomers. Structural analysis revealed that the C-terminal region played a dominant role in A $\beta$ 42 oligomer formation, whereas A $\beta$ 40 oligomerization was primarily driven by intermolecular interactions among the central hydrophobic regions. The N-terminal region (2)AEF played a prominent role in A $\beta$ 40 oligomerization but did not contribute to the oligomerization of A $\beta$ 42 or the mutants.

Studies conducted *in vitro* and *in vivo* suggest that administration of flavonoids, compounds naturally present in many foods including wine and tea, can prevent and reverse A $\beta$  aggregation, but the mechanism of their action is unknown. Lemkul and Bevan [148] employing atomistic, explicit solvent MD simulations investigated the mechanism of A $\beta$  fibril destabilization by morin which is one of the most effective anti-aggregation flavonoids. They used a model of mature A $\beta$  and through the course of 24 simulations found that morin could bind to the ends of the fibrils to block the attachment of an incoming monomeric peptide and can penetrate into the hydrophobic core to disrupt the D23-K28 salt bridges. It also modified the backbone hydrogen bonding.

The stability of A $\beta$ 42 fibrils and thermodynamics of peptide dissociation were investigated in [149] using all-atom molecular dynamics simulations and pulling one monomer from the pentameric protofibril of A $\beta$ 42. Results indicated that the presence of water molecules around the D23-K28 salt bridge is crucial to protofibril stability. The extent of packing between hydrophobic residues regulates the level of hydration in the core of the protofibril and thus rigidifies the D23-K28 salt bridge. Such studies explore the mechanism of destabilization of amyloid aggregates which may be important because numerous studies have found that the insoluble fibrillated form of the peptide also contributes to neurotoxicity, although the principal toxic species in Alzheimer's disease are believed to be the soluble, oligomeric aggregates of A $\beta$ . Membrane disruption and increased ion conductance have been observed *in vitro* in the presence of A $\beta$ , and it is assumed that the same phenomena occur in the brain neurons of Alzheimer's disease patients. Simulations of A $\beta$  in a membrane bilayer revealed how the peptide interacts with the surrounding lipids and to what extent it affects lipid behavior and contributes to membrane damage. The results showed that A $\beta$ 40 is capable of disordering the nearby lipids, as well as of decreasing the thickness of the membrane. During simulations the peptide unfolded and finally acquired a disordered, extended conformation allowing for extensive electrostatic and hydrogen bonding interactions with lipids [149]. The stability and



conformational dynamics of trimeric and pentameric full-length A $\beta$ 42 peptides were investigated by Masman et al. [150] for the purpose of defining structural elements influencing their stability. The N-terminal part not detected in NMR was treated as a disordered domain. The models of the oligomer were stable during 100-ns simulations while the  $\beta$ -strand acquired a characteristic twist which facilitated a compact packing of the side chains from the neighboring  $\beta$ -sheets. It seems that the hydrophobic core comprising the  $\beta$ 2 fragment of the oligomer  $\beta$ -sheet is a stabilizing element in the process of A $\beta$  aggregation. Destabilization of this crucial  $\beta$ -sheet fragment emerges as a prospective target for anti-amyloid drugs.

### **1.5.6 Amyloid Can Be Beneficial for Cells and Also Convenient for Engineers**

Amyloid fibrils are cross- $\beta$ -sheet structures that are primarily associated with several neurodegenerative diseases. However, amyloid is also a fundamental nonpathological protein structure (or conformation) utilized by organisms from bacteria to humans. The cross- $\beta$ -sheet motif is composed of intermolecular  $\beta$ -sheets arranged along the fibril axis with the  $\beta$  strands aligned perpendicularly to the fibril axis. Amyloid fibril formation also provides biologically important entities termed functional amyloids [151] that are present in silkworms [152,153] and in mammalian skin [154]. It is also known that pituitary hormones are functioning in an amyloid state. Riek and coworkers [155] found that peptide and protein hormones in secretory granules of the endocrine system are stored in an amyloid-like conformation composed of cross- $\beta$ -sheets. Thus, functional amyloids in the pituitary and other organs can contribute to normal cell and tissue physiology. The hormone amyloids are stored inside the granules, an “inert” membrane container, and the amyloid fibrils dissociate only upon secretion. Additionally, the amyloid aggregation of these hormones must be highly regulated. This regulation may include the processing of prohormones that aggregate more slowly than their hormone counterparts [156] or require the presence of helper molecules to induce aggregation; the latter was demonstrated for prolactin, which lacks a prohormone stage.

Amyloid, a fibrillar quaternary structure, was first discovered in the context of human disease and tissue damage. Therefore it was long thought to be detrimental to the host. However, recent studies have identified functional amyloid fibers in bacteria, fungi, insects, invertebrates, and humans. Nevertheless, physiological amyloidogenesis requires tight regulation to avoid toxicity of the produced amyloids. Diverse physiological applications of amyloids can change our views on the potential treatment of amyloid diseases [151]. The discovery of native amyloids in mammals provides a key insight into the molecular basis of both the physiological and pathological role of amyloids.

The examples of useful amyloids include fungal prions, which are involved in prion replication, the amyloid protein Pmel17 which is involved in biosynthesis of the pigment melanin in mammals, and the factor XII protein of the hemostatic system which is activated by amyloid. The Pmel17 protein

forms amyloid fibrils that act as a template and accelerate the covalent polymerization of small reactive molecules into melanin—a critically important biopolymer that protects against a broad range of cytotoxic insults including UV and oxidative damage. The Pmel17 amyloid also appears to play a role in diminishing the toxicity associated with melanin formation by sequestering and minimizing diffusion of highly reactive melanin precursors [156]. The silkworm chorion protein is also a natural protective amyloid. This is the major component of the eggshell, a structure with extraordinary physiological and mechanical properties. Other natural, protective amyloids are fish chorion, the hydrophobins, and the antifreeze protein from winter flounder.

The phenomenon of the self-assembly of molecules is more and more frequently exploited to invent new supramolecular structures and materials inspired by biological systems such as novel biocompatible polymeric structures with excellent physicochemical properties for new biomedical and industrial applications [153]. A variety of protein and peptide molecules with various amino acid sequences form highly stable and well-organized amyloid assemblies under diverse conditions. They display phase states ranging from liquid crystals to rigid nanotubes. The potential applications of these supramolecular assemblies can be broader than those of synthetic polymers since one can easily introduce biological function in addition to their mechanical properties [157]. Self-assembly is a powerful mechanism for organizing molecular building blocks into complex structures and aromatic groups can facilitate this process [158,159]. For example, the Phe-Phe dipeptide motif from Alzheimer's disease  $\beta$ -amyloid protein was able to self-assemble into peptide-based nanotubes [160]. The Phe-Phe peptide is of special interest due to its ability to form ordered nano-assemblies of unique physical, chemical, and mechanical properties [160,161]. It was shown that the thermal stability of diphenylalanine peptide nanotubes is significantly higher than that of a nonassembling dipeptide, dialanine. In addition to thermal stability, the peptide nanotubes were chemically stable in many organic solvents. Other aromatic dipeptides can also self-assemble into ordered structures such as tubes, spheres, plates, and hydrogels [161,162]. Moreover, Phe-Phe nanotube-based electrochemical biosensors have shown a large increase in their sensitivity upon the modification of the electrode surfaces with the forest-like nanotube arrays [163,164]. Such bio-inspired materials can be composed of chemically synthesized biomolecules. In the recent work [165], diphenylalanine nanotubes have been used to modify carbon electrodes, by physical vapor deposition of peptide nanotubes, of the electrochemical energy storage devices called supercapacitors. The structural motif of Phe-Phe forms discrete and stiff nanotubes that can be used for production of discrete nanowires with a long persistence length. The same dipeptide building block, made of d-phenylalanine, resulted in the production of enzymatically stable nanotubes [160]. It was shown that a non-charged peptide analogue, Ac-Phe-Phe-NH<sub>2</sub>, self-assembled into similar tubular structures as did diphenylalanine. A similar peptide, diphenylglycine, self-assembled into ordered nanospherical assemblies. Other homo-aromatic dipeptides, in which phenyl side chains were modified with halogen atoms,

additional phenyl groups or by alteration of the phenyl groups and naphthyl groups, or by nitro substitutions, were also investigated. In all cases, the well-ordered nanostructures were formed in the shape of tubular, spherical, and two-dimensional structures [161]. Peptide-based nanostructures represent nano-objects of particular interest, as they are biocompatible, can be easily synthesized in large amounts, decorated with functional elements, and used in various biological and non-biological applications.

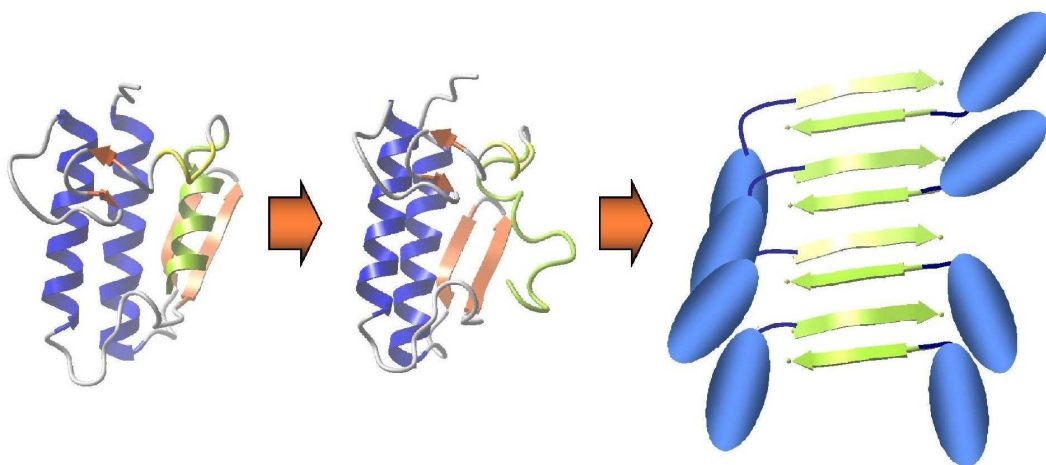
The significant thermal and chemical stability of the peptide nanotubes could be potentially useful in microelectronics and microelectromechanics as well as for fabrication of functional nanotechnological devices [166]. Amyloids have unusual properties, for instance, rigidities varying over four orders of magnitude depending on the nature of intermolecular forces. The major contribution to their rigidity stems from a generic interbackbone hydrogen bonding network that can be modulated by the variability of side chain interactions [167]. Especially the aromatic residue, side chain interactions play a role in Phe-Phe and related peptide nanotubes and in biological processes such as collagen self-assembly which involves the hydrophobic interactions of Tyr and Phe residues within the C-terminal chain [82, 83]. Usage of bionanostructures in industrial applications requires precise control over self-assembly of monomeric units and the ability to scale up production of these materials. A significant challenge is to control the formation of large, homogeneous arrays of bionanostructures on macroscopic surfaces. The example is the self-assembly of large arrays of aromatic peptide nanotubes using vapor deposition methods. This approach allows controlling of the length and density of the nanotubes by supplying the building blocks from the gas phase. The nanotube arrays can be used to develop high-surface-area electrodes for energy storage applications, microfluidic chips, and also highly hydrophobic self-cleaning surfaces [168]. Other interesting applications are supramolecular gels in nonpolar solvents which are composed of self-assembled nanowires. Such studies highlight the role of self-assembly and gelation in the electronic properties of semiconducting molecular gelators and opens the window for a new class of conducting materials which may find a wide application in organic electronic devices [159]. There was also a proposition of using amyloid fibrils as new nanoscale biomaterials to investigate cell adhesion, migration, and differentiation *in vitro*. Gras et al. [169][85] used peptides with an additional segment motif of the biological cell adhesion sequence (RGD) or a control sequence (RAD) at the C-terminus of an 11-residue peptide taken from the amyloidogenic protein transthyretin. The fibrils containing such sequences are bioactive and interact specifically with cells via the incorporated sequences exposed on the fibril surface. Such functionalized fibrils can be systematically altered, so it could be possible to generate nanomaterials based on amyloids to promote interactions for a range of cell types.

One of the most recent and striking examples of the usefulness of a potentially dangerous, fibril-forming protein was described in [170,171]. The protein is  $\alpha$ -synuclein which participates in the Lewy body formation in Parkinson's disease. The authors fabricated a pea pod-type gold

nanoparticle (AuNP) arrangement into one-dimensional chain structures within the dielectric amyloid fibrils of  $\alpha$ -synuclein. The assembly units composed of  $\alpha$ -synuclein encapsulating AuNPs were manipulated by either hexane or the pH value to induce structural rearrangement within the protein coat. The method of encapsulation of noble metal nanoparticles within dielectric matrices is used to develop fast optoelectric response systems near the surface plasmon resonance frequency. Light energy can be transported through nanoparticles whose sizes are substantially smaller than the wavelength of the corresponding light. These AuNP-embedded amyloid protein nanofibrils exhibited photoconductivity with visible light—such property is crucial for the development of a subwavelength size light-guiding nano-optics systems.

### Conclusions

The genome-wide analysis revealed that self-complementary amyloidogenic segments are found in almost all proteins [97]; however, not all proteins form amyloids. There are 40–50 amyloid-associated diseases identified so far, but only a few proteins were identified to be causative in such diseases. Such an observation may result from the fact that only the most vulnerable proteins convert into amyloids. There are protective mechanisms that shield other proteins from this dangerous behavior. About 500–600 genes/proteins protect young organisms from such diseases, but their role is diminishing with age, so the ultimate goal would be to find a way to restore their protective function. Amyloids can be devastating but also beneficial when kept under control by specific cellular systems. Finally, the unusual properties of amyloid—mechanical, electronic, and other—can be exploited in many industrial applications. These biological nanostructures do not cease to inspire new ideas on how to protect the organism against their detrimental effects but also on how to use them in practical devices.



*Fig 1.15 A scheme displaying the process of amyloid formation resulting from instability of part of the protein structure (in yellow-green). The rest of proteins not participating in amyloid  $\beta$ -sheet development are shown as blue ellipses. All figures were created in Yasara (YASARA Biosciences) [450].*

## Part I: Cell-free Expression and Structural Studies of the $\gamma$ -Secretase Components

### 2 Cell-free Expression of Membrane Proteins

#### 2.1 Overview

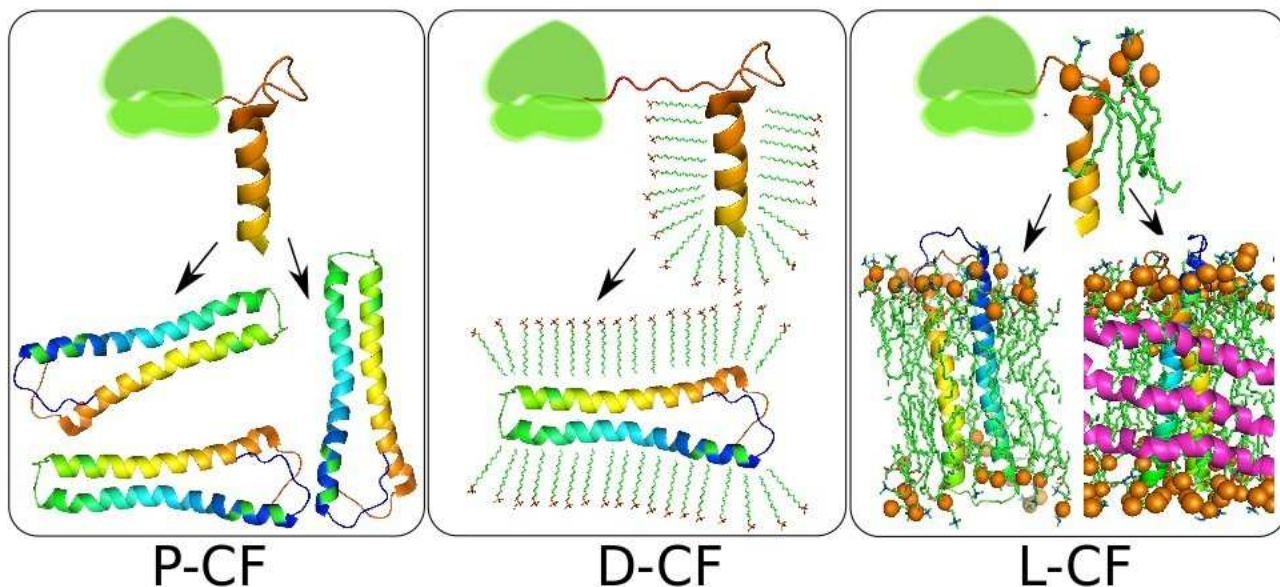


Fig 2.1 Various Cell-free expression modes: P-CF (as precipitate), D-CF (in detergent) and L-CF (in liposomes/bicelles and nanodiscs)

Cell-free (CF) expression has been proven to be an efficient way to express high amounts of not only soluble (cytosolic) but also functional membrane protein targets, which is impossible by traditional *E. coli* based *in vivo* expression system [172–176]. CF expressed proteins are suitable for structural and functional characterization as well as for high-throughput drug screening and in proteomics. CF is emerging as an important tool in synthetic biology which allows synthesis of biological building blocks and mimic their *in vivo* activity.

Recently mathematical modeling of gene expression dynamics was used to engineer CF system for high yield, reproducibility and predictability [177]. Kuruma et al showed the artificial synthesis of phospholipids by two membrane proteins embedded in liposomes [178]. In recent times functional membrane proteins have been successfully associated into nanodiscs [174,175,179]. Kim et al proposed the use of fructose-1,6-bisphosphate as an energy source in CF system to reduce the operating cost [180]. A myriad of membrane protein targets have been successfully expressed by CF system as described in our previous review [173].

Due to the openness of the CF system, it can be classified into three types depending on the type of external additives: P-CF (no additives and protein is expressed as precipitate), D-CF (addition of various Brij detergents in RM), and L-CF (addition of liposomes or bicelles or nanodiscs or

combination of lipid and detergent). This allows improved expression or functionality or direct incorporation of expressed proteins in micelles (D-CF) or lipid bilayers (L-CF) or expression as a precipitate.

## 2.2 Cell-free expressed membrane proteins

Over the years an efficient CECF expression system especially suitable for membrane protein synthesis has been set up. Since last few years, we have successfully expressed and purified and structurally or functionally classified various membrane protein targets such as GPCRs (bacteriorhodopsin into nanodiscs [181], solved NMR structure of proteorhodopsin [182], aquaporins [183,184] endothelin A receptor [185] etc.), channels (PorA, PorH [186]), transporters (organic cation transporters OCT1 and OCT2 and rat organic anion transporter OAT1[187,188]), enzymes (MraY [189] , GNA1 [190]), tail anchored protein Get3 [191] etc. A collection of 134  $\alpha$ -helical integral membrane proteins from *E. coli* inner membrane proteome mostly consisting of larger transporters were also expressed [188]. Furthermore we reconstituted the mitochondrial 542 kDa ATP-Synthase complex, comprising 16 membrane embedded and 9 soluble proteins, *in vitro* and resolved its electron microscopic structure by single particle reconstitution which is almost identical to the *in vivo* expressed complex [192]. We recently did N-terminal tag variation before the start codon in order to improve the expression of several GPCR targets [193].

## 2.3 Solved membrane protein structures of CF expressed proteins

Presenilin 1 is the catalytic subunit of the  $\gamma$ -secretase complex which cleaves the amyloid precursor protein into amyloid beta peptides (Ab 37-43), that accumulate as senile plaques in Alzheimer's patients, affecting more than 30 million people worldwide. The structure of CF expressed Presenilin 1 CTF in SDS micelles was solved using NMR spectroscopy [53]. Furthermore the structure of proteorhodopsin expressed in CF system was solved by NMR spectroscopy[182]. Other solved structures include multidrug transporter EmrE, voltage gated anion channel VDAC1 (expressed in D-CF mode).

Due to the emergence of multidrug resistant bacteria, new antibacterial targets are urgently required. By means of CF, we reconstituted the bacterial peptidoglycan (cell wall) synthesis pathway involving MraY and MurA,...,G *in vitro*. In our system, adding substrate of MurA: UDP-GlcNAc yields Lipid II which is the product of MurG. MraY translocase is involved in the formation of Lipid I. CF expressed MraY [189] can be useful for throughput anti bacterial screening because it is easily accessible and contains ten transmembrane helices that can be easily targeted. Several members of Mur family and MurY are already being targeted by existing antibiotics. Our system would allow discovery of novel drugs against Mur family and MraY because it allows throughput screen of a single lead compound against all of these protein targets.

Glucosamine 6-phosphate N-acetyltransferase (GNA1) is involved in UDP-GlcNAc biosynthesis pathway, and is indispensable for the viability of the organism. It can be a useful drug target against

variety of diseases including aspergillosis and cancer. We proposed a novel CF expression strategy for functional and inhibition assay of human GNA1 and GNA-sGFP in 96-well microplate format. In addition, we carried out inhibition assay of GNA1 and GNA1-sGFP by Glucose-6-phosphate inhibitor. From molecular modeling, we suggested binding mode of Glucose-6-phosphate to GNA1 and the differences to its original substrate and pathogenic *Aspergillus* GNA1 which could be helpful in novel lead design [190].

## 2.4 Expression of Membrane Proteins (MPs):

Membrane Proteins (MPs) represent 20-30% open reading frames of an genome and serve a myriad of functions including receptors, channels, transporters, enzymes, signal transducers, cell adhesion etc.  $\alpha$ -helical membrane proteins represent more than a quarter of all MPs in human genome. Unlike soluble proteins, it is not a trivial task to express membrane proteins. Integral membrane proteins are targeted to the membrane through endoplasmic reticulum (ER) and golgi apparatus (GB). The insertion of membrane proteins into the lipid bilayer is mediated by translocons i.e. Sec61 (eukaryotes) and SecYEG (prokaryotes).

	<i>E. coli</i>	<i>L. lactis</i>	Yeast	Insect	Mammalian	E-CF/W-CF
<b>Basic properties<sup>2</sup></b>						
Set up requirements	1	1	1	3	3	1/1
Time investment	1	1	2	3	3	1/3
Costs	1	1	2	3	3	2/2
Robustness	3	3	3	1	1	2/2
Preparative scale	3	3	3	2	1	3/3
Bio-safety risk	1	1	1	1-2	1-3	1/1
<b>Characteristics</b>						
Throughput option	+	+	+	-	-	+/?
<u>Expression environment</u>						
Native membranes	+	+	+	+	+	??
Micelles	-	-	-	-	-	+/?
Artificial liposomes	-	-	-	-	-	+/?
IBs/precipitate	+	?	(+)	(+)	-	+/?
<u>MP labelling</u>						
Uniform	+	+	+	-	-	+/?
Specific	+	+	+	-	-	+/?
Combinatorial	(+)	(+)	(+)	-	-	+/?
<u>PTMs</u>						
Glycosylation	-	-	+	+	+	-/?
Prenylation	-	-	+	+	+	-/?
Disulfid bonds	(+)	-	+	+	+	+/?

<sup>1</sup> IBs, inclusion bodies; PTMs, post-translational modifications; ?, not analysed yet, E-CF, *E. coli* cell-free, W-CF, wheat germ cell-free

<sup>2</sup> 1, low; 2, medium; 3, high

Table 2: Comparison of membrane protein expression systems

Rhodopsin is highly abundant in the retina of bovine eye. Such naturally evolved system specialized in producing high amounts of membrane proteins. However it is hard to express other proteins in such systems.

## 2.5 Cell-free expression of MPs

CF systems can easily be set up in biochemical laboratories due to the lack of requirement of advanced instruments. Initially the CF technique was employed to uncover the genetic code. Later it was used for analytical scale synthesis of soluble proteins. Preparative scale expression by means of CF was established in the 80s. In present time, CF systems are routinely used for expression and structural characterization of soluble proteins.

Preparative scale CF expression of MPs was first reported by [194]. Because CF expression is an open system, it is subjected to modifications at any time point. Furthermore, toxic and growth inhibitory effects of MPs in case of expression *in vivo* can readily be avoided. In cell-free extracts, cellular metabolism pathways are diminished.

## 2.6 CF extract sources:

The most efficient way to obtain mg amount of protein from reaction mixture (RM) is by using prokaryotic *E. coli* or eukaryotic wheat germ extracts (WGE). However, the quality of WGE is dependent on the source of the wheat seeds. Consequently, variations in expression can be observed from different sources of WGE, and it is also laborious and quite lengthy to prepare WGE. However, due to the eukaryotic origin of the extract, post-translational modifications can be observed, and also the protein expression can last upto a week resulting in 10 mg/ml yield.

In case of *E. coli* extract, various strains BL21 derivatives or strains deficient in endogenous RNAases like A19 or D10 can be used to obtain S30 extract within a day. Also endogenous mRNAs and amino acids are removed during S30 extract preparation resulting in minimization of unwanted protein expression and also efficient amino acid labeling schemes can be used for NMR spectroscopy [195]. WGE and S30 extracts are equally efficient in producing large proteins [188].

Other eukaryotic systems which are based upon rabbit reticulocyte extracts (RRL), insect cells or parasitic protozoan *Leishmania* can only be used for analytical scale productions. Furthermore a hybrid CF extract called PURE (Protein Synthesis Using Recombinant Elements) have been constructed where purified recombinant components are used to make the *E. coli* translation machinery. It can be useful to study the kinetics of protein translation to name a few.



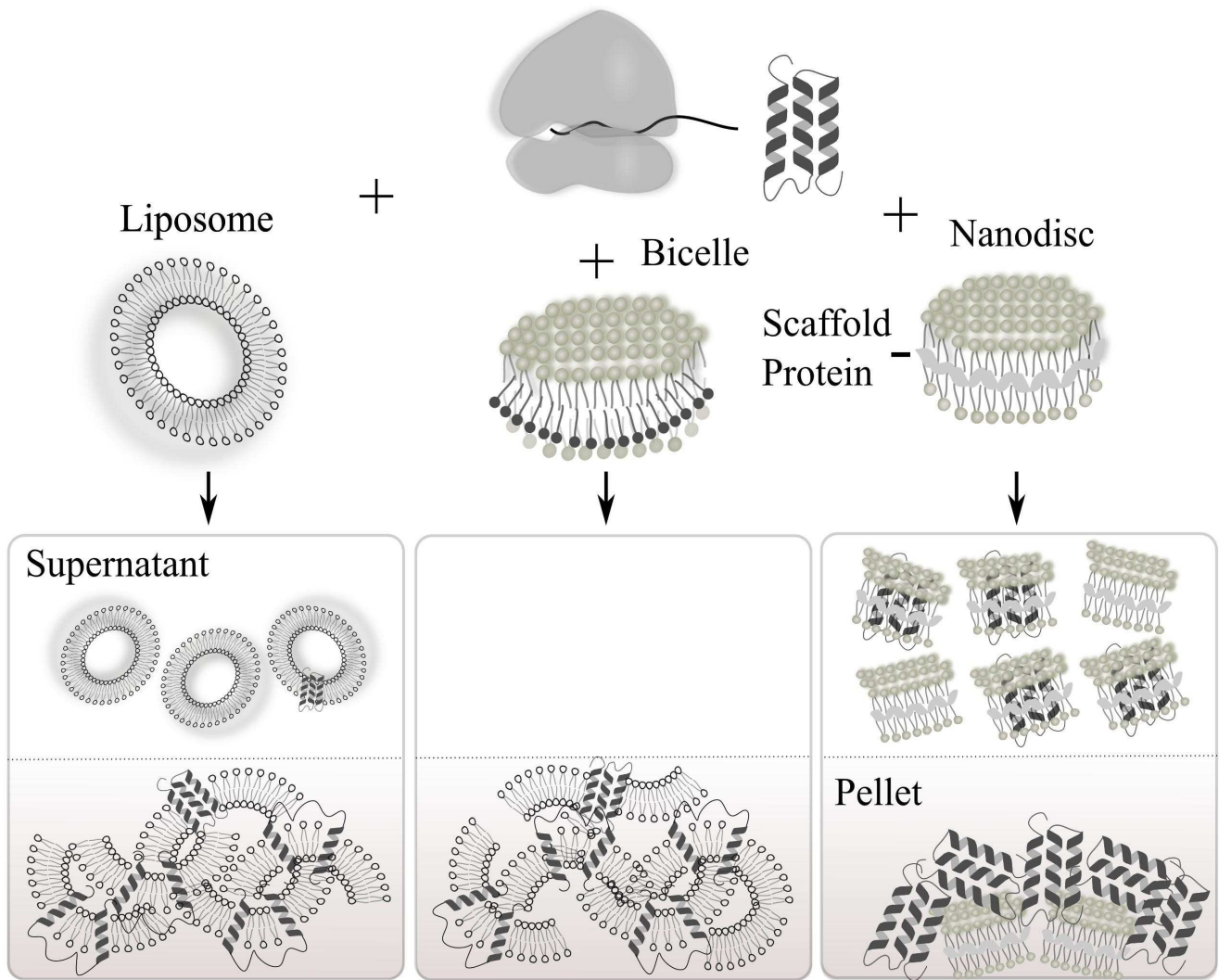
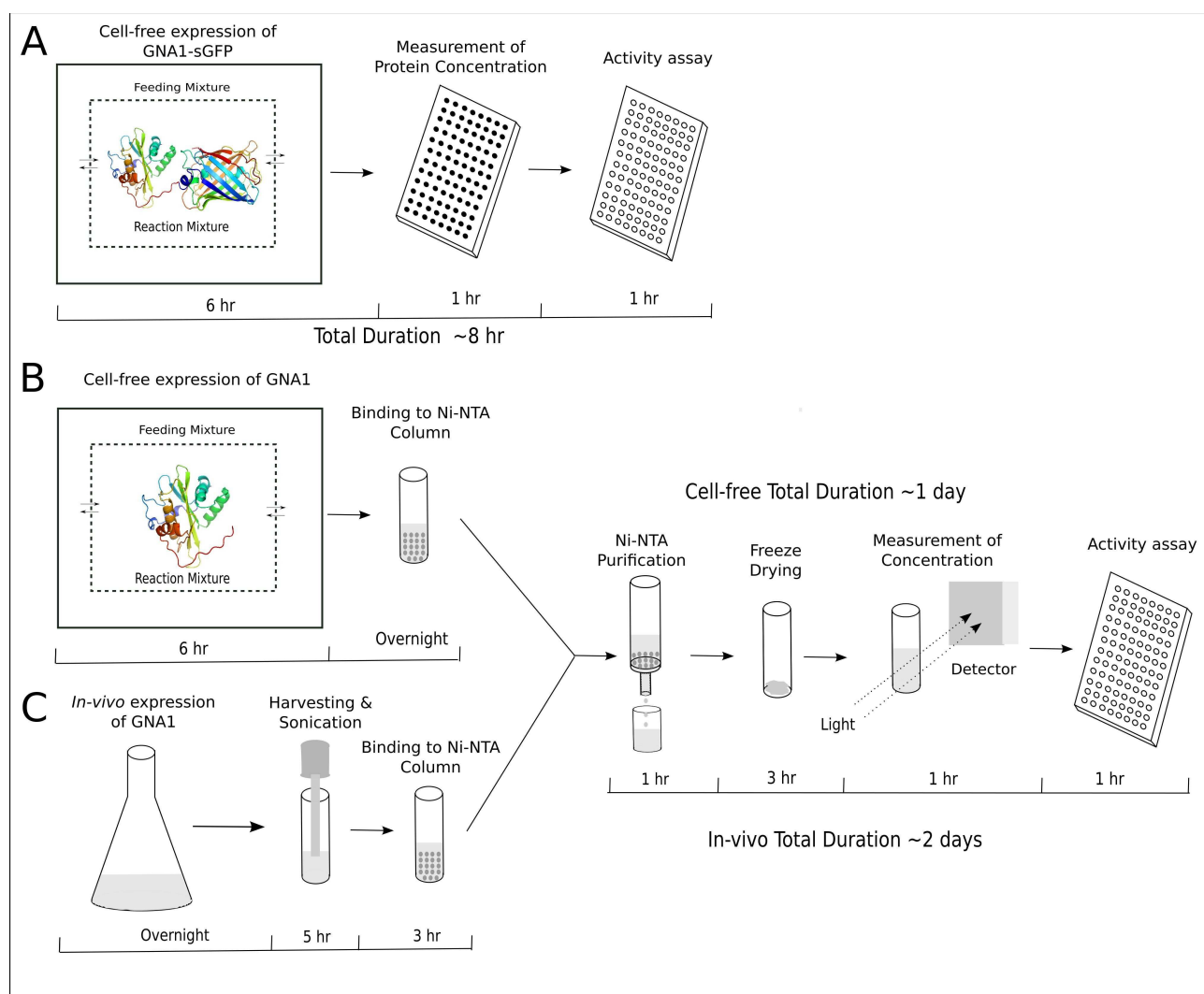


Fig 2.2: Various modes of L-CF expression



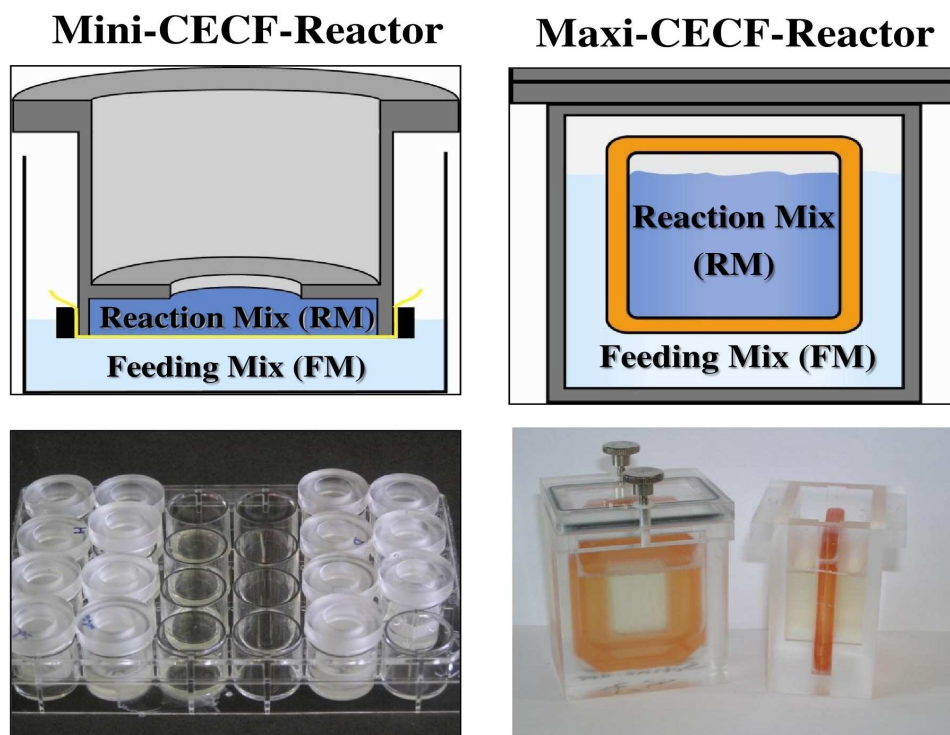
*Fig 2.3 Flow-chart of HsGNA1-sGFP and HsGNA1 production by CECF expression and in vivo. A. CECF expression of HsGNA1-sGFP which only takes ~6 hours followed by measurement of its concentration and assay of activity each for 1 hour. The whole procedure can be completed in 8 hours. B. In contrast, CECF expression of HsGNA1 alone followed by binding to Ni-NTA column and affinity purification, freeze drying, measurement of concentration and finally activity assay. The whole procedure takes around one day. C. In-vivo expression of HsGNA1, followed by harvesting the cells, sonication, centrifugation, resuspending the pellet and binding to the Ni-NTA column followed by purification and activity measurement steps similar to CECF expression of HsGNA1. Here binding to the Ni-NTA resin takes around 3 hours, totally, while it takes 2 days at least to finish this process for HsGNA1 in the traditional way.[190]*

## 2.7 E. coli based CF expression system

In the continuous exchange CF (CECF) system, milligram amounts of protein targets can be synthesized within a few hours. CF system utilizing the *E. coli* extract retains its transcription and translational machinery of ribosomes and ER that allows proper folding of newly synthesized proteins as well as an open system. During CF expression, plasmid/operon of the target protein(s) is added to the reaction mixture (RM) containing all higher molecular weight compounds like tRNA, T7 polymerase etc and *E. coli* extract. RM is separated from the feeding mixture (FM) holding the lower molecular weight precursors such as amino acids, energy sources (nucleotides) by a semipermeable membrane. This allows diffusion of toxic chemicals and aminoacids/energy sources

to FM and RM respectively and thereby improving the expression yield.

*E. coli* based CF expression systems are primarily based on T7 or SP6 polymerase coupling. So the target gene must contain T7 or SP6 promoter. This does not hold true for WGE CF system because the optimum  $Mg^{++}$  concentration is 15-18 mM in case of transcription but only 3 mM for translation. Also the translation efficiencies depend on the nature of 5' and 3' untranslated regions of mRNA. Therefore the 5' and 3' ends of the mRNA are replaced by AC rich sequences derived from tobacco mosaic virus and viral leaders found in plant virus genomes.



*Fig 2.4: Pictures showing Mini(left) and Maxi(right) cell-free reactors used for analytical and preparative scale reactions respectively*

CF extracts contain the translation machinery: ribosomes, aminoacyl-tRNA synthases, translation factors, acetate kinase etc. Additionally, transcriptional and translational precursors like amino acids, tRNAs, nucleotide triphosphates (ATP, GTP, CTP, TTP or collectively NTPs), DNA template are supplied during the CF reaction. Amino acids concentrations can vary from 0.3 – 2 mM and unstable amino acids (R, C, W, M, D, E) are given in higher amounts.

Commonly used component for ATP regeneration during CF reaction is phosphoenolpyruvate (PEP) and acetylphosphate (AcP) together with their enzymes. Protease inhibitors, ligands, stabilizers, chaperones and any compound which is beneficial for the expression, folding and stabilization of the protein can be added to the CF system due to its open nature. Additional tRNAs can replenish the effect of heterogeneous codon usage during translation. Formation of disulfide bonds can also be prevented. Chaperone and microsomal extracts can improve folding of the target MP. Moreover, composition of the supplied amino acid can be modified according to their distribution in the target MP. Incorporation of isotope labelled amino acids in the expressed protein makes them suitable for NMR and diminishes scrambling backgrounds.

CECF reactions can be performed in two ways. One is with small amount of RM (i.e. 30 -100 $\mu$ l) in order to optimize the expression by varying N-terminal tag of the target protein, vector type, Mg<sup>++</sup>, K<sup>+</sup> ion, plasmid concentrations etc for maximum yield. This is called analytical scale CECF reaction. It is required for screening of various conditions for maximal expression with minimal chemical usage. However for >80% of the targets only varying Mg<sup>++</sup> ion concentration from 12-22 mM and keeping K<sup>+</sup> ion concentration at a fixed 290 mM. Separate optimization of K<sup>+</sup> ion concentration in RM is required for different cell-free extracts because they have different concentrations of K<sup>+</sup> ion. After all required optimizations, one can proceed to CECF reactions in much higher scale up to 1-3 ml of RM to obtain target proteins in mg amounts per ml which is suitable for further functional (protein purification, activity assay, circular dichroism, gel filtration etc), structural studies (NMR, crystallization) etc.

Analytical scale reactions are usually carried out in 24 well plate trying 12 conditions in small mini-reactors (as shown in Fig 2.4) The membrane containing RM is manually sealed with a teflon ring and put into the well containing FM (note). High throughput robotic system for analytical CECF reaction is still under development. The RM is separated by FM with membrane (MWCO <14 kDa) so that all the toxic byproducts during protein synthesis can escape to the FM giving drastic improvements to expression yield in comparison to batch scale reaction. The volume of FM is kept at 15-20 times more than RM and it primarily consists of low molecular weight compounds like amino acids, and energy sources like NTPs, AcP, PEP etc as shown in Table No. 2 One disadvantage of manual setup of mini reaction chamber is the chances of leakage of RM which can be avoided with manual skill. Therefore the same reaction in mini reactors are done in duplicate to avoid error due to leakage of membrane. The reproducibility of analytical scale reaction in preparative scale reaction is quite high e.g. the conditions optimized in analytical scale can be used without alteration for comparable protein yields in preparative scale reactions. During preparative scale reactions the maxi reactors (Side-A-Lyser, Thermo Scientific, USA with MWCO <10 kDa) are used and it improves the diffusion rate because the RM is contained in flat membrane in two sides. The concentrations of required chemicals in stock solutions and their final concentrations in FM and RM are given in table 2. A pipetting scheme can be made using a spreadsheet softwares like proprietary MS Excel or free and opensource LibreOffice Calc or Gnumeric. Avoid air bubbles in the RM while pipetting it to mini or maxi reactors. Also carefully check for any leakage from mini or maxi reactors containing RM.

## 2.8 Various CF expression modes for MPs

Only in CF, Mps can be expressed in three novel ways. In the simplest method, Mps can be expressed as precipitate (P-CF) mode without providing any hydrophobic environment. Further detergents can be added for ready incorporation of MPs as micelles (D-CF) or in lipid based (L-CF) mode.

### 2.8.1 P-CF Mode

In P-CF mode of CF expression, MPs precipitate just after translation because the CF extract lacks any amphiphilic or hydrophobic substance (only 50-100 µg lipid/ml of extract which might not be admissible to the MP). The P-CF precipitates can readily be solubilized in detergents within few hours. No extensive refolding mechanism is required like in inclusion bodies. The detergents which are often used for resolubilization include lipid like n-dodecylphosphocholine (Fos-12) , lyso-phosphoglycerols 1-myristoyl-2-hydroxy-sn-glycero-3-[phospho-rac-(1-glycerol)] (LMPG) and 1-palmitoyl-2-hydroxy-sn-glycero-3-[phospho-rac-(1-glycerol)] (LPPG) . These detergents are suitable for structural investigation by NMR.

### 2.8.2 D-CF Mode

In D-CF mode the RM contains artificial amphiphilic environment comprising detergents in which the synthesized protein is inserted just after translation. As a result hardly any precipitate of expressed protein is obtained. However it requires to screen a variety of detergents to ascertain in which detergent the protein solubility is maximum and functional. However some detergents like Fos-12 and Cholamidopropyl)dimethylammonio]propanesulfonic acid (CHAPS) were found to inhibit protein expression despite being present in low concentration (<critical micelle concentration (CMC)) in the RM. In case of other well used detergents n-dodecyl-β-D-maltoside (DDM) or Triton X-100 were found to be efficient until 10 times of their respective CMCs. Having said that, most suitable detergents for Mps expression in D-CF mode are Brij family which are mild polyoxyethylene-alkyl-ethers, and also the steroid derivative digitonin. In case of Brij, the tolerance exceeds 100 times of CMC [196].

Moreover, amphipols or fluorinated surfactants can be used in the RM during D-CF expression. Unlike detergents, they don't disintegrate membranes and can promote partitioning of membrane proteins into lipid bilayers. In addition, various detergents can be combined (mixed micelles) or lipid and detergent mixture (lipomicelles) can be used to render ideal environment for proper solubilization and functionality of the nascent protein expressed in the RM. Lipids might have beneficial effect to membrane protein folding and can work like chaperones. One advantage of D-CF mode is that the expressed proteins are not required to be resolubilized. So an important requirement to obtain MP s in soluble form can be achieved.

### 2.8.3 L-CF Mode

Co-translational insertion of expressed MPs into supplied lipid bilayers can be achieved in L-CF mode. Lipids are natural environment for Mps and required for their folding, stability and function.

P-CF or D-CF expressed MPs can be integrated to the lipid bilayer in two different ways. One is post-translational reconstitution based on published protocols for the specific protein. The alternative is L-CF mode of expression. The charge of the lipid plays decisive role in determining the translational efficiencies of Mps e.g. cationic lipids inhibit translation. Lipids can be added to the RM as preformed liposomes in defined compositions, isolated fractions of cell membranes, detergent solubilized lipo-micelles allowing mixed environment to the expressed MP, bicelles comprising planar bilayers surrounded by detergents, nano-lipid particles or nanodiscs providing highly soluble bilayer. Mechanism of co-translational MP translocation can be assessed by screened by all these options. In L-CF mode, the insertion of membrane protein to the lipid bilayer is more specific compared to the post-translational reconstitution in P-CF/D-Cf mode.

Having said that, the L-CF mode of expression is still not well established method for MP expression and currently various possibilities are being investigated. In 2007 Kalmbach et al showed co-translational insertion of bacteriorhodopsin in various lipids and lipid mixtures. What is more, cholate solubilized phosphatidylcholines can be mixed to the RM and can easily be dialyzed out to assist bacteriorhodopsin proteoliposome formation. Combining detergents to lipids in RM is useful in improving the functionality of the synthesized protein. Effect of chaperones, crude inner membrane vesicles and microsomes in MP integration to membranes have been reported. But the added chaperones did not improve the post-translational insertion of aquaporin Z into synthetic liposomes. Expression of membrane proteins can further be improved by means of variation of lipid composition, polarity of the lipid head group, alkyl chain length in order to change the lipid bilayer thickness. Recently use of nanodiscs have become useful tools to improve MP integration to lipid bilayer in aqueous environment.

## 3 Materials

### 3.1 Laboratory equipments

Name	Source Company
0.22- $\mu$ m polysulfone filters	Roth
10-liter fermenter	B.Braun
96F Nunclon Delta Black Microwell SI	Nunc, ON 137101
Äkta purifier	GE Healthcare Amersham
Autoclave	Getinge
Balance	Sartorius
Centrifuge	
Centriprep devices YM-10	Amicon
Cooled table top centrifuge	
Dialysis tubes type 27/32 MWCO 14 kDa	Spectrum, Rancho Dominguez
DispoDialyser 25 kDa MWCO, regenerated cellulose Spectrum,	Rancho Dominguez
French press / Cell disrupter	
Gel imager	Biometra
Glass vials	e.g., Rotilabo vials; Roth
Heating block	VWR
ITHACO electrometer	ITHACO, NY, USA
Lumi Imager P1	Roche Diagnostics
Maxi PCR Purification Kit	Machery & Nagel
MicroDispoDialyser 25 kDa MWCO, regenerated cellulose	e.g., Roth
Microwave	Alaska
Midi DNA preparation kit	e.g., Qiagen
Mini-extruder	Avanti lipids
Nanodrop 100	peqlab
Ni-NTA Superflow resin	Qiagen
pH-meter PHM 210	Radiometer
Pipettes	Abimed, Gilson
Plate Centrifuge Rotana 46RSC-Robotic	Hettich
PS-microplate 96well V-shape ON 651101	Greiner bio-one
QIAquick gel extraction kit	Qiagen
QIAquick PCR purification kit	Qiagen
Rolling device	e.g., Frö Labortechnik
SDS-polyacrylamide gel electrophoresis system	BioRad
SEC columns: Superdex200 3.2/30	
Shaker	Heidolph
Shaking incubator	New Brunswick
Slide-A-lyzer dialysis cassette 10 kDa MWCO	Pierce
Sonifier Labsonic U	B.Braun
Temperature-controlled shaking incubator	
Thermocycler	Bio-Rad, Eppendorf
UV/Vis spectrometer	Cary
Vacuum pump	Abm
Vortexer Reax 2000	Heidolph
Waterbath	
Western blotting system	BioRad

## 3.2 Chemicals

Chemicals were purchased from Roth (Darmstadt) unless otherwise stated. Molecular biology enzymes were obtained from NewEngland Biolabs (Frankfurt, Germany).

1,4-Dihydrobutane dihydrochloride (Putrescine) (Sigma)  
1,4-Dithiothreitol (Roth)  
2-Mercaptoethanol (Sigma Aldrich)  
2x yeast tryptophan peptone glucose (YTPG) medium (see Buffers)  
Acetyl phosphate lithium potassium salt (Sigma Aldrich)  
Adenosine 5-triphosphate disodium salt trihydrate (Roche Diagnostics)  
Amino acids (Sigma Aldrich)  
Antibodies: anti-biotin peroxidase conjugate (Sigma); anti-green fluorescent protein from rabbit (Calbiochem); anti-mouse IgG HRP conjugate from goat (Sigma Aldrich); anti-penta His IgG from mouse (Qiagen); goat anti-rabbit IgG (Calbiochem); T7 tag antibody HRP conjugate (Novagen)  
Antifoam Y-30 emulsion (Sigma Aldrich)  
Bactotryptone (Roth)  
Benchmark protein ladder (Invitrogen)  
Coenzyme A sodium salt hydrate (Sigma)  
Complete protease inhibitor mix (Roche Diagnostics)  
Cytidine 5`triphosphate disodium salt hydrate (Fluka Sigma Aldrich)  
Detergents (Sigma Aldrich, Anatrace, Avanti Polar Lipids, Glycon): Brij-35, polyoxyethylene-(23)-lauryl-ether; Brij-56, polyoxyethylene-(10)-cetyl-ether; Brij-58, polyoxyethylene-(20)-cetyl-ether; Brij-72, polyoxyethylene-(2)-stearylether; Brij-78, polyoxyethylene-(20)-stearyl-ether; Brij-97, polyoxyethylene-(10)-oleyl-ether; Brij-98, polyoxyethylene-(20)-oleyl-ether;  $\beta$ -OG, n-octyl- $\beta$ -glucopyranoside; FG, n-Heptyl- $\beta$ -D-Glucopyranoside; NG, n-Nonyl- $\beta$ -D-Glucopyranoside; CHAPS, 3-[(3-cholamidopropyl)dimethylammonio]-1-propansulfonat; DHPC, 1,2-diheptanoyl-sn-glycero-3-phosphocholine; diC6PC, 1,2-dihexanoyl-sn-glycero-3-phosphocholine; diC8PC, 1,2-dioctanoyl-sn-glycero-3-phosphocholine; DM, n-decyl- $\beta$ -maltoside; DDM, n-dodecyl- $\beta$ -D-maltoside; NM, n-Nonyl- $\beta$ -maltoside; DMPC, 1,2-dimyristoyl-sn-glycero-3-phosphocholine; DPC, dodecyl-phosphocholine; LMPG, 1-myristoyl-2-hydroxy-snglycero-3-[phospho-rac-(1-glycerol)]; LPPG, 1-palmitoyl-2-hydroxy-sn-glycero-3-[phospho-rac-(1-glycerol)]; Lauryl-MNG, 2,2-didecylpropane-1,3-bis- $\beta$ -D-maltopyranoside; Decyl- MNG, 2,2-dioctylpropane-1,3-bis-b-D-maltopyranoside; C6F-TAC, C6F13C2H4-S-poly[tris(hydroxymethyl)aminomethane]; C8F-TAC, C8F17C2H4-S-poly[tris(hydroxymethyl)]; SDS, sodium dodecylsulfate; Triton X-100,; Tween 20, polyoxy-ethylene sorbitan monolaurate 20; Tyloxapol. E. coli A19 (E. coli Genetic Stock Center)  
Ethidiumbromide (Roth)  
Folinic acid calcium salt (Sigma Aldrich)  
Gene ruler 1kb DNA ladder (Fermentas)  
Glucose monohydrate (Roth)  
Guanosine 5`triphosphate disodium salt hydrate (Fluka Sigma Aldrich)  
Hemin (Sigma)  
HEPES (Roth) Imidazole (Roth)

## 3.3 Software

Gemini 4.2.17.304  
Tecan Magellan 5.03  
UNICORN 5.11



### 3.4 Buffers and Media for S30 extract and T7 polymerase preparation:

40 x S30-A/B buffer: 400 mM Tris-acetate, pH 8.2, 560 mM Mg(OAc)<sub>2</sub>, 2.4 M KCl.

Supplement 1 x S30-A buffer with 6 mM β-mercaptoethanol.

Supplement 1 x S30-B buffer with 1 mM DTT and 1 mM PMSF.

40 x S30-C buffer: 400 mM Tris-acetate, pH 8.2, 560 mM Mg(OAc)<sub>2</sub>, 2.4 M KOAc.

Supplement 1 x S30-C buffer with 0.5 mM DTT.

2x YTPG medium: 22 mM KH<sub>2</sub>PO<sub>4</sub>, 40 mM K<sub>2</sub>HPO<sub>4</sub>, 100 mM glucose, tryptone 16 g/l, yeast extract 10 g/l, NaCl 5 g/l.

LB medium: Peptone 10 g/l, yeast extract 5 g/l, NaCl 5 g/l.

Buffer T7RNAP-A: 30 mM Tris-HCl, pH 8.0, 50 mM NaCl, 10 mM EDTA, 10 mM β-mercaptoethanol, 5% glycerol.

Buffer T7RNAP-B: 30 mM Tris-HCl, pH 8.0, 50 mM NaCl, 1 mM EDTA, 10 mM β-mercaptoethanol, 5% glycerol.

Buffer T7RNAP-C: 30 mM Tris-HCl, pH 8.0, 1 M NaCl, 1 mM EDTA, 10 mM β-mercaptoethanol, 5% glycerol.

Buffer T7RNAP-D: 10 mM K<sub>2</sub>HPO<sub>4</sub>/KH<sub>2</sub>PO<sub>4</sub>, pH 8.0, 10 mM NaCl, 0.5 mM EDTA, 1 mM DTT, 5% glycerol.

20% streptomycin sulphate.

### 3.5 Reagents for CF reaction:

All stock solutions should be prepared with MilliQ ultrapure water and stored at -20°C if not otherwise stated.

50 x Complete® protease inhibitor cocktail (Roche Diagnostics) 1 tablet per 1 ml of MilliQ water.

Amino acid mixtures containing 4 mM or 8 mM of each of the 20 natural amino acids.

RCWMDE mixture containing 16.7 mM of each amino acid.

1 M acetyl phosphate lithium potassium salt (AcP) (Sigma-Aldrich), adjusted to pH 7.0 with KOH.

1 M phospho(enol)pyruvic acid K<sup>+</sup> salt (PEP) (Sigma-Aldrich), adjusted to pH 7.0 with KOH.

NTP mixture containing 90 mM ATP, 60 mM CTP, 60 mM GTP and 60 mM UTP, adjusted to pH 7.0 with NaOH.

Pyruvate kinase (Roche Diagnostics), 10 mg/ml.

RiboLock® RNase inhibitor (Fermentas), 40 U/μl.

Total E. coli tRNA (Roche Diagnostics), 40 mg/ml.

Folinic acid, Ca<sup>2+</sup> salt, 10 mg/ml (Sigma-Aldrich).

Polyethylene glycol 8000 (PEG 8000), 40% (w/v).

4 M potassium acetate (KOAc).

2.4 M HEPES/ 20 mM EDTA, pH 8.0 adjusted with KOH.

500 mM 1,4-dithiothreitol (DTT).

E. coli S30 extract, store frozen at -80°C

T7-RNA polymerase (T7RNAP), store frozen at -80°C

Template DNA (plasmid DNA or linear PCR products) 200-500 ng/μl

Reaction container: Analytical and preparative scale reaction container; D-tube containers, 12-14 kDa MWCO (Merck Biosciences); Slide-A-Lyzer, 10 kDa MWCO (Pierce); dialysis tubes, 14 kDa MWCO.

10 x Premix: 15 mM putrescine, 15 mM spermidine, 2.5 M K<sup>+</sup>-glutamate, 100 mM NH<sub>4</sub><sup>+</sup>-glutamate,

100 mM Mg<sup>2+</sup>-glutamate, 40 mM Na<sup>+</sup>-oxalate, 330 mM Na<sup>+</sup>-pyruvate, 340 μg/ml folinic acid, 10 mM DTT, 5.3 mM NAD<sup>+</sup>. 30 mM CoA-Na<sup>+</sup>.

GFP assay buffer: 20 mM Tris, 150 mM NaCl, pH 7.8.

1000× Ampicillin stock: Dissolve 100 mg/ml Na<sup>+</sup>-Ampicillin salt in 50% H<sub>2</sub>O and 50% EtOH. Store at -20°C.

1000× Kanamycin stock: Dissolve 30 mg/ml kanamycin sulfate in H<sub>2</sub>O. Store at -20°C.

all other chemicals are from Sigma-Aldrich if not otherwise stated.

### 3.6 SDS-gel buffers:

Stacking gel buffer: 0.4% (w/v) SDS, 0.5 M Tris-HCl, pH 6.8.

Separating gel buffer: 0.4% (w/v) SDS, 1.5 M Tris-HCl, pH 8.9.

5× SDS-PAGE sample buffer: 25% (w/v) glycerol, 25% (v/v) β-mercaptoethanol, 7.5% (v/v) SDS, 0.05% (w/v) bromphenol blue, 300 M Tris-HCl, pH 6.8.

Running buffer: 0.025 M Tris-HCl, pH 8, 0.1% (w/v) SDS and 0.2 M glycine.

Tricine-SDS- PAGE buffers:

10× Anode buffer: 1.0 M Tris-HCl, pH 8.9

10× Cathode buffer: 1.0 M Tris, 1.0 M Tricine, 1.0% (w/v) SDS, pH ~8.25

3× buffer: 3.0 M Tris-HCl, 0.3% (w/v) SDS, pH 8.45

### 3.7 Gel

Coomassie brilliant blue-staining solution for SDS gels:

50% (v/v) ethanol (96%), 10% (v/v) acetic acid (100%) and 0.1% (w/v) Coomassie Brilliant Blue G250 + R250. Dissolve in H<sub>2</sub>O and store at RT in a dark bottle.

Western-blotting buffer (Towbin): Dissolve 25 mM Tris, 192 mM Glycin, 3.5 mM (1%) SDS, 15% MeOH in H<sub>2</sub>O and adjust to pH 8.3 by HCl. Store at 4° C.

ECL1: 100mM Tris (pH 8.5), 2.5 mM Luminol, 0.4 mM p-cumaric acid. Store at 4° C.

ECL2: 100mM Tris (pH 8.5), 0.0183% H<sub>2</sub>O<sub>2</sub>. Store at 4° C

### 3.8 Sequencing primers for pET-Vectors

T7-promotor : 5' TAA TAC GAC TCA CTA TAG GG 3' TM= 53.2 °C

T7-terminator: 5' GCT AGT TAT TGC TCA GCG G 3' TM= 56.6 °C

### 3.9 Materials for Nanodisc Preparation

pET28b vector (Merck, Germany)

1 M IPTG in H<sub>2</sub>O

30 mM kanamycine

500 mM phenylmethanesulfonyl fluoride (PMSF) in 100% ethanol

Syringe filter, sterile, 0.45 μm (Roth)

Buffer MSP-A: 40 mM Tris, pH 8.0, 300 mM NaCl, 1% TritonX-100 (v/v).

Buffer MSP-B: 40 mM Tris, pH 8.9, 300 mM NaCl, 50 mM cholic acid.

Buffer MSP-C: 40 mM Tris, pH 8.0, 300 mM NaCl.

Buffer MSP-D: 40 mM Tris, pH 8.0, 300 mM NaCl, 50 mM imidazol.

Buffer MSP-E: 40 mM Tris, pH 8.0, 300 mM NaCl, 300 mM imidazol.

Buffer MSP-F: 40 mM Tris, pH 8.0, 300 mM NaCl, 10% glycerol (v/v).

Buffer ND-A: 10 mM Tris, pH 8.0, 100 mM NaCl

Strep-buffer: 20 mM Tris-Cl, pH 7.5, 100 mM NaCl

Sodium cholate (Carl Roth)

D-desthiobiotin (IBA)

Strep-elution buffer

### 3.10 Materials for Tag Variation Screen and Template Preparation

PCR purification kit (Qiagen)

Plasmid DNA purification kit (Machery-Nagel)

Vent polymerase (New England Biolabs)

Agarose (Rotigrose, Carl Roth).

### 3.11 Microbial Strains

Strain	Relevant genotype	Reference
<i>E. coli</i> DH5 $\alpha$	[F', <i>endA1</i> , <i>hsdR17</i> ( <i>r<sub>k</sub>'m<sub>k</sub>'</i> ), <i>supE44</i> , <i>thi-1</i> , <i>recA1</i> , <i>gyrA</i> (Nal <sup>r</sup> ), <i>relA1</i> , $\Delta$ ( <i>lacZYA-argF</i> )U169, $\Phi$ 80 <i>lacZ</i> $\Delta$ M15]	(Hanahan et al. 1983)
<i>E. coli</i> BL21 (DE3)	<i>hsdS</i> , <i>gal</i> [ $\lambda$ cI, <i>ts857</i> , <i>cnd1</i> , <i>hsdR17</i> , <i>recA1</i> , <i>endA1</i> , <i>gyrA96</i> , <i>thi-1</i> , <i>relA1</i> ]	(Studier and Moffat 1986)
<i>E. coli</i> K-12 strain A19	[ <i>rna19</i> <i>gdh A2</i> <i>his95</i> <i>relA1</i> <i>spoT1</i> <i>metB1</i> ]	<i>E. coli</i> Genetic stock center, New Haven, CT

### 3.12 Oligonucleotides

All primers were ordered from MWG Biotech AG or Biospring unless otherwise stated.

## 4 Methods

### 4.1 Cell-free Expression of GS Components

#### 4.1.1 S30 extract preparation

In order to obtain highly efficient cellular extracts, cultivation of the cells in a fermenter with good aeration is recommended. Growth curve of selected strain should first be determined in the individual facility and an efficient chilling of the broth after fermentation should be ensured. A brief summary of S30 extract preparation is listed in the following table, detailed protocols can be found in previous reports [197].

Protocol	Step	Comments
Cell fermentation	Inoculation	100 ml freshly E. coli overnight culture into 10 liters of sterilized medium in a fermenter.
	Incubation	37oC with intensive aeration and stirring
	Chilling	Start to chill the cell broth before the cells reach mid-log phase (OD600 3-5) (see Note 6).
	Harvesting	Harvest the cells at 18oC by centrifugation at 7,000 x g for 15 min at 4oC.
	Storage	Keep the cell pellets at 4oC for the following steps or freeze the cell paste in -80oC for later processing.
Cell extraction	Washing	Gently but completely resuspend the cell pellet with 300 ml pre-cooled S30-A buffer. Centrifuge at 7,000 x g for 10 min at 4oC. Discard the supernatant and repeat washing step twice.
	Cell disruption	Resuspend the pellet in 110% (v/w) pre-cooled S30-B buffer and disrupt the cell by high pressure, e.g, French press at 1,000 psi (see Note 7).
	Lysate clearing	Centrifuge the lysate at 30,000 x g for 30 min at 4oC. Transfer the supernatant to a fresh tube and repeat the centrifuge step.
	Removal of endogenous mRNA	Harvest the supernatant and adjust to a final concentration of 400 mM NaCl. Incubate at 42oC for 45 min in a water bath (see Note 8).
	Dialysis	Dialyse the turbid extract overnight against 100-fold excess of pre-cooled S30-C buffer (12-14 kDa MWCO membrane) with two changes of dialysis buffer.
	Harvesting	Centrifuge the extract at 30,000 x g for 30 min at 4oC. Harvest the supernatant and transfer to a clean plastic tubes.
	Storage	Aliquot into small reaction tubes, shock-freeze in liquid nitrogen and store at -80oC.

## Pre-preparation

Make pre-culture in 100ml LB medium from LB plates prepared the previous day.

### 5L 2xYTPG:

2xYPTG: 160g Tryptone, 100g Hefeextract, 50g NaCl to 5L (gets autoclaved in fermenter)

2L Phosphate: 29.9g KH<sub>2</sub>PO<sub>4</sub>, 91.3g K<sub>2</sub>HPO<sub>4</sub> to 2L (autoclave separately) pH 7

1M Glucose: 198g to 1L (filter separately)

2L DH<sub>2</sub>O: autoclave separately to add later

### S30 buffers (50x)

A/B (50x): 60.6g Tris, 150.1g Mg(OAc)<sub>2</sub>, 224g KCl to 1L (adjust pH to 8.2 with acetic acid)

C (50x): 60.6g Tris, 150.1g Mg(OAc)<sub>2</sub>, 294g KOAc to 1L (adjust pH to 8.2 with acetic acid)

### 2M DTT

1.54g in 5ml H<sub>2</sub>O

### 100mM PMSF

34.8mg in 2ml EtOH

### Other preparations

Chill two 5L beaker of ddH<sub>2</sub>O

Autoclave necessary glassware

Collect six 1L centrifuge tubes

Collect ten 35ml centrifuge tubes

Collect two 500ml centrifuge tubes (pre-weigh one of these)

## Preparation of S30 extract

First day

1. Drain water from fermenter and rinse once more with dH<sub>2</sub>O
2. Pour media in fermenter (w/o glucose and phosphate; they ppt out if autoclaved) and add 2L of H<sub>2</sub>O (to a total volume of 7L)
3. Turn fermenter on with green switch and press Ack to disengage alarm
4. Set fermenter switch to sterilization position
5. Re-attach third pipe from the left to remove dirty water.
6. Turn on the water and make sure small black faucet at on top of metal tubing is open (otherwise the water cannot enter) and press thermostat fill a couple of times
7. Close off the pressure sensor, close off the compressor (black knob to the side) and open the Aeration valve. Turn on compressor and wait until the meter is at 8-10 bar. At this point open both the pressure sensor and compressor, and close the aeration valve until a pressure of 1.6 bar is reached
8. Close metal cage, go to batch and sequence F1, then press 1 and enter to activate autoclave sequence (at this point, the ball at top of the pressure sensor should drop down)
9. Leave the room while autoclaving
10. After autoclaving, turn water off (once temperature comes back down)
11. Leave compressor on O/N (sterilization mode and stirring) so that nothing grows meanwhile

Second day

1. Put fermenter in fermentation position
2. release the pressure valve and open one of the openings to fermenter
3. place in glucose (filtered 0,2u) and phosphate buffer (should make it up to 10L)
4. place in antifoam for excessive foaming (usually 2ml)
5. release pressure and close opening
6. Go to batch control and begin stirring (~500rpm)
7. Add bacterial culture and take "0" sample for OD
8. Take samples every 0.5h for O.D (595nm).
9. When reaching about 2.4-2.8 O.D., the temp is set to ~ 0°. Once a temp of ~ 20° is reached, the water bath is filled with ice water and connected to fermenter (switch on top button of bath, not bottom). You then wait for the temp to reach less than 15° and take one last O.D.

10. Drain the bacteria into centrifuge flasks (six of these so you refill each once).
  11. Centrifuge (Ludwig Lab's) at 6500rpm for 15 min
  12. Pour supernatant again into fermenter and autoclave
  13. Take pellet and stir with glass rod. At this step, add buffer A (20ml 50x stock, 420ul  $\beta$ -mercapto to 1L H<sub>2</sub>O)
  14. After uniform, add ~ 10ml buffer A
  15. Pour contents into centrifuge flask (weighed) and top up with buffer to 300ml.
  16. Wash 2x at 8000xg for 10min and the last time for 30min
  17. Weigh the pellet and add 1.1x buffer B (2ml 50x stock, 50ul (2M stock) DTT, 100ul (100mM stock) PMSF in 100ml H<sub>2</sub>O)
  18. re-dissolve pellet in buffer B and take for cell disruption
  19. use french press as described
    - assemble french press upside down, place in nozzle and nob
    - make sure the opening is closed before placing it on apparatus
    - place so that knob and nozzle face outward and that the handles do not collide
    - turn on pump, place switch on high until up to platform roof and the press on medium.
    - turn the knob slightly so that pressure does not fall below 1000 and collect lysate
    - after bubbles come out, close knob and refill
  20. Alternatively, use the cell-breaker (Glaubitz)
    - Turn machine on, set pressure to 0.5 kbar, and press start to wash apparatus:
    - wash once with water, 2x with NaOH, and 2x with water (pressure 0.5 kbar)
    - pour in cells, press stop, change pressure to 1.5-1.7 kbar, and press start
    - run cells 2-3x through in order to get complete lysis
    - after finishing, wash machine as before (above)
  21. Take the supernatant and spin it at 30xg for 30min
  22. Take supernatant and transfer to another container
  23. Spin again at 30xg for 30min.
  24. Collect supernatant and add from a 5M NaCl stock to final conc of 400mM, and incubate at 42° for 45min; meanwhile dialysis tubes:
    - add the tip of NaH<sub>2</sub>PO<sub>4</sub> to 500ml H<sub>2</sub>O and microwave for 2min
    - add 1ml 0.5M EDTA to 500 ml H<sub>2</sub>O and microwave for 2min
    - rinse several times with water (if needed, store in 30% EtOH, 4°)
  25. Place extract inside membrane and let equilibrate in buffer C (100ml 50x stock, 1.25ml (2M) DTT in 5L H<sub>2</sub>O) a few hours. Then exchange buffer (another 5L) and let exchange O/N.
- Third day
- A. Take out dialyzed lysate and centrifuge at 30xg for 30min at 4°
  - B. Collect supernatant, transfer to eppendorf tubes, freeze in liquid nitrogen, and store at -80°

#### 4.1.2 Production of T7RNAP

T7RNAP is produced by conventional overexpression in *E. coli* cells in Erlenmeyer flasks. On average, approximately 20,000 - 40,000 units can be isolated out of one liter culture. A summary of the expression protocol is given in .

Protocol	Step	Comments
Fermentation	Inoculation	Inoculate on liter of LB medium 1:100 with a fresh overnight culture of strain BL21 (DE3) Star x pAR1219.
	Incubation and induction	Incubate at 37°C until OD <sub>600</sub> of 0.6 – 0.8 and induce with 1 mM IPTG. Continue to grow at 37°C.
	Harvesting	Harvest the cells by centrifugation at 8,000 x g for 15 min at 4°C after inducing for five hours. Cell pellets

can be stored at -80oC for later usage.

Purification	Cell disruption	Resuspend the cell pellet with 30 ml of T7RNAP-A buffer. Disrupt cells by French press at 1,000 psig or by sonication. Remove cell debris by centrifugation at 20,000 x g for 30 min at 4oC.
	Precipitation of DNA	Adjust the supernatant to a final concentration of 4% streptomycin sulphate. Mix gently, inoculate on ice for 5 min and centrifuge at 20,000 x g for 30 min at 4oC.
	Ion exchange chromatography	Load the supernatant on a 40 ml Q-sepharose column equilibrated with T7RNAP-B buffer and wash the column with the same buffer extensively. Elute the T7RNAP with a gradient from 50-500 mM NaCl using T7RNAP-C buffer for 10 cv. Collect the fractions and analyze by SDS-PAGE (see Note 9).
Concentration	Dialysis	Pool T7RNAP containing fractions and dialysis against T7RNAP-D buffer overnight.
	Concentration	Adjust to a final concentration of 10% glycerol and concentrate the T7RNAP pool to a total protein concentration of 3-4 mg/ml by ultrafiltration (see Note 10). Adjust to a final concentration of 50 % glycerol and store aliquots at -80oC.

### 4.1.3 CECF reaction preparation for 1 ml RM and 16 ml FM

Table 3

Compound	Stock	Final concentration	Volume <sup>a</sup> [μl]
<b>RFM-Mix</b>			
RCWMDE	16.7 mM	1 mM	1,020
Amino acid mix	25 mM	0.5 mM	374
Acetyl phosphate (Li <sup>+</sup> , K <sup>+</sup> )	1 M	20 mM	340
Phospho(enol)pyruvic acid (K <sup>+</sup> )	1 M	20 mM	340
75 x NTP mix	90 mM ATP	1.2 mM	226.7
	60 mM G/C/UTP	0.8 mM	
1,4 Dithiothreitol	500 mM	2 mM	68
Folinic acid (Ca <sup>2+</sup> )	10 mg/ml	0.1 mg/ml	170
Complete® protease inhibitor	50 x	1 x	340
Hepes/EDTA buffer	24 x	1 x	623.3
Mg(oAc)	1 M	11.1 16, mM <sup>b</sup>	274
KOAc	4 M	110, 270, mM <sup>b</sup>	382.5
PEG 8000	40%	2%	850
NaN <sub>3</sub>	10%	0.05%	85
MilliQ water			1,963.5
Total:			7,057
<b>RM-Mix</b>			
Master mix RFM			415 μl
Pyruvat kinase	10 mg/ml	0.04 mg/ml	4 μl
t-RNA ( <i>E. coli</i> )	40 mg/ml	0.5 mg/ml	12.5 μl
T7RNAP	1.4 mg/ml	0.05 mg/ml	35,7 μl
Ribolock	40 U/μl	0.3 U/μl	7.5 μl

DNA template	0.2 - 0.5 mg/ml	0.015 – 0.03 mg/ml	60 $\mu$ l
<i>E. coli</i> S30 extract	1 x	0.35 x	350 $\mu$ l
MilliQ water			115.3 $\mu$ l
			Total: 1 ml
<b>FM-Mix</b>			
Master mix RFM			6,642 $\mu$ l
S30-C Puffer	1 x	0.35 x	5,600 $\mu$ l
Amino acid mix	4 mM	0.5 mM	2,000 $\mu$ l
MilliQ water			1,758 $\mu$ l
			Total: 16 ml

<sup>a</sup> Calculated for 1 ml RM and 16 ml FM = 17 ml RFM master mix.

<sup>b</sup> Subject to optimization. Volumes are calculated for final total concentrations of  $Mg^{2+}$  of 16 mM and  $K^+$  of 270 mM as additional amounts of 4.9 mM  $Mg^{2+}$  and 160 mM  $K^+$  are contributed by other compounds.

#### 4.1.4 Template Production and Yield Optimization by Tag Variation Screen

CF system using *E. coli* extract require either circular plasmids or linear DNA as template for expression. No or insufficient expression of proteins is usually caused by an inefficient translation/transcription process. Rational template design is therefore important for successful protein synthesis.

#### 4.1.5 General Template Design

High level protein production in the *E. coli* CF system is usually directed by standard T7 regulatory sequences. T7 promoter as well as T7 terminator must be present in the DNA template. Suitable vectors containing the T7 promoter sequence include the pET (Merck Biosciences) and pIVEX (Roche Diagnostic) series. In addition, few further parameters regarding template design should be considered. (I) Purification tags. The presence of small tags at the C-terminal of the target protein is a valuable tool for the detection of expression and for full-length protein purification. Poly(His)<sub>10</sub>-tags or StrepII-tags may be used for protein detection by immunoblotting or for efficient membrane protein purification. (II) Expression monitoring. Fusion with reporter proteins such as GFP allows the fast monitoring of expression and may accelerate the optimization of the reaction conditions for improved protein target production. (III) Codon usage. In rare cases, cluster of non-frequent codons or the formation of unfavorable secondary structures within the coding sequence may cause amino acid misincorporation or premature terminations of the protein product. Such problems should be addressed by expression of synthetic genes. (IV) Expression tags. In most cases, inefficient initiation of translation is the reason for low expression efficiency. This problem is addressed by the tag variation screen (see Subheading 3.3.3) by fusing small tags to the N-terminus of the target protein. A small set of templates containing different optimized tags is synthesized by a two step PCR approach and the generated linear PCR products are directly screened in CF reactions.



### 4.1.6 DNA Template Preparation

High quality and purity of the plasmid or linear DNA is crucial for efficient CF expression. As the vectors do not replicate during the reaction, specific selection markers are not of importance. The final template concentration for optimal expression efficiency should be determined for each new target with an initial concentration screen in the range in between 0.1 and 20 ng/ $\mu$ l of reaction mixture.

### 4.1.7 Plasmid DNA

1. Inoculate 100–200 mL of LB medium supplemented with the specific antibiotic with the *E. coli* strain containing the desired plasmid and incubate overnight at 37°C on a shaker.
2. Purify the plasmid using commercial available kits like “Midi” or “Maxi” kits. “Mini” preparations are not suitable due to the low quality of the purified DNA.
3. Dry the DNA overnight and dissolve it into MilliQ water. Optimal concentration of stock DNA solution is in between 0.2 and 0.5 mg/ml. DNA stocks can be stored at -20°C. In case of precipitation due to freezing/thawing cycles, the DNA concentration must be checked again.

### 4.1.8 Linear DNA

If the target gene is already present in a suitable vector under control of T7 elements, fragments containing the T7 regulatory sequence and the target gene can be amplified by standard PCR and directly used in the CF reaction. If the target gene is not under T7 control or if expression/purification tags or expression monitors have to be attached, a multistep PCR strategy such as for the tag variation screen can be applied.

1. For the PCR two primers are needed: a forward primer annealing at the T7 promoter and a reverse primer annealing at the T7 terminator.
2. The reaction can be performed in 50  $\mu$ l volume with 20 ng of plasmid DNA as template, 0.4  $\mu$ M each of the forward and reverse primers, 200  $\mu$ M of dNTPs, 0.5 units of Vent DNA polymerase (NEB) and 1x ThermoPol reaction buffer. The PCR program has 2 min denaturation step at 94 °C followed by 32 cycles of denaturation for 15 s at 94°C, primer annealing for 30 s at 55°C and extension for 2.5 min. at 72°C. A last step of final elongation is performed for 10 min. at 72°C.
3. PCR products should be purified using standard PCR purification kits (e.g. Qiagen).

### 4.1.9 Tag variation screen for improved expression

Small expression tags comprising 6-12 codons are fused to the translational start site of the target coding sequence by an overlap PCR approach. The expression tags are optimized in suppressing secondary structure formation which may prevent the initiation of translation. As secondary structures also depend on the individual target sequence, it is recommended to empirically screen a

mini library of optimized tag fragments (Table 5). The whole tag variation screen can be performed in one day. The finally identified most efficient expression tag may further be modified [193]. The small expression tags have only minimal impact on structure and function and can usually remain on the target protein.

#### 4.1.10 Primer Design

Four different primers are needed for the first step of the PCR reaction in order to generate a library of tags fragments and a target fragment.

Primer 1 (P1): A universal forward primer annealing upstream of the T7 promoter.

Primer 2 (P2): Reverse tag-specific primer, for each tag a specific P2 primer is needed. P2 anneals at the ribosome binding site and contains a 5-prime linker with a specific tag sequence followed by an PCR overlap region. For the PCR overlap region, protease cleavage sites e.g. for TEV or PreScission protease could be used, allowing the option to remove the tags after protein purification.

Primer 3 (P3): This primer is target gene specific annealing at the start of the coding sequence and carrying a 5-prime linker with the PCR overlap region.

Primer 4 (P4): A universal reverse primer annealing downstream of the T7 terminator.

#### 4.1.11 Linear DNA Template Preparation for Tag Variation Screen

1. First a tag fragment library is produced by PCR using a vector containing the T7 promoter as template and the primers P1 and P2.
2. Second a linear target fragment is produced by PCR with the primers P3 and P4 and with a vector carrying the coding sequence of the target protein under T7 control.
3. All synthesized PCR fragments are purified using standard PCR purification kits (e.g. Qiagen).
4. In the overlap PCR step, the linear target fragment is mixed with the individual fragments from the tag library in equal molar ratios using 200 ng target fragment and corresponding amounts of tag fragments. The overlap PCR is performed with the primers P1 and P4. The resulting PCR fragments are purified and analysed for purity and integrity on a 1% agarose gel. The purified PCR fragments can be directly added into the CF reaction as templates. Expression tags may be removed after CF expression and purification of the target protein by taking advantage of the protease recognition site used as overlap region. Alternatively, identified beneficial expression tags may be further modified by truncations.

### 5. Expression tags for tag variation screen

Name	Nucleotide sequence	Amino acid sequence	AT (%)
AT	AAATAT TATAAATATTAT	KYYKYY	100
SER	AAATCATCATCATCA	KSSSSS	72
H	AAACCATACGATGGTCCA	KPYDGP	55
G	AAAAGTAAAGGAGAA GAA	KSKGEE	72

#### 4.1.12 General Setup of CF Expression Reactions

Analytical scale CECF reactions are suitable for any kind of screens such as  $Mg^{2+}$  optimization screens, tag variation screens or ND lipid screens and can be performed in Mini-CECF reactors or in suitable D-tubes (Merck Biosciences) with RM volumes of 50-100  $\mu$ l. The identified optimal reaction parameters can then be scaled up in a 1:1 ratio into preparative scale CECF reactions with RM volumes of several ml and by using Maxi-CECF reactors. Mastermixes of common compounds of RM and FM should be prepared in order to minimize variations in between identical reactions. Nevertheless, reactions should be always prepared at least as duplicates. We recommend a RM : FM ratio in between 1:15 and 1:20.

#### 4.1.13 Analytical Scale CECF Reactions

1. Calculate the individual compound volumes according to the desired number of reactions and design an appropriate pipetting scheme using standard software packages such as MS Excel, LibreOffice Calc or Gnumeric.
2. Prepare standard 24-well microplates and appropriate numbers of Mini-CECF reactors and pieces of dialysis membrane with a MWCO of 14 kDa.
3. Prepare a common master mix RFM for the RM and FM and mix by vortexing (Table 2).
4. Reconstitute RM and FM with the appropriate volumes of RFM and complete by addition of MilliQ water. Mix FM by vortexing and the RM by gentle shaking.
5. Fill RM and FM aliquots into reaction containers. A: Mini-CECF-Reactors [172] with RM volumes of 30 - 100  $\mu$ l can be used in combination with standard 24-well microplates holding FM volumes of up to 1.5 ml in their cavities. A piece of fresh dialysis membrane is fixed to the Mini-CECF-Reactors with a Teflon ring. Avoid air bubbles or residual water in the container and check for leakage. B: Commercial D-tube dialyzers (Novagen) for RM of 100  $\mu$ l can be used in combination with suitable tubes (e.g. 2 ml Eppendorf tubes) holding the appropriate volume of FM.
6. CECF reactions are incubated over night at temperatures in between 20 - 30°C with gentle shaking or rolling in order to ensure efficient substance exchange between RM and FM.

#### 4.1.14 Preparative Scale CECF Reactions

For preparative scale CECF reactions, commercial Slide-A-Lyzer devices (Pierce) with a MWCO of 10 kDa and holding up to 3 ml RM volumes may be used. As FM container, either Maxi-CECF-Reactors [197] or small suitable plastic boxes or beakers may be used. Alternatively, suitable sized D-tube dialyzer or appropriate pieces of dialysis tubes sealed at both ends by knots may be used as RM container in combination with 15 - 50 ml Falcon tubes as FM container.

#### **4.1.15 Quantification of Target Production by <sup>35</sup>S-Met Incorporation**

If exact determination of expression efficiency is necessary, the labelling with <sup>35</sup>S methionine is recommended. Refer to [197] for details.

#### **4.1.16 Preparation of NDs as Supplement for CF Reactions**

NDs are highly water soluble and stable. Different diameters of NDs are possible depending on the selected type of MSP [198,199]. Most frequently used are MSP1 (9.7 nm with POPC) fitting MPs containing at least up to 21 transmembrane segments [181] and MSP1E3 (12.1 nm with POPC). Essential for the efficient solubilization of CF expressed MPs are highly concentrated and homogenous ND stock solutions. The preformed NDs can be filled with a large variety of different lipids and the selected type of lipid can have crucial effects on the solubilization efficiency of individual MPs. We describe the basic protocol for MSP1 production containing an N-terminal His<sub>6</sub>-tag cleavable by the TEV protease. The protocol can also be applied to the preparation of the larger MSP1E1, MSP1E2 and MSP1E3 constructs.

#### **MSP1 Expression**

1. For high yield expression transform the MSP1-gene containing pET28b-plasmid (Merck Biosciences) into BL21 Star (DE3) cells (Invitrogen). Inoculate 4 x 600 ml sterile LB medium supplemented with sterile filtrated glucose (0.5%; w/v) in 2 liter Erlenmeyer flasks with 50 ml of fresh LB overnight cultures.
2. Grow cells at 37°C and induce expression with 1 mM Isopropyl-β-D-thiogalactopyranoside (IPTG) at OD<sub>600</sub> = 1. Incubate cells for 1 h at 37°C and then reduce temperature to 28°C for additional 4 hrs. Harvest cells by centrifugation (10 min, 6,000xg). The pellet can be stored at -20° C.

#### **MSP1 Purification**

1. Thaw pellet of 2.4 liter fermentation on ice and resuspend in 50 ml buffer MSP-C supplemented with one tablet Complete<sup>®</sup> protease inhibitor cocktail (Roche) and 1 mM freshly prepared PMSF. Add Triton X-100 from 10% (v/v) stock to a final concentration of 1% (v/v).
2. Disrupt cells by ultrasonification (3x 60 sec and 3x 45 sec) with equal cooling times in between each cycle. Centrifuge at 30,000xg for 20 min. The supernatant is filtered through 0.45 μm prior further purification.
3. Equilibrate ion metal affinity chromatography (IMAC) column (10 ml bed volume, Sepharose 6 FF, GE Healthcare) with 5 column volumes (CV) buffer MSP-A. Load the filtered supernatant of cell disruption onto the column and wash the column with 5 CV buffer MSP-A, 5 CV buffer MSP-B, 5 CV buffer MSP-C and 5 CV buffer MSP-D.
4. Elute MSP1 with buffer MSP-E in 1 ml fractions. Pool MSP1 containing fractions and

immediately set to 10% glycerol (v/v) to prevent precipitation. Dialyze pooled fractions over night at 4°C against 5 liters MSP dialysis buffer (40 mM Tris-HCl, pH 8.0, 300 mM NaCl, 10% glycerol (v/v)) with one buffer exchange after 2 hrs. MSP concentration is determined using  $A_{280}$  with the molar extinction coefficient  $\epsilon = 24,750 \text{ L}\cdot\text{mol}^{-1}\cdot\text{cm}^{-1}$  and MW = 25.3 kDa. MSP1 is aliquoted and frozen at -80°C for long time storage.

### ND Assembly

The assembly of empty NDs requires extensive optimization if homogenous samples are required. As the MSP associates with a fixed number of lipids, the MSP to lipid ratio needs to be adjusted for each MSP/lipid combination.

1. Lipid stocks are prepared by suspending lipids in water to a final concentration of 50 mM supplemented with sodium cholate (Roth) for complete solubilization. For each individual lipid specific final concentrations of sodium cholate may be required in order to obtain a clear suspension.
2. For analytical assembly, combine MSP with lipids and detergent in a final volume of 100  $\mu\text{l}$ . MSP (25  $\mu\text{M}$ ), lipids (e.g. 2 mM with MSP1 and DMPC) and detergent (e.g. DPC 0.1% w/v) are mixed and incubated for 1 h at room temperature. Induce ND assembly by dialysis against a large volume (e.g. 5 liters) of disc formation (DF) buffer (100 mM NaCl, 10 mM Tris-HCl, pH 8.0) at room temperature for 12 hrs (10 kDa MWCO, Slide-A-Lyzer, ThermoFischer). Refresh buffer and dialyse for further 24 hrs at 4°C. Centrifuge supernatant (20 min, 22,000xg) and apply to SEC analysis.
3. Load supernatant of the assembly reaction to a suitable SEC column (e.g. Superdex 200 3.2/30). Equilibrate the column with buffer ND-A. At optimal MSP to lipid ratios, the SEC elution profiles should show evenly shaped elution peaks indicating homogeneous ND populations.
4. For preparative scale ND assembly, reaction volumes can be scaled to 3 ml. Larger volumes (e.g. 10 ml) might require extended dialysis time and more frequent buffer changes.

### ND Concentration, Storage and Stability

1. Equilibrate Centricon filter devices (10 kDa MWCO, Millipore) with buffer ND-A and concentrate at 4°C to the minimal volume of 200  $\mu\text{l}$  as recommended by the manufacturer. After concentration centrifuge the ND stock for 10 min at 22,000xg. The final protein concentration is determined by the Bradford assay.
2. For short time storage (< 12 hrs) the ND stock can be kept on ice. Long time storage (up to weeks) can be achieved by flash freezing in liquid nitrogen and storage at -80°C. Avoid extensive numbers of freeze-thaw cycles as this will destroy the NDs. Thaw ND stock solutions on ice.

3. Stability depends on the type of generated ND. While MSP1-DMPC NDs can be stored at 4°C for weeks, other types will start to precipitate after a few days. Stability generally decreases with increasing numbers of double bonds in the lipid chains (e.g. DMPC NDs are more stable than DOPC NDs) and with increasing MSP size.

### **Co-translational Formation of MP/ND Complexes**

A protocol for the co-translational association/insertion of CF expressed MPs with NDs is described. Solubilization with detergents or the initial precipitation of the MPs can thus be avoided. NDs are highly tolerated by CF systems and usually do not have inhibitory effects on the expression efficiency. C-terminal fusion-constructs of MPs to sGFP allow a fast initial monitoring of the co-translational solubilization of MPs with the NDs, although the folding of sGFP may not always be completely correlated with the folding of the N-terminal MP fusion.

### **Quantifying sGFP**

1. After expression, centrifuge the reaction mixture (10 min, 22,000xg) and use supernatant for fluorescence assay.
2. Add 3 µl of the supernatant in 300 µl of GFP-assay buffer (20 mM Tris-HCl, pH 7.4, 150 mM NaCl). Measure fluorescence intensity of each reaction by at least triplicates and convert into protein concentration by a calibration curve with purified sGFP.

### **CF Production of MP/ND Complexes**

1. The synthesized MPs and the NDs must be present at least in stoichiometric amounts in order to improve the sample homogeneity. In addition, the type of lipid in the NDs can strongly affect the efficiency of complex formation with individual MPs. Initial ND concentration screens should therefore be performed for each new MP/ND combination in order to determine conditions for best solubilization. The MP is quantified in the soluble and non-soluble fraction after reactions with increasing amounts of supplied NDs. For quantification, either MP-sGFP fusions, detection of immuno-tags after western blotting or <sup>35</sup>S-Met incorporation can be used.
2. Prepare duplicate CF reactions for each ND concentration. A recommended initial screen would comprise final ND concentrations of 0 µM, 20 µM, 40 µM, 60 µM, 80 µM, 100 µM and 120 µM in the RM.
3. After reaction, centrifuge the reaction at 22,000xg for 10 min. to separate soluble and insoluble fractions and quantify the expressed MP in supernatant and in the pellet fraction.
4. If sufficient solubilization of the MP cannot be achieved, the effects of different ND types and/or different lipids should be analysed.

## Purification of MP/ND Complexes

1. For functional analysis in most cases purified MP/ND complexes are required. For optimal sample homogeneity, the NDs are usually added in excess to the CF reaction and empty NDs might need to be removed. Providing unique affinity tags at the MP, e.g. a Strep-II-tag, is therefore recommended.
2. Centrifuge reaction mixture after expression at 22,000xg for 10 min. Apply 5-fold diluted supernatant in Strep-buffer to Strep-Tactin resin (IBA) equilibrated in Strep-buffer. Perform binding in batch setup with 500  $\mu$ l Strep-Tactin resin (IBA) for each 1 ml supernatant of reaction mixture. Incubate for 12 hrs at 4°C with slightly shaking.
3. Load binding solution on an empty gravity flow column. Wash resin with 5 CV Strep-buffer. Elute protein with Strep-elution-buffer (Strep-buffer supplemented with 2.5 mM desthiobiotin). Apply 1x 0.5 CV and 4x 1 CV of Strep-elution buffer and collect main fractions.

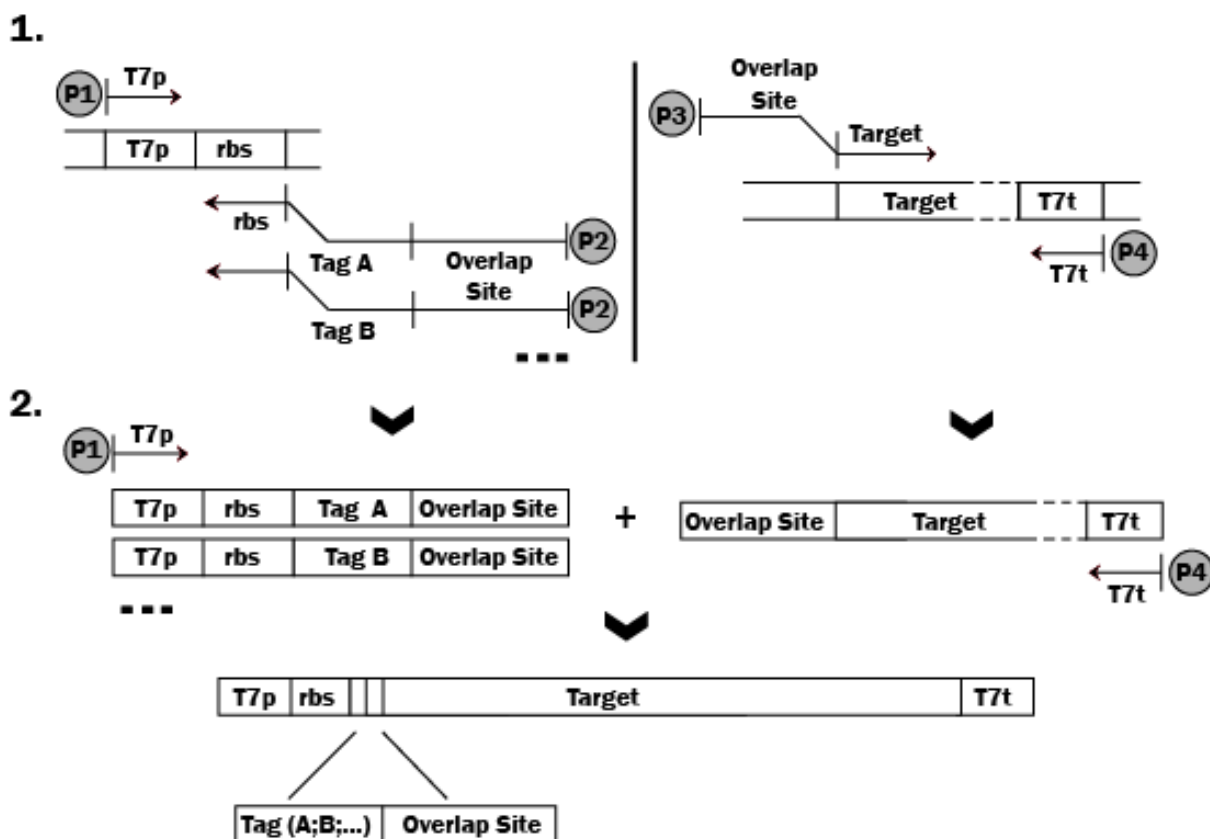


Fig 4.1 Strategy for the tag variation screen. Two subsequent PCR steps are illustrated. 1: The tag fragment(s) with primers P1/P2 and the target fragment(s) with primers P3/P4 are generated with appropriate overlap sites. 2: Corresponding tag fragments and target fragments are fused by an overlap PCR reaction using primers P1/P4. The products are suitable as CF expression templates for the tag variation screen.

## 4.2 Protein Purification

Samples were separated on 16% (w/v) acrylamide/bisacrylamide gel and stained with Coomassie blue or used for immunodetection via Western blotting and by using mouse penta-His antibodies

(Qiagen, Hilden, Germany). The CF expressed GS components were purified by immobilized metal chelate affinity chromatography (IMAC). 1 ml RM with expressed HsGNA1-His6 was mixed with 300  $\mu$ l of Ni<sup>2+</sup>-NTA resin (Qiagen, Hilden, Germany), 10-fold equilibration buffer (20 mM phosphate buffer, pH 7.2, 300 mM NaCl, 20 mM imidazole) and incubated overnight at 4 °C with gentle shaking. The mixture was then poured into an empty column and washed with 10-fold column volumes of equilibration buffer supplemented with 60 mM imidazole. GS components were then eluted with equilibration buffer supplemented with 300 mM imidazole.

For analysis by size exclusion chromatography (SEC), IMAC purified GS samples were loaded on a Superdex 75 30/ 100 column (GE Healthcare, Munich, Germany) with a bed volume of 24 ml pre-equilibrated with SEC-buffer (150 mM NaCl, 20 mM Tris-Cl, pH 8.0) at a flow rate of 0.4 ml/min.

### 4.3 SDS-PAGE and Western Blot:

The soluble fractions in CF expression system were harvested by centrifuging the reaction solution for 10 min at 13,000 rpm and precipitated with the addition of 10 times cold acetone. Then it was redissolved in 25  $\mu$ l sample loading buffer after centrifugation, air drying and heated for 10 mins at 95 °C. About 15  $\mu$ l samples were loaded onto a 16% (w/v) acrylamide/bisacrylamide gel and stained with Coomassie blue. For western blot analysis, proteins were then transferred to a Immobilon-poly (vinylidene difluoride) membrane (Millipore GmbH, Schwalbach, Germany) by electroblotting for 30 min at 350 mA. The membrane was blocked with 4% milk in PBST (0.137 M NaCl, 0.003 M KCl, 8 mM Na<sub>2</sub>HPO<sub>4</sub> and 1.5 mM KH<sub>2</sub>PO<sub>4</sub> and 0.05% (v/v) Tween 20) and probed with mouse Penta-His Antibody (Qiagen GmbH, Hilden, Germany) followed by HRP-conjugated goat antimouse IgG (Santa Cruz biotechnology, Inc., Santa Cruz, CA, USA). The protein bands were visualized with the Western Blotting Luminal Reagent (Santa Cruz biotechnology, Santa Cruz, USA) and developed using an luminescence detection system Lumi-Imager F1™ (Roche Diagnostics, Mannheim, Germany).

Compounds	Resolving gel [ $\mu$ l]		Stacking gel 4% [ $\mu$ l]
	12%	16.5%	
1.5 M Tris-HCl (pH 8.9)	2500	2500	-
0.5 M Tris-HCl (pH 6.8)	-	-	1000
Protogel™ 30%	4000	5500	1000
1% (w/v) SDS	1000	1000	500
50% Glycerol	1000	1000	-
H <sub>2</sub> O	1500		1500
TEMED	25	25	10
10%APS (w/v)	40	40	25



#### 4.4 Laser-induced liquid bead ion desorption-Mass spectrometry (LILBID)

LILBID is a novel mass spectrometry method that allows an exact mass determination of single macromolecules dissolved in droplets of solution containing an adequate buffer, pH, ion strength, etc., as described previously [1]. Briefly, droplets of solution of analyte are ejected by a piezo-driven droplet generator and transferred into a high vacuum. There, they are irradiated droplet by droplet ( $\Phi = 50 \mu\text{m}$ ,  $V \sim 65 \text{ pl}$ ,  $10 \text{ Hz}$ ) by a pulsed IR laser tuned to the stretching vibration of water at  $2.9 \mu\text{m}$ . By laser ablation the droplets explode, ejecting preformed biomolecular ions into the vacuum. The total volume of solution required for the mass determination is only few  $\mu\text{l}$  in typically  $\mu\text{M}$  concentration. The method is ideal for studying biomolecules of low availability [1]. The amount of energy transferred into noncovalent complexes by the IR desorption/ablation process can be controlled in a wide range, starting from ultrasoft to harsh conditions, just by varying the laser intensity [2]. At ultrasoft conditions large macromolecules can be detected in their native stoichiometry. The complexes are detected in different charged states, preferentially as anions. The number of charge states observed increases with the size of the molecules but is less than those observed in electrospray ionization and considerably more than in MALDI. For buffer exchange Zeba Spin columns were used following the standard procedure.

#### 4.5 Blue Native PAGE

Gel-chamber, 4-16% Bis-Tris gels (NativePAGE Novex 4-16%, 1.0 mm, 10 well) and native marker (NativeMark) also from invitrogen were used.

Anode-buffer: 50 mM Bis-Tris/HCl, pH 7

Cathode-buffer 1: 50 mM Tricine, 15 mM Bis-Tris/HCl, pH 7, 0.02% Coomassie

Cathode-buffer 2: 50 mM Tricine, 15 mM Bis-Tris/HCl, pH 7, 0.002% Coomassie

The gel was run for 1 h at 100 V with cathode buffer 1 containing 0.02% Coomassie and 1 more hour at 400 V with cathode buffer 2 containing 0.002% Coomassie

##### 4.5.1 Silver Staining

The following protocol was used from [200]

Procedure for 30-min silver-staining method <sup>a</sup>

	Steps	Solution	Time
1	Fixation	60 ml acetone stock <sup>b</sup> ; 1.5 ml TCA stock <sup>b</sup> ; 25 $\mu$ l 37% HCHO	5 min
2	Rinse	ddH <sub>2</sub> O	3 $\times$ 5 s
3	Wash	ddH <sub>2</sub> O	5 min
4	Rinse	ddH <sub>2</sub> O	3 $\times$ 5 s
5	Pretreat	60 ml acetone stock <sup>b</sup>	5 min
6	Pretreat	100 $\mu$ l Na <sub>2</sub> S <sub>2</sub> O <sub>3</sub> ·5H <sub>2</sub> O stock <sup>b</sup> in 60 ml ddH <sub>2</sub> O	1 min
7	Rinse	ddH <sub>2</sub> O	3 $\times$ 5 s
8	Impregnate	0.8 ml AgNO <sub>3</sub> stock <sup>b</sup> ; 0.6 ml 37% HCHO; 60 ml ddH <sub>2</sub> O	8 min
9	Rinse	ddH <sub>2</sub> O	2 $\times$ 5 s
10	Develop <sup>c</sup>	1.2 g Na <sub>2</sub> CO <sub>3</sub> ; 25 $\mu$ l 37% HCHO; 25 $\mu$ l Na <sub>2</sub> S <sub>2</sub> O <sub>3</sub> ·5H <sub>2</sub> O stock <sup>a</sup> ; 60 ml ddH <sub>2</sub> O	10–20 s
11	Stop	1% glacial acetic acid in ddH <sub>2</sub> O	30 s
12	Rinsed	ddH <sub>2</sub> O	10 s

<sup>a</sup> All steps were performed in glass or plastic containers on a shaker at room temperature (approx. 23°C). The volumes of all solutions were 60 ml for mini-gels 0.75 mm thick.

<sup>b</sup> Stock solutions: 50% acetone in ddH<sub>2</sub>O; 50% TCA in ddH<sub>2</sub>O; 20% AgNO<sub>3</sub> in ddH<sub>2</sub>O (store in dark; shelf-life up to 4 months); 10% Na<sub>2</sub>S<sub>2</sub>O<sub>3</sub>·5H<sub>2</sub>O in ddH<sub>2</sub>O (shelf-life about 4 months).

<sup>c</sup> A brown precipitate may appear upon contact of the gel with the developer. It can be dissolved by vigorous shaking.

## 5 Results

### 5.1 Cell-free Expression of GS Operon

$\gamma$ -secretase is an intramembrane cleaving protease that provides the final cut to Amyloid Precursor Protein to form Amyloid beta peptides (A $\beta$  37-43) that are stored as plaques in the brains of Alzheimer's patients. Therefore  $\gamma$ -secretase can be an important drug target but there is no way to measure its activity in vitro and assay inhibitors or modulators. A  $\gamma$ -secretase operon carrying T7 promoter, Pen-2, PS1, Aph1-(His)<sub>10</sub>, Nicastrin, T7 terminator respectively connected by linkers was synthesized and cell-free expression protocol was optimized to express it. After initial Mg<sup>2+</sup> ion concentration optimization, the operon was expressed as precipitate (P-CF), in detergent micelles (D-CF), and in nanodiscs (L-CF). All proteins of the complex were pulled down and purified by Ni<sup>2+</sup> metal-chelate affinity chromatography. Expression yield and solubility in various detergents were measured and compared. The presence of Aph-1 was confirmed by western blotting with Anti His antibody and other subunits were detected mass spectrometry. Electron microscopy analysis reveals the presence of heterogeneous complexes although blue native PAGE resulted in smearing. Currently further optimizations and in vitro activity assay is in progress.

#### Cell-free Expression of GS

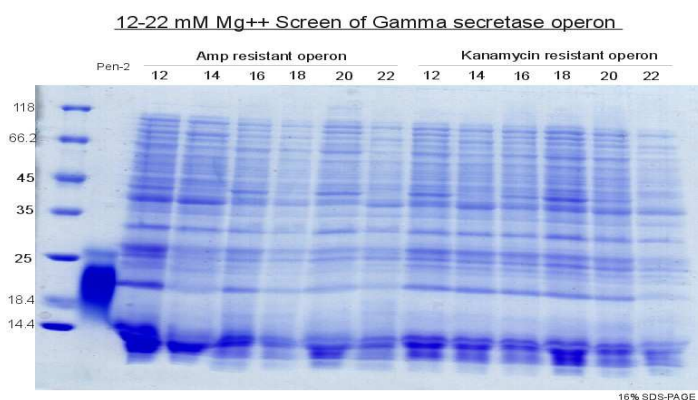


Fig 5.2 Mg<sup>++</sup> Screen of GS operon

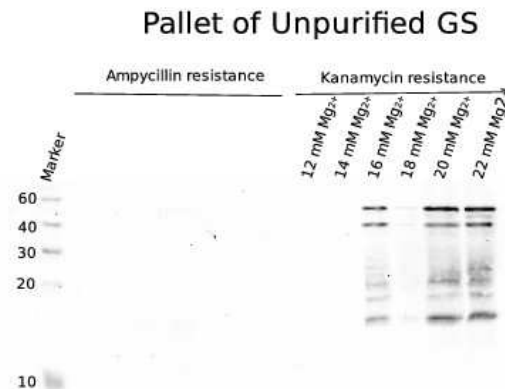


Fig 5.1 Western blot of pallet of unpurified GS

Synthesized GS operon comprises T7 promoter, Pen-2, Presenilin 1, Aph1-His 10, Nicastrin, T7 terminator respectively from N- to C-termini. Each protein is separated by linkers. Assuming only monomers in the final  $\gamma$ -secretase complex, it gives a molecular weight of  $(13.6 + 52.7 + 28.8 + 13.7) = 108.8$  kDa. Two constructs with Ampicillin Resistance (pTriEx1.1 vector- 8145 bp) and Kanamycin Resistance (pMK-RQ-Bb vector- 5328 bp) were made. The sequences of the operon were verified by sequencing with T7 promoter and terminator primer respectively. In P-CF mode expression, first the Mg<sup>2+</sup> ion concentration was optimized from 12-22 mM (Fig. 5.2) for both of the constructs. A clear lane was visible at the same level of Pen-2 control. Western blotting (WB) with anti-His antibody indicated the highest expression of Aph-1 at 20-22 mM Mg<sup>2+</sup> ion concentration

for the operon with Kanamycin resistance (Fig. 5.1). 20-22 mM Mg<sup>2+</sup> concentration was further used for expression and purification of the complex.

### Purification of GS Operon

After preparative scale P-CF expression of 1 ml, GS was resolubilized with 0.5 % LMPG. It was incubated overnight to Ni-NTA column at 4°C with 150 mM NaCl, 20 mM HEPES pH 7.0, 2 mM beta Mercaptoethanol, and 20 mM Imidazole. Thereafter the resin was poured into an empty column and the flow through due to gravity was discarded. It was washed with 10 column volumes of the same buffer supplemented with 20, 40, 60 & 80 mM Imidazole and 0.05 % LMPG respectively. The protein was eluted at 300 mM imidazole. The obtained protein concentration detected by Bradford assay was (0.15 mg/ml)

Because the amount of purified protein was low, Invitrogen's SilverQuest Silver staining kit was used to stain the 16 % SDS-PAGE gel. It is ~30 times more sensitive than coomassie G-250 staining. The purified protein migrates at 4 different MW i.e. two around 15 kDa and two around 50 kDa (Fig. 5.3).

In order to verify the presence of GS components, WB was carried out with anti-His antibody which recognizes the His tag. Since Aph-1 only contains His tag, clear bands in the WB indicates that Aph-1 is present in the expression. A thick band is missing in WB where Pen-2 migrates in SDS-PAGE of purified GS. So one can purify Pen-2 along with Aph-1. However in SDS-PAGE membrane proteins don't run at the same level of their molecular weight.

We did not attach His-tag to any other GS component because it might inhibit their activity. Pen-2 and PS1 should be also expressed because they are situated before Aph-1 in the operon. Still it is not clear whether we have all components of GS complex expressed well. Also during washing with 80 mM Imidazole there is no signal in WB whereas in the first elution fraction one can see strong signal indicating that the proteins of the GS complex could be pulled down by Aph-1 (Fig 5.4, 5.3).

Presenilin-1 could show up in the WB because it has several histidines and if they come close during folding, antibody would bind to it. Western blotting with PS1, Pen-2 and Aph-1 antibodies was undertaken without any signal probably due to the fact that the antibodies were raised against endogenous protein. However Pen-2 control expressed by cell-free expression gives signal.

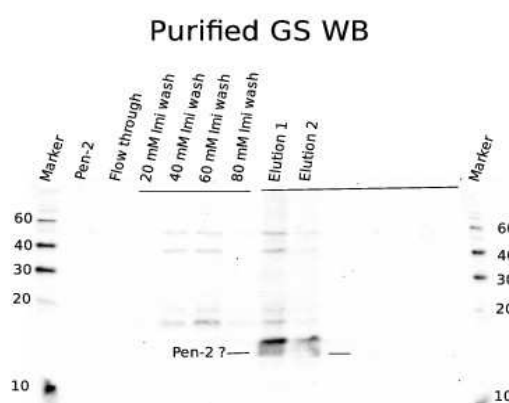


Fig 5.4

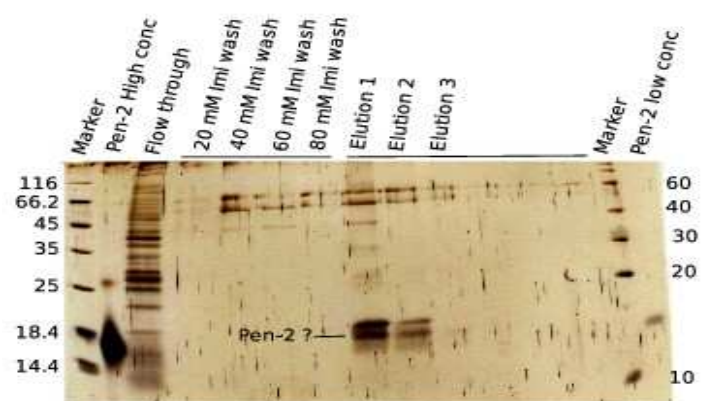


Fig 5.3

In order to further verify the presence all components of the GS complex in the pull-down, Laser-induced Liquid Bead Ion Desorption-MS (LILBID) was undertaken. LILBID is a novel mass spectrometry method that allows an exact mass determination of single macromolecules dissolved in droplets of solution containing an adequate buffer, pH, ion strength, etc.

LILBID was performed on the elution fractions after dialysis to 50 mM NaCl. Before the experiment buffer was exchange from HEPES/LMPG into TRIS/DDM. Desalting in HEPES/LMPG gave no signals. Desalting in Tris/DDM gave signals, but only at harsh conditions - i.e. most likely breaking any complexes. Softer condition gave worse signals. So the LMPG to DDM transition was only partially successful. However in harsh conditions signals were better (Fig. 5.5).

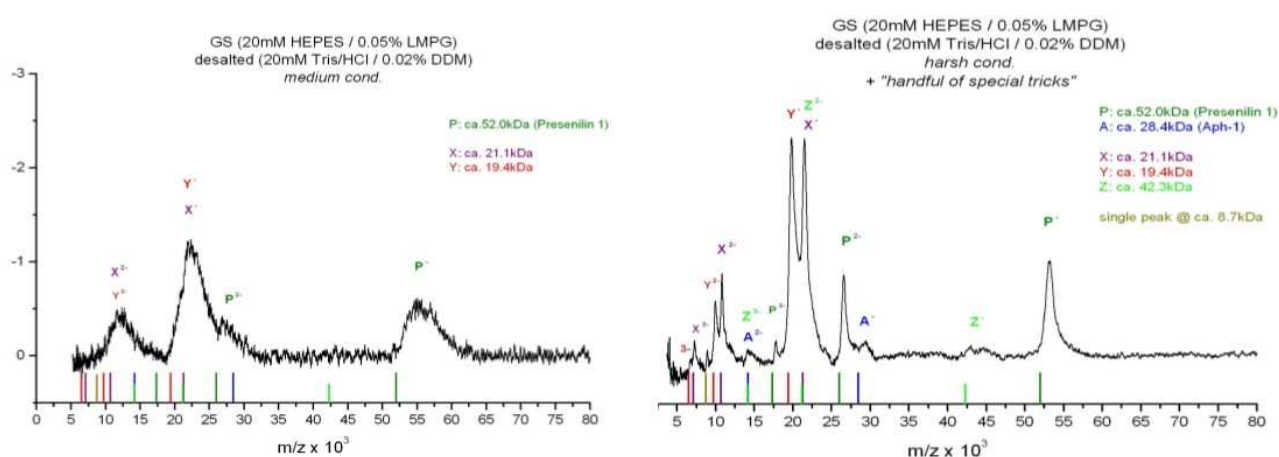


Fig 5.5: LILBID mass spectra of expressed GS components. (left) In medium condition showing only the Presenilin 1 (right) in harsh condition showing Presenilin 1 and Aph-1.

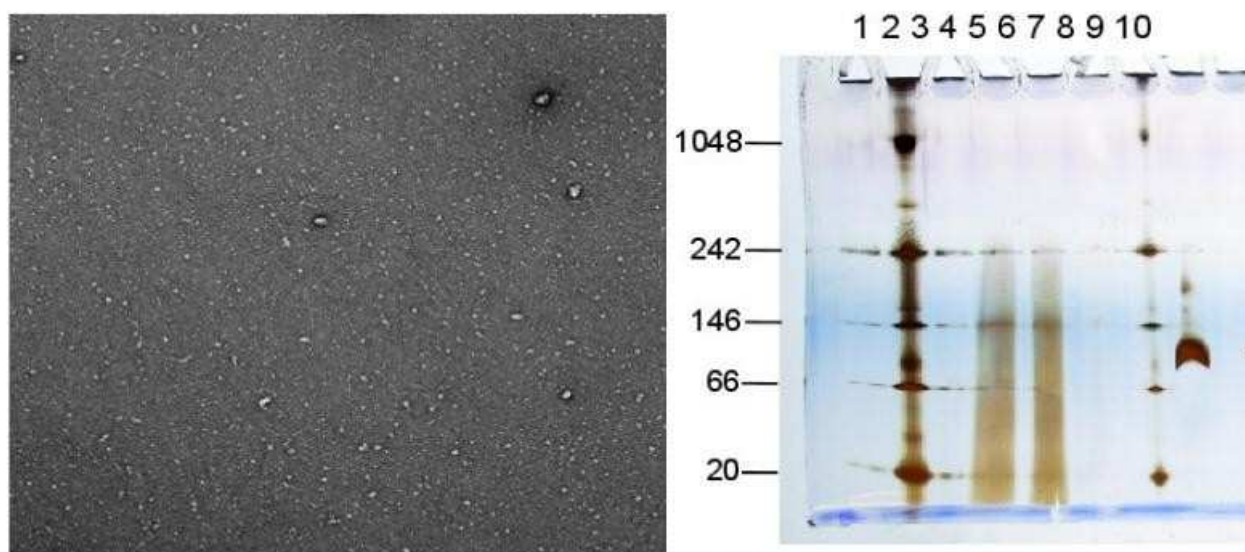


Fig 5.6: (left) Electron Microscopy (EM) of GS showing presence of heterogeneous particles. (right) Blue Native Page of expressed GS. Lane 2: Invitrogen BN-marker, 4: GS LMPG 5: 0.05% DDM GS 6: 0.02 % DDM GS 7: Invitrogen BN-marker 8: c-ring (91 kDa)

In order to identify the existence of GS complex after pull down, EM images were taken after

negative staining (Fig.5.6). In the EM images, many seem to be single particles and they seem to lie in different orientations on the grid, which is good if one is aiming for 3-D reconstruction. But also larger objects (dimer and trimer) can be seen. One has to take many more images of highly purified protein and do single particle analysis. However the sample needs to be optimized, which is still heterogeneous.

Also blue native page was carried out (Fig. 5.6) to identify the presence of complex after pull down. However only a smear was obtained in case of GS solubilized in DDM and LMPG. So it needs to be optimized and need to be performed just after purification.

Furthermore, GS operon was expressed in D-CF mode with detergents like Brij 58, 78 etc and in WB of pallet had similar pattern like in P-CF expression mode.

### **Conclusions and outlook**

- GS operon consisting of Pen-2, Presenilin 1, Aph1-His10, Nicastrin was expressed and purified by continuous-exchange cell-free expression strategy.
- Presenilin 1 and Aph1 proteins were identified by LILBID-MS and Pen-2 by SDS-PAGE and WB.
- Activity assay of the purified complex is required to establish its enzymatic activity in vitro. Further Electron Microscopy needs to be performed for single particle reconstitution.

## 5.2 Structural Investigation of Pen-2

Presenilin enhancer 2 (Pen-2) is a structural subunit of the  $\gamma$ -secretase complex. Pen-2 is 101 amino acid long integral membrane protein with two membrane spanning helices and both N-terminal and C-terminal facing extracellularly. It regulates the endoproteolytic cleavage of Presenilin 1 into C-terminal and N-terminal fragments required for its functioning. Recent studies show that Pen-2 and Presenilin 1 are able to cleave the substrates in vitro.

### Methods:

*Expression and Purification of Pen-2:* Human gene of Pen-2 was cloned into the pET21a(His)10 vector. PCR based N-terminal tag variation of Pen-2 followed by  $Mg^{2+}$  ion concentration optimization was performed in order to improve its expression.

*Quality and Stability assay of Pen-2 and NMR:* Homogeneity of purified Pen-2 in various detergents was checked by Superdex 200 3.2/30 column and stability was checked performing Circular Dichroism (Secondary Structure and Melting Curve) experiments.

*NMR studies from Pen-2 expressed in P-CF mode:* Pen-2 was expressed as precipitate in 3 ml reaction mix of continuous exchange cell free (CECF) expression mode. It was resolubilized in 2% Fos-14 and bound overnight to Ni-NTA column with 20mM Tris-HCl pH 7.0, 300 mM NaCl, 20 mM  $\beta$ -Mercaptoethanol and 20 mM Imidazole. The column was washed gradually with the same buffer and 10 column volume of 20 mM, 40 mM and 60 mM Imidazole and eluted with 5 column volume of 300 mM Imidazole. Purified Pen-2 was dialyzed to 20 mM sodium acetate buffer pH 5.0 and 5 mM  $\beta$ -Mercaptoethanol and concentrated before recording [ $^{15}N,^1H$ ] TROSY spectra of uniformly [ $^{15}N,^1H$ ] labeled Pen-2.

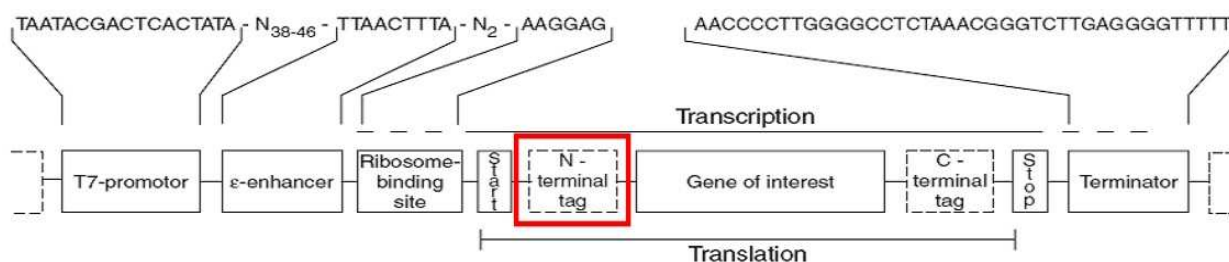
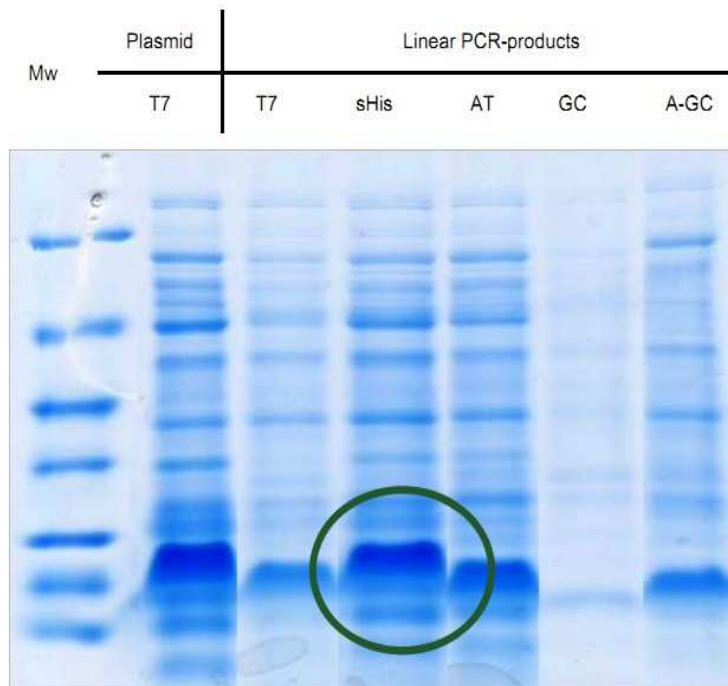


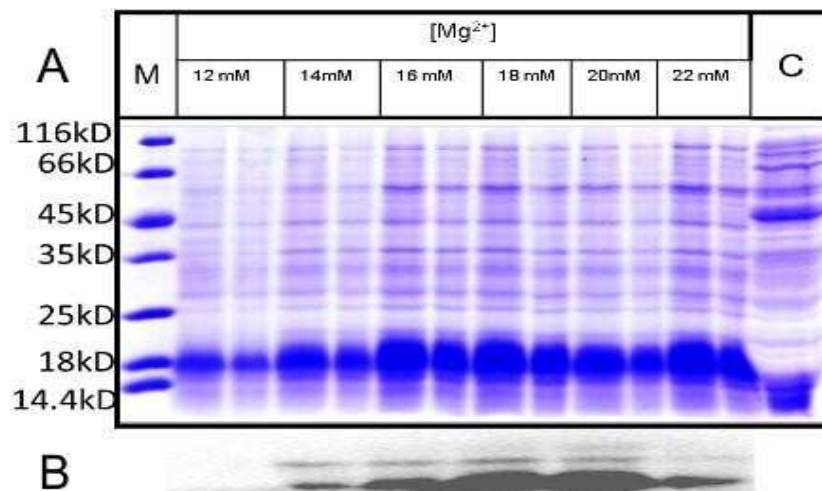
Fig 5.7: Scheme for Tag variation

The scheme of tag-variation is shown above where the N-terminal tag of the gene which is preceded by T7-promoter, epsilon enhancer, ribosome binding site and start codon respectively and followed by Gene of Interest, C-terminal tag, stop codon and terminator respectively. This N-terminal tag was

varied in order to boost the expression of Pen-2.



**Fig 5.8: Tag variation screen of Pen-2.** The expression with sHis tag was the highest compared to T7, AT, GC, A-GC tag. sHis tag consists of 6 Serine residues (DNA sequence: TGA TGA TGA TGA TGA TGA TGA ). Final sequence of Pen-2 is as follows: **M S S S S S L E V L F Q G P M N L E R V S N E E K L N L C R K Y Y L G G F A F L P F L W L V N I F W F F R E A F L V P A Y T E Q S Q I K G Y V W R S A V G F L F W V I V L T S W I T I F Q I Y R P R W G A L G D Y L S F T I P L G T P L E H H H H H H H H H H H**



**Fig 5.9 Optimization of Mg<sup>++</sup> ion concentration.** From 16-18 mM of Mg<sup>2+</sup> the expression of Pen-2 was the highest. Moreover optimal K<sup>+</sup> ion concentration was 290 mM. The results were confirmed by western blot.



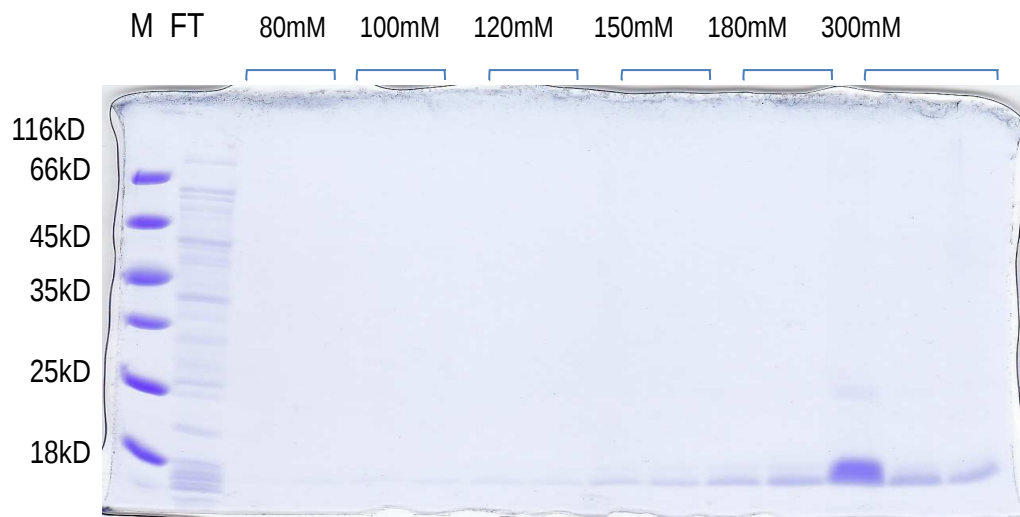


Fig 5.10: **Purification of Pen-2.** NiNTA elution profile of Pen-2 : After purification Pen-2 gives band below 18 kDa.

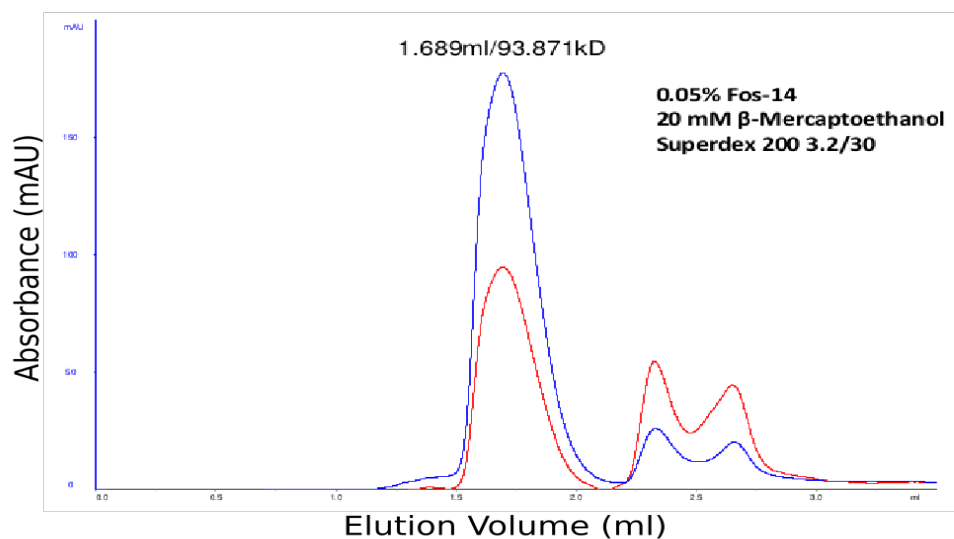


Fig 5.11 In order to check the homogeneity and stability of the purified Pen-2 sample in various detergents, size-exclusion chromatography was carried out after keeping it at room temperature for two days. First Gaussian like peak at 1.6 ml elution volume corresponds to Pen-2 micelle. Running Buffer of 0.05% Fos-14, 20 mM  $\beta$ -Mercaptoethanol, Superdex 200 3.2/30 column was used for separation. x-axis: elution volume (ml) y-axis: blue line- absorbance at 280 nm (arbitrary units), red line: - absorbance at 260 nm (arbitrary units)

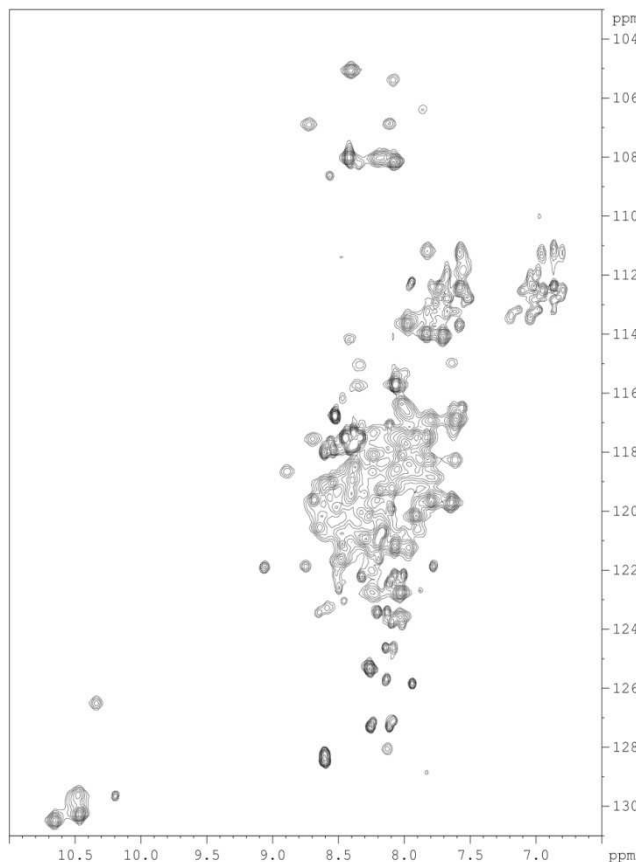
**[<sup>15</sup>N,<sup>1</sup>H] TROSY HSQC Spectra of Pen-2 in various detergents:**

Fig 5.12 2% Fos-14

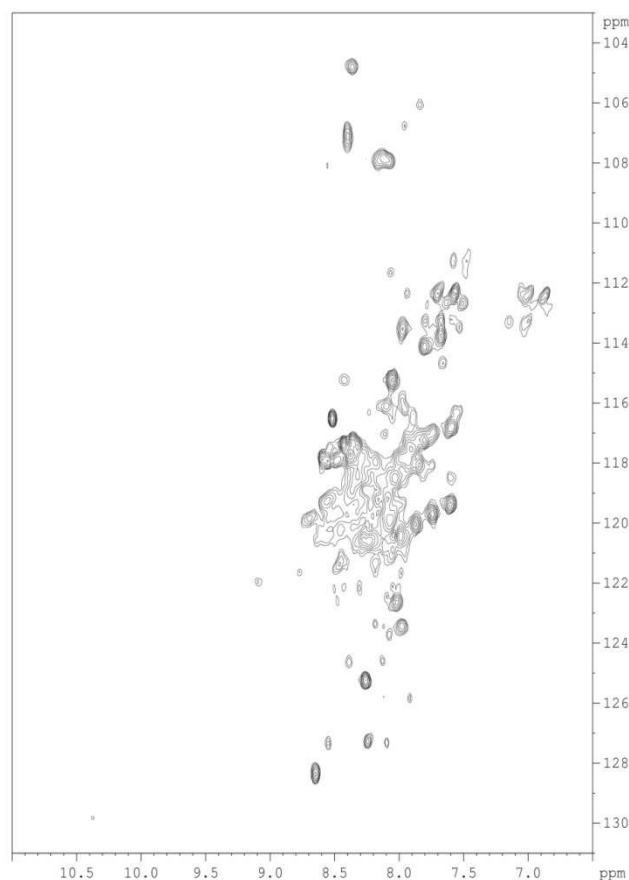


Fig 5.13 0.4% Fos-16

Fig 5.15 [<sup>15</sup>N,<sup>1</sup>H] TROSY spectra of uniformly [<sup>15</sup>N,<sup>1</sup>H] labeled 0.2 mM Pen-2 resolubilized in 2% Fos-14 , 0.4% Fos-16 and 1% LDAO, 20 mM sodium acetate buffer pH 5.0 and 5 mM β-Mercaptoethanol recorded at 303K in 600 MHz Bruker spectrometer. For Fos-14, Fos-16, & LDAO 9, 5,6 glycine and 5, no & 6 tryptophan peaks are visible respectively.

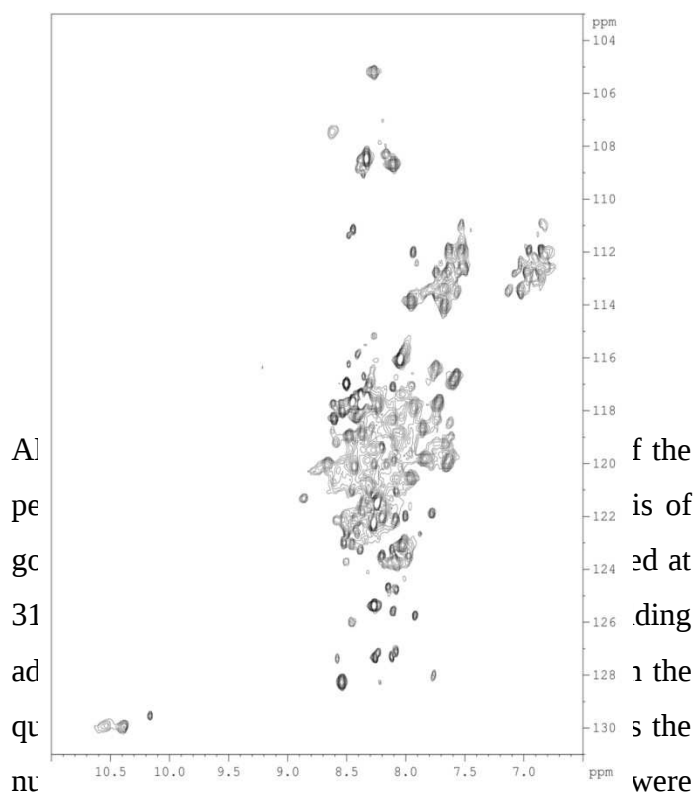


Fig 5.14: 1% LDAO

All  
pe  
gc  
31  
ad  
qu  
nu

f the  
is of  
ed at  
ding  
1 the  
s the  
were

obtained for 1% starting concentration of Fos-14 although with around 4% the spectra were little bit worse. Spectra for Fos-14 at pH 6.5 did not provide any improvement.

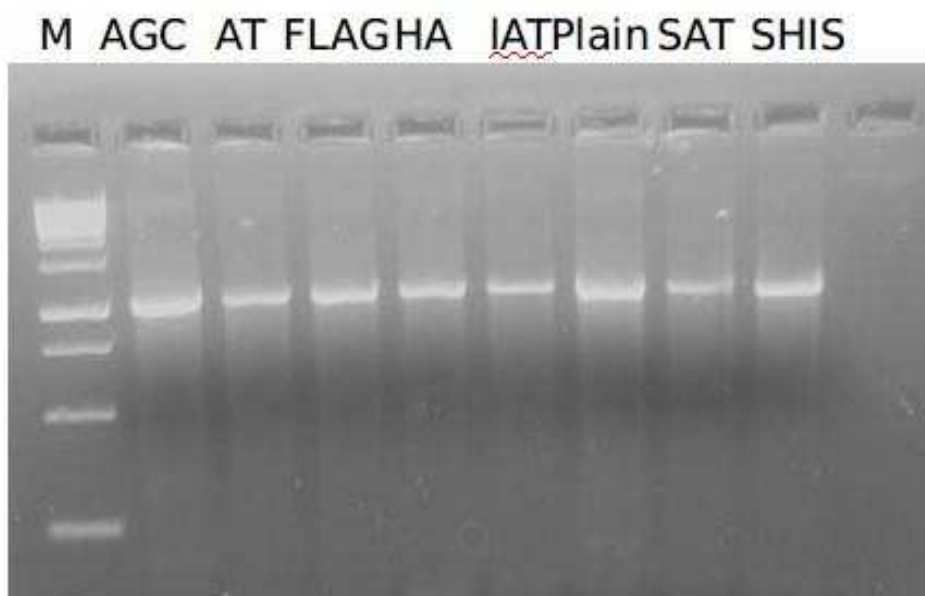
The possibility of degradation and oligomerisation of Pen-2 was addressed by adding 1x Complete protease inhibitor and TCEP bond breaker respectively without any change in recorded spectra. However all the recorded spectra are not quite well resolved and lack homogeneous peaks.

### Conclusions and Perspectives

- Pen-2 can be highly expressed in P-CF, D-CF and L-CF mode of cell-free expression.
- Among various detergent conditions, best [ $^{15}\text{N}$ , $^1\text{H}$ ]-TROSY spectra were observed at 600 MHz with Pen-2 dissolved in 2 % Fos-choline-14. However it needs further improvement before starting 3D experiments.

### 5.3 Aph-1 tag variation

N-terminal tag of Aph-1 was varied to different tags as given the figure below. However no expression was detected in the western blot after analytical scale CF reaction.



*Fig 5.16: Agarose gel electrophoresis of linear Aph-1 constructs after N-terminal tag variation*

## Part II: Molecular Modeling of the interactions of $\gamma$ -Secretase Components

### 6 Introduction

#### 6.1 Molecular Dynamics (MD) Simulations

MD simulations are used to describe time evolution of a system containing protein(s), DNA/RNA, water, ions etc by numerically solving Newton's equations of motion for all atoms in the system. However, MD simulations have the following approximations:

- Born-Oppenheimer approximation where the wave functions describing nuclear and electronic motions are decoupled
- Nuclei is treated classically by Newtonian mechanics
- Interactions of particles are described by empirical force field (FF)

##### 6.1.1 Born-Oppenheimer approximation

Dynamical time dependent evolution of a system can be described by time dependent Schrödinger

equation:  $i\hbar \frac{\partial}{\partial t} \Psi = \hat{H} \Psi$  where the reduced Planck's constant  $\hbar = h/2\pi$ ,  $i$  is the imaginary unit,  $\Psi$  is the wave function,  $\hat{H}$  is the Hamiltonian operator describing the total energy of the wave function. It

takes various form depending on the system. In three dimensions, it becomes  $-\frac{\hbar^2}{2m} \nabla^2 + V(\mathbf{r})$  where  $\nabla^2$  is laplacian and position of the particle is  $\mathbf{r} = (x, y, z)$ ,  $V(\mathbf{r})$  is the potential energy of the particle and  $m$  is the mass.

In the Born-Oppenheimer approximation, the wave functions representing the fast motion of the electrons are separated from the slow moving nucleus. So the total wave function can be expressed as multiplication of electronic and nucleus wave function:

$$\Psi_{\text{total}} = \psi_{\text{electronic}} \times \psi_{\text{nuclear}} \text{ to be more precise: } \Psi_{\text{tot}}(\mathbf{r}, \mathbf{R}) = \psi_e(\mathbf{r}; \mathbf{R}) \times \psi_n(\mathbf{R})$$

where  $\mathbf{R}$  and  $\mathbf{r}$  denotes the coordinates of nuclei and electrons respectively. Due the separation, movement of nucleus in the potential energy surface can be given by solving time independent Schrödinger equation for the electrons:  $\hat{H}_e \Psi_e = E_e \Psi_e$

##### 6.1.2 Classical Newtonian Mechanics of Particles

Despite separating nuclear and electronic wavefunctions, solving Schrödinger equation for each nucleus in a macromolecule of protein/DNA/RNA is still computationally too expensive. Therefore an additional approximation that nucleus obeys Newton's equations of motion has been taken:

$\mathbf{F}_i = m_i \frac{d^2 \mathbf{R}_i(t)}{dt^2}$  where  $\mathbf{F}_i$  is the force on atom  $i$ ,  $m_i$  is its mass,  $\mathbf{R}_i$  its coordinates. The force on the atom  $i$  is given by the negative distance derivative of the potential energy.  $\mathbf{F}_i = -\nabla_i V(\mathbf{R})$  The acceleration  $\frac{d^2 \mathbf{R}_i(t)}{dt^2}$  results in the change in velocity and position of the atom.  $\Delta t$  is chosen in such a way that fastest motions in the system can be captured which arise from the bond and angle vibrations. The bond vibrations due to light hydrogen atoms occur in femtosecond(fs) timescale limiting the  $\Delta t$  to 1 fs. To extend the timestep, algorithms like SHAKE [201] and LINCS [202] were developed to constrain the bond length. A parallel version of LINCS called P-LINCS has been used throughout the text [203].

Leap-frog integration algorithm was used for numerical integration of Newton's equations of motion: [204] It uses positions  $\mathbf{r}$  at time  $t$  and velocities  $\mathbf{v}$  at time  $t - \Delta t / 2$  and it updates positions and velocities using the forces  $\mathbf{F}(t)$  determined by the positions at time  $t$  using the following equations:

$$v(t + \frac{1}{2}\Delta t) = v(t - \frac{1}{2}\Delta t) + \frac{\Delta t}{m} F(t) \quad r(t + \Delta t) = r(t) + \Delta t v(t + \frac{1}{2}\Delta t)$$

### 6.1.3 Force Fields (FF)

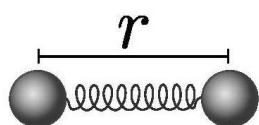
Although Born-Oppenheimer approximation provides a way to express electronic wave function as a function of nuclear coordinates, still the solution of time-independent Schrödinger equation to obtain electronic wave function is computationally too expensive. Therefore the potential energy of the system is expressed in terms of the following functions:

$$V = V_{bond} + V_{angle} + V_{dihedral} + V_{improper} + V_{nonbonded}$$

In case of AMBER force-field,

$$E_{total} = \sum_{bonds} K_r (r - r_{eq})^2 + \sum_{angles} K_\theta (\theta - \theta_{eq})^2 + \sum_{torsions} \frac{V_n}{2} [1 + \cos(n\phi - \gamma)] + \sum_{nb\ pairs} \frac{q_i \cdot q_j}{4 \pi \epsilon_0 D(r) r_{ij}} + \sum_{nb\ pairs} \left( \frac{A}{r_{ij}^{12}} - \frac{B}{r_{ij}^6} \right)$$

### Bond potentials

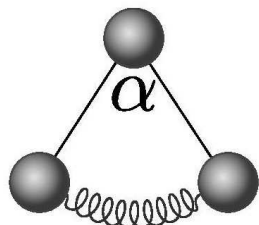


Harmonic:  $V_{\text{B}_{\text{harm}}}(\mathbf{r}) = \frac{1}{2}k_{\text{B}}(\mathbf{r} - \mathbf{r}_0)^2$

standard potential for covalent bond-stretching.

Morse:  $V_{\text{B}_{\text{morse}}}(\mathbf{r}) = k_{\text{B}}[1 - \exp(\beta(\mathbf{r} - \mathbf{r}_0))]^2$

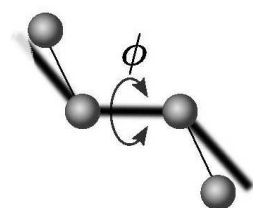
allows anharmonic bond-stretching and breaking of covalent bonds.



### Angle potential

Harmonic:  $V_{\text{B}_{\text{harm}}}(\alpha) = \frac{1}{2}k_{\alpha}(\alpha - \alpha_0)^2$

standard angle potential, e.g. modeling tetrahedral bond angles from  $sp^3$  hybridization.



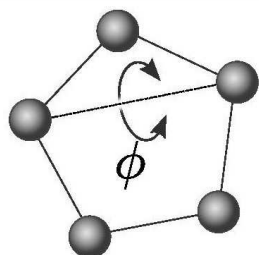
### Dihedral potentials

Periodic:  $V_{\text{dih}}(\phi) = k_{\phi}[1 - \cos(n \cdot \phi - \phi_0)]$

Ryckaert-Bellemans:  $V_{\text{RBdih}}(\phi) = \sum_{n=0}^5 C_n(\cos(\phi - \pi))^n$

cosine power series potential.

dihedral potentials for sterical restriction, e.g. due to Pauli-repulsion of orbitals from first and fourth atom.

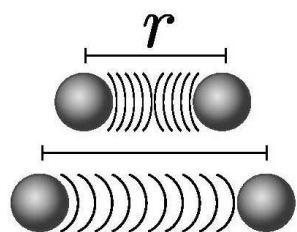


### Improper dihedral potential

Harmonic:  $V_{\text{imp}}(\phi) = \frac{1}{2}k_{\phi}(\phi - \phi_0)^2$

Improper dihedral potentials atoms restrain atoms to a plane, e.g. in conjugated planar ring systems with delocalized  $\pi$ -electron systems.

### Combined repulsion and dispersion potentials



Lennard Jones:  $V_{\text{LJ}}(\mathbf{r}) = \frac{C_{(12)}}{r^{12}} - \frac{C_{(6)}}{r^6}$

standard potential for repulsion and dispersion

Buckingham:  $V_{\text{B}}(\mathbf{r}) = A \cdot \exp(-B \cdot |\mathbf{r}|) - \frac{C}{r^6}$

more realistic repulsion (3 free parameters), computationally expensive due to expansion of the exponential function.

repulsion and dispersion usually cut-off at distances of  $\approx 1$  nm

### Limitations

In a force field, the electrons are at ground state, force fields are pair additive, the boundary of the system is periodic (using periodic boxes of different shapes like cubic, rhombic dodecahedron, truncated octahedron etc), the long range interactions are cut-off.

## 6.2 Reduced Representation of the system

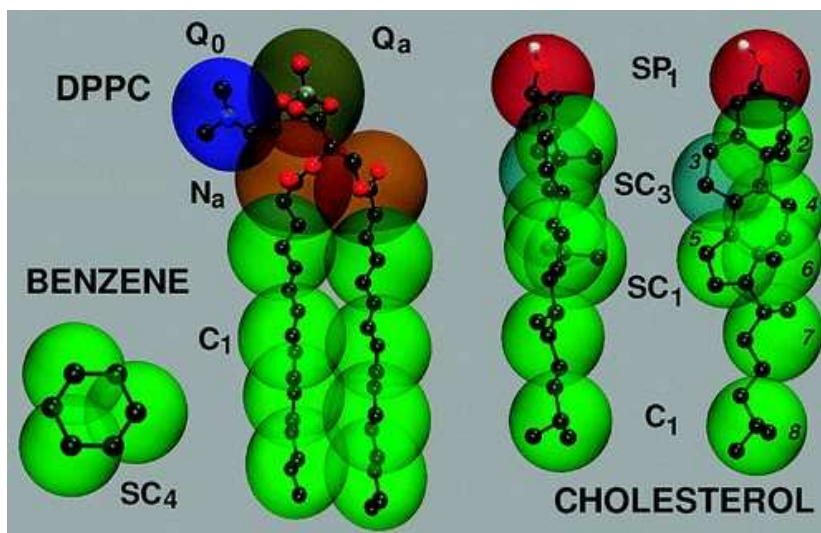
Length of simulation that can be achieved by all-atom or united atom (no non-polar hydrogens) MD simulations is limited only to few  $\mu$ s range. However, most of the biological processes take place in much longer time scales like protein folding which happens in millisecond time scale, whereas membrane protein folding can take up to a second. Also dynamics of large protein complexes and assembly of protein molecules are also very computationally demanding. This issue can be addressed by two ways:

- Coarse graining : by reducing the degrees of freedom of the system by representing a set of atoms by one grain (i.e. one grain for 4 atoms or 1:4 mapping like in Martini FF)
- Implicit Solvation: by removing the water which often require up to 70% of all atoms in the system by implicit solvent (e.g. in IMM1 or EEF1 FF)

Here such FFs for membrane protein simulation in lipid bilayer will be explained.

### 6.2.1 Coarse-graining

#### Martini



*Fig 6.1 Mapping between the chemical structure and the coarse grained model for DPPC, cholesterol, and benzene. The coarse grained bead types which determine their relative hydrophilicity are indicated. The prefix “S” denotes a special class of CG sites introduced to model rings.*

Martini is a popular coarse grained force field used in biomolecular simulations involving membrane proteins in lipid bilayer [205–207] In It is parametrized by reproduction of partitioning free energies between polar and apolar phases of many chemical compounds. Martini, the coarse-graining is achieved by mapping 4:1 which means that 1 grain substitutes 4 heavy (not hydrogen) atoms with total 20 beadtypes. Such mapping is applied for all parts of the system: protein, water molecules (one

water grain is formed from 4 all-atom water molecules) and phospholipids or detergents. This approach reduces the number of building blocks which results in faster computation, longer timestep and much longer overall time of simulation compared to all atom simulations (even three orders of magnitude). Therefore, MARTINI is an excellent method to investigate processes occurring at a timescale of microseconds or higher. It has been implemented in leading MD packages like Gromacs and NAMD. MARTINI was primarily developed to study the behavior of large biological membrane systems. Using this method a range of processes have already been investigated including vesicle fusion [208,209], membrane domain formation [210] and also formation of micelles and membranes from randomly distributed lipid or detergent molecules [211]. In recent times, it has been utilized to investigate the syntaxin-1A sequestering by ionic protein-lipid interactions [212] and mechanism of opening of mechano-sensitive MscL channels embedded in liposome [213].

Therefore, in Martini coarse grained forcefield :

- Four heavy atoms are represented by a single interaction center (including the water molecules as the solvent is explicit in this model)
- There are four main types of interaction sites: polar (P), nonpolar (N), apolar (C) and charged (Q)
- A shifted Lennard Jones 12-6 potential energy function has been used to describe the non-bonded interactions. The value of  $\epsilon$  ranges from 5.6 kJ/mol to 2.0 kJ/mol,  $\sigma=4.7\text{\AA}$

$$V_{Lennard-Jones} = 4\epsilon_{ij} \left[ \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{r_{ij}} \right)^6 \right]$$

$$V_{el} = \frac{q_i q_j}{4\pi\epsilon_0\epsilon_{rel}r_{ij}}$$

- The charge groups (type Q) bear a full charge  $q_{ij}$  interacting by Coulombing potential energy function. Coulombic energy function  $\epsilon_{rel} = 15$  for explicit screening

- The bonded interactions are given by weak harmonic potential similar to the one explained before.

See Fig 6.1 and Fig 6.2 for the diagram of Martini representations of lipids and proteins respectively.



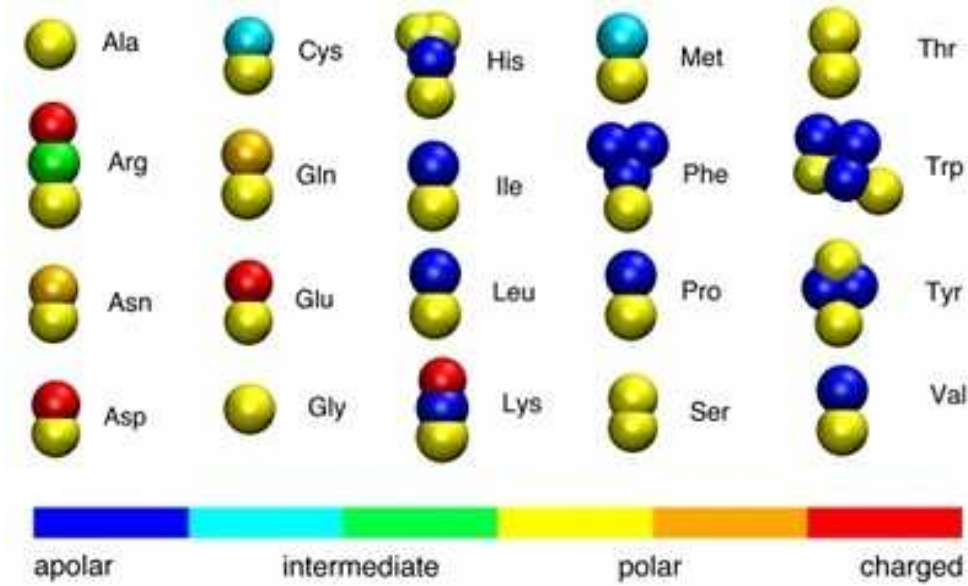


Fig 6.2: Coarse-grained representation of all amino acids. Different colors represent different particle types.

## 6.2.2 Implicit Solvation

### Effective Energy Function for Proteins in Solution (EEF1)

EEF1 [214] is based upon the CHARMM19 polar hydrogen force-field with modification in the implicit solvation terms (interactions of the atoms with solvent):

$$\Delta G_i^{solv} = \Delta G_i^{ref} - \sum_j \int_{V_j} f_i(r) dr \quad \text{or more simply} \quad \Delta G_i^{solv} = \Delta G_i^{ref} - \sum_{j \neq i} f_i(r_{ij}) V_j$$

$\Delta G^{solv}$  is the solvation free energy of atom atom,  $r_{ij}$  is the distance between i and j.  $\Delta G^{solv}$  is the solvation free energy of the atom fully exposed to solvent,  $f(r)$  is the solvation free energy density

which is modeled as a Gaussian function :  $4\pi r^2 f_i(r) = \alpha_i \exp(-x_i^2)$ ,  $x_i = \left(\frac{r - R_i}{\lambda_i}\right)$  where  $R_i$  is the van der Waals radius of i (one half the distance to the energy minimum in the Lennard–Jones potential),  $\lambda_i$  is a correlation length (3.5 Å for most atoms), and  $\alpha_i$  is a proportionality coefficient.

The distance dependent dielectric constant ( $\epsilon=r$ ) was used for representing electrostatic interactions.

### Implicit Membrane Model 1 (IMM1)

In the IMM1 model [215], the membrane is positioned as parallel to the xy plane with its center at  $z=0$ . The solvation parameters ( $\Delta G_i^{ref}$ ,  $\Delta G_i^{free}$ ) of the atoms depend on the vertical direction  $z$  or  $z' = |z|/(T/2)$  where  $T$  is the thickness of the nonpolar core of the membrane (20-30 Å) depending of the type of lipid.

$\Delta G_i^{ref}(z') = f(z')\Delta G_i^{ref,H_2O} + (1 - f(z'))\Delta G_i^{ref,mem}$  where the function  $f(z')$  describes the phase transition.  $f(z') = \frac{z'^n}{1 + z'^n}$ ,  $n$  controls the stepness of the transition. For example,  $n=0$  gives the region of 6 Å over which the environment goes from 90% nonpolar to 90% polar.

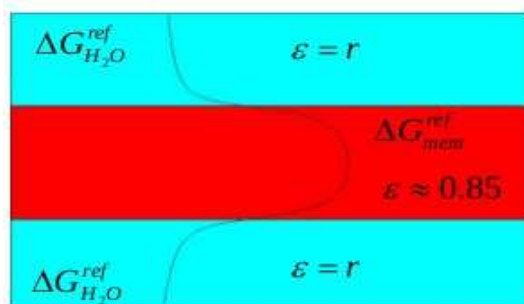


Fig 6.3: The IIM1 model, red is the hydrophobic core and cyan blue is the implicit water. The transition between polar to nonpolar environment is given by the Gaussian curve.

The electrostatic interactions are represented by a modified distance dependent dielectric screening in order to strengthen the electrostatic interactions in the membrane:  $\epsilon = r^{f_{ij}}$  where  $f_{ij}$  depends on the position of the interacting atoms with respect to the membrane which is given by an empirical model:  $f_{ij} = a + (1 - a)\sqrt{f_i f_j}$  Far from membrane  $f_{ij}=1$ , and  $\epsilon=r$ . In the membrane  $f_{ij}$  is equal to  $a$ , and empirically that value was set to  $a=0.85$ .

### 6.3 Analysis of MD Trajectories

#### 6.3.1 Root Mean Square Deviation (RMSD)

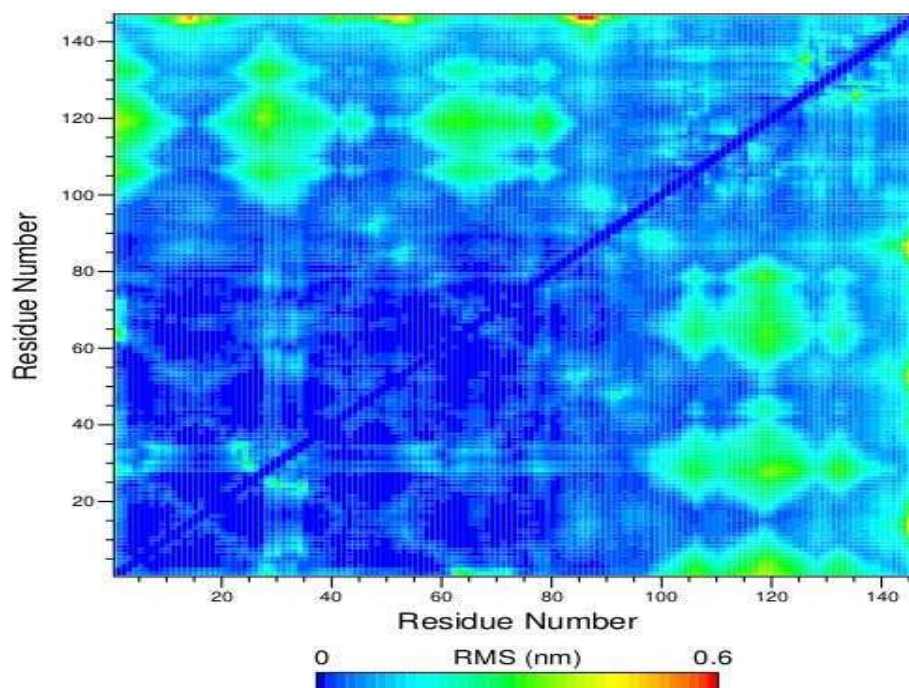


Fig 6.4: Matrix of RMS Deviations

RMSD of specific atoms in a molecule with respect to a reference structure can be obtained by least-square fitting the structure to the reference structure ( $t_2=0$ ) and calculating RMSD by

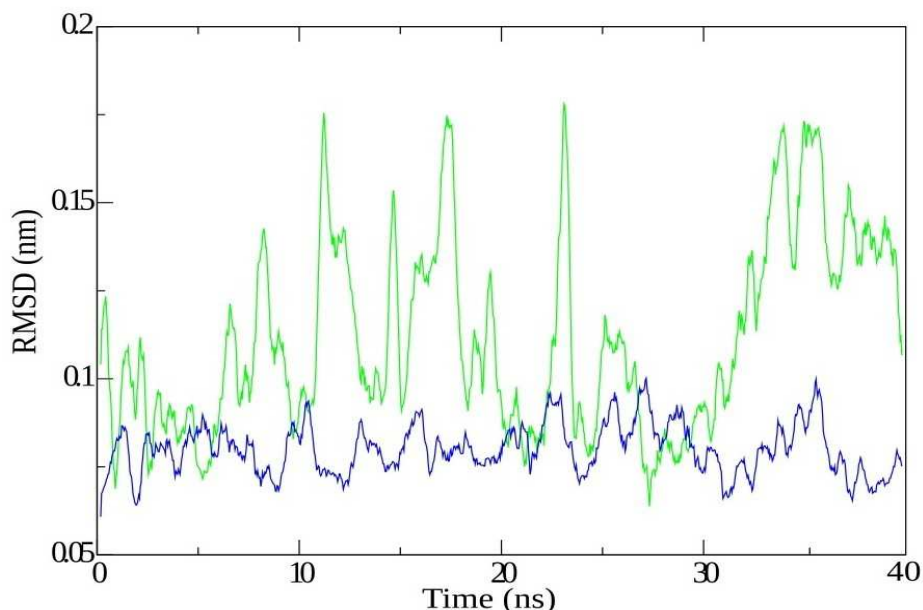


Fig 6.5: Time evolution of RMSD for two group of atoms (green and blue)

$$RMSD(t_1, t_2) = \sqrt{\frac{1}{M} \sum_{i=1}^N m_i \|r_i(t_1) - r_i(t_2)\|^2} \text{ where } M = \sum_{i=1}^N m_i \text{ and } r_i(t) \text{ is the position of the atom } i \text{ at time } t.$$

Fitting does not use the same atoms for calculation since the fitting is usually obtained by using backbone atoms (N, C-alpha, C) although the RMSD is calculated for the whole protein or other atoms than backbone.

### 6.3.2 Secondary Structure

The change in secondary structure content (Helix, Beta Sheet, Coil etc) during simulation can be obtained by the DSSP program [216]. The DSSP output assigns each residue a letter according to its secondary structure: H = alpha helix, B = residue in isolated beta-bridge, E = extended strand, participates in beta ladder, G = 3-helix (3/10 helix), I = 5 helix (pi helix), T = hydrogen bonded turn, S = bend

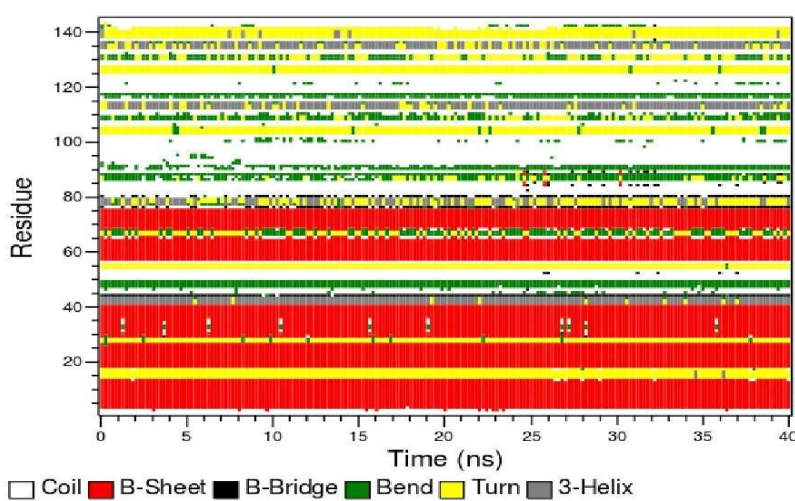


Fig 6.6: Change in secondary structure content over 40 ns simulation

A Ramachandran plot is the projection of the structure between two residues on the two dihedral angles  $\phi$  and  $\psi$  of the protein backbone ( $\phi$  : C-N-CA-C,  $\psi$  : N-CA-C-N).

### 6.3.3 Root Mean Square Fluctuations (RMSF)

RMSF is defined as :  $RMSF = \sqrt{\frac{1}{T} \sum_{t_j=1}^T |x_i(t_j) - \tilde{x}_i|^2}$  where T is the total time over which

average will be taken,  $\tilde{x}_i$  is the reference position of the particle i which is often the time averaged position of the particle i.

In case of RMSF, the average is calculated over time. In case of RMSD, the average is calculated over particles giving rise to time dependent values. RMSF is useful for checking the flexibility of the protein especially at the active site and then color the structure according to b-factor

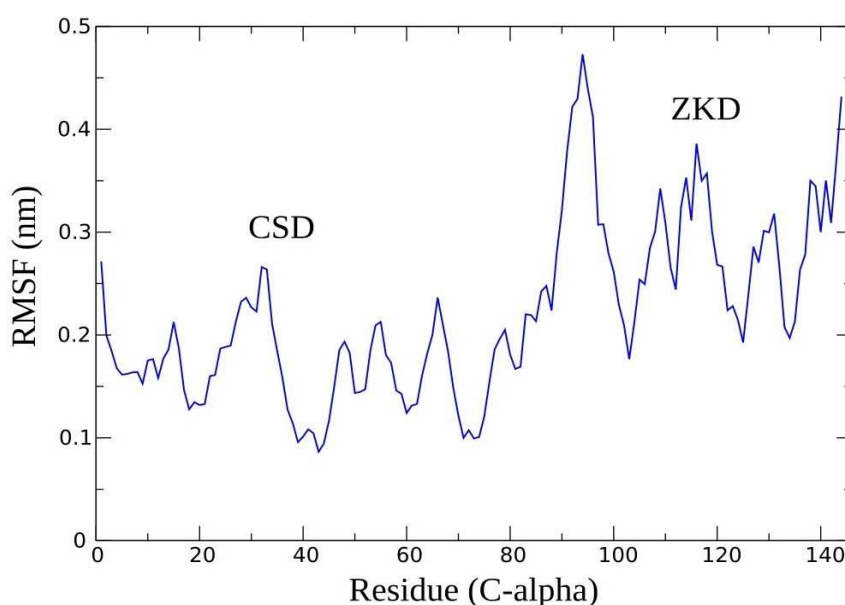


Fig 6.7: RMS Fluctuations

### 6.3.4 Solvent Accessible Surface Area

#### Area

Solvent accessible surface area around protein (both hydrophobic and hydrophilic) can be calculated.

### 6.3.5 Principal Component Analysis (PCA)

Covariance analysis or PCA or essential dynamics is used to find correlated motions of atoms. It approximate the

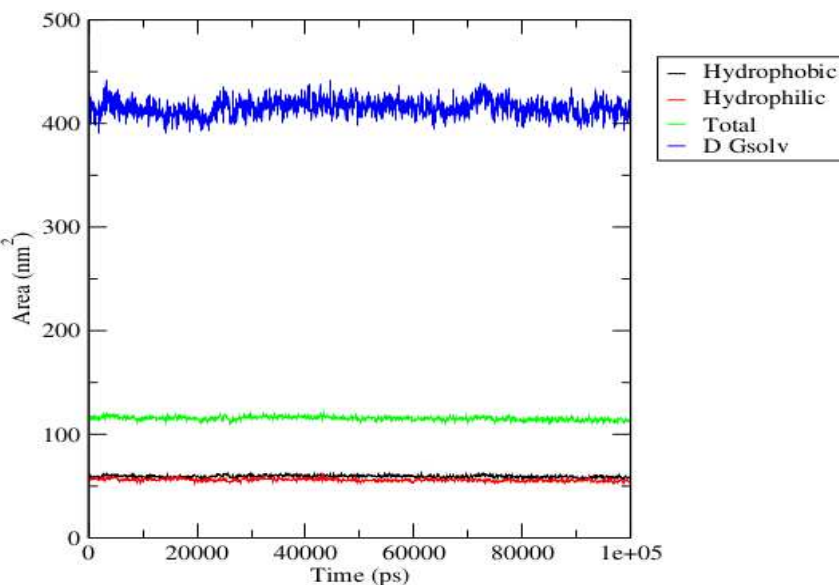


Fig 6.8: Solvent accesible surface areas and free energy of solvation (D Gsolv)

motion of the complex into a set of eigenvectors with corresponding eigenvalues after matrix diagonalization. Then first two or following eigenvectors can be plotted for information on Gibbs free energy, entropy landscape etc.

It uses the covariance matrix  $C$  of the atomic coordinates:

$$C_{ij} = \langle \sqrt{M_{ii}(x_i - \langle x_i \rangle)} M_{jj}(x_j - \langle x_j \rangle) \rangle$$

where  $M$  is a diagonal matrix containing masses of atoms in case of mass-weighted analysis or unit matrix in case of non-mass weighted analysis.  $C$  is a symmetric  $3N \times 3N$  matrix which can be diagonalized by an orthonormal transformation matrix  $R$ :

$$R^T C R = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_{3N}) \quad \text{where}$$

$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_{3N}$  the columns of  $R$  are eigenvectors and are called principal or essential nodes. Now the trajectory can be projected on the eigenvectors to obtain principal component  $p_i(t)$

$$p(t) = R^T \sqrt{M}(x(t) - \langle x \rangle)$$

eigenvalue  $\lambda_i$  is the mean square fluctuation of the principle component  $i$ .

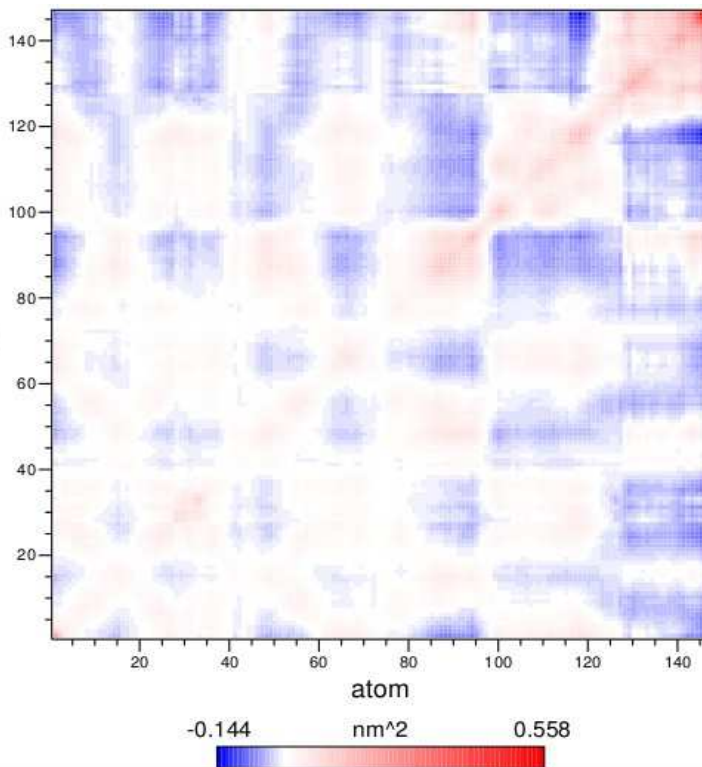
The collective motion of the system can be represented by the first few eigenvectors.

For example, the trajectory can be fitted to the eigenvector  $i$  by:

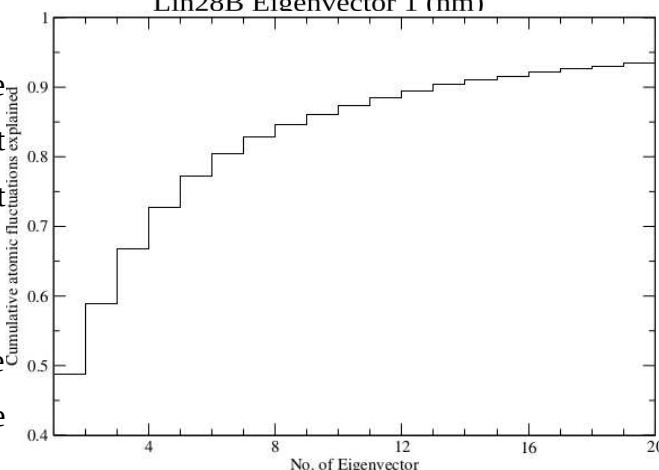
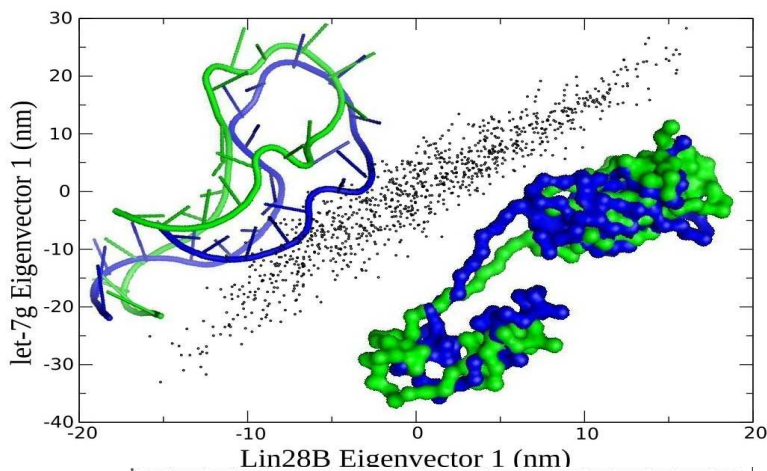
$$x^f(t) = \langle x \rangle + M^{-0.5} R_i p_i(t)$$

In order to get rid of rotational and translational motion of the molecule, least-square fit can be performed on the reference structure during the simulation. However care must be taken to choose the reference structure since it influences the covariance matrix.

In order to find out if the eigenvectors or the principle components actually don't represent noise



The Fig 6.9: Covariance matrix of atomic fluctuations



or random motion, cosine content has to be determined. It has been shown that the principal component of random diffusion are cosines with number of periods equal to the half of the principal component index [217]. The eigenvalues are proportional to the index to the power -2. The cosine content is defined as:

$\frac{2}{T} \left( \int_0^T \cos\left(\frac{i\pi t}{T}\right) p_i(t) dt \right)^2 \left( \int_0^T p_i^2(t) dt \right)^{-1}$  when the cosine content of the first few principal components is  $\sim 0$ , the largest fluctuation is actually random diffusion.

### 6.3.6 Change in radius of gyration ( $R_g$ )

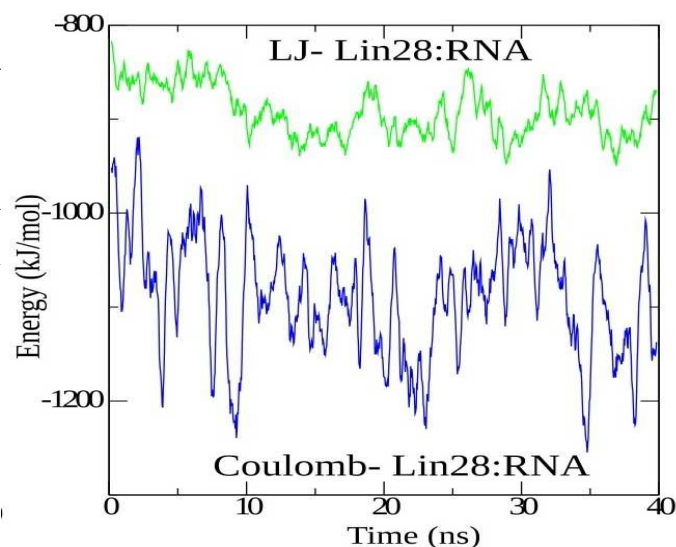
$R_g$  is used to check how compact or loose is the structure during the simulation.

$R_g = \sqrt{\frac{\sum_i \|r_i\|^2 m_i}{\sum_i m_i}}$  where  $m_i$  is the mass of the atom  $i$ , and  $r_i$  is the position of the atom  $i$  with

respect to the center of mass of the molecule.  $R_g$  is really useful in characterizing proteins (see PS1 CTF simulations.)

### 6.3.7 Interaction energies plotted over time.

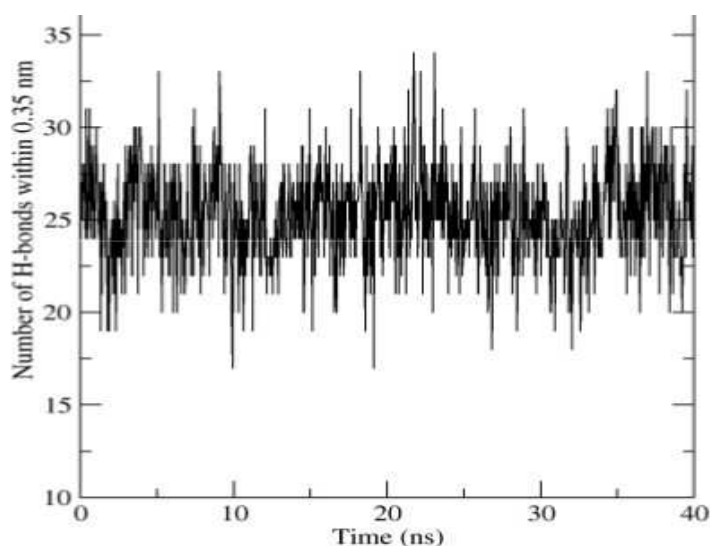
Lennard-Jones and Coulombic interaction energies can be plotted over time between a pair of atoms throughout the simulation.



### 6.3.8 Hydrogen bonding pattern

Hydrogen bonds can be analyzed between all possible donors and acceptors using the geometric criteria of maximum distance of 0.35 nm and angle of  $30^\circ$ . Using the average of autocorrelation function  $C(\tau) = \langle s_i(t) s_i(t + \tau) \rangle$  of the existence function (either 0 or 1) of all H bond at time  $t$ , the lifetime of hydrogen bond can be calculated by

$$\tau_{h-bond} = \int_0^\infty C(\tau) d\tau$$



### 6.3.9 k-means or hierarchical clustering of trajectories

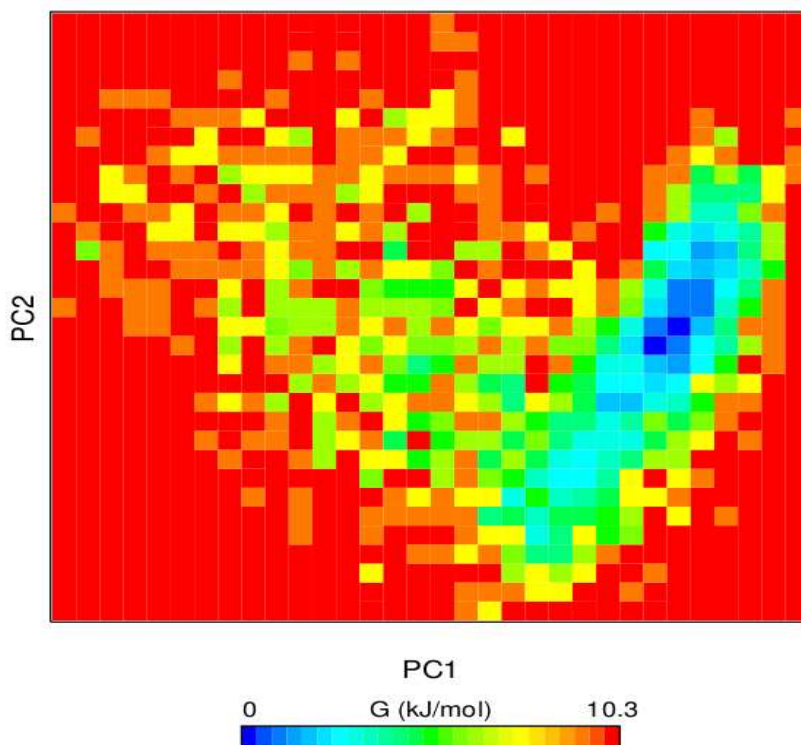
#### 6.3.10 Free energy

Biomolecular processes such as folding or aggregation can be described as molecule's free energy:  $\Delta G(R) = -k_B T [\ln P(R) - \ln P_{max}]$  where  $k_B$  is the Boltzmann constant,  $P$  is the probability distribution of the molecular system along coordinate  $R$  in which  $P_{max}$  gives the maximum which is subtracted to make sure that  $\Delta G=0$  for the lowest free energy minimum.  $R$  (order parameters) can be defined by various order parameters like  $R_g$ , RMSD, no. of h-bonds etc. The reduced free energy surface can be obtained by plotting two such order parameters.

The Gibbs free energy can be obtained from the potential of mean force or thermodynamic integration or free energy perturbation. In order to obtain absolute free energy, a reference state is required whose absolute free energy is known (state B) i.e. in ideal gas. This can be achieved by using a parameter  $\lambda$  to simulate the transition between small molecules (state A with  $\lambda=0$ ) and non-interacting particles in harmonic wells (state B with  $\lambda=1$ ) whose absolute free energy is known. The free energy difference can be expressed

$$\Delta G = \int_0^1 \left\langle \frac{\partial H(R, \lambda)}{\partial \lambda} \right\rangle_\lambda d\lambda$$

where the total Hamiltonian of the system at any particular value of  $\lambda$  is  $H = (1 - \lambda) H_B + \lambda H_A$



as: Fig 6.10: Free energy surface projected along principle component (PC) 1 and 2

A myriad of other analysis can be conducted depending on the research problem.

## 6.4 Modeling Membrane Proteins

### 6.4.1 Membrane proteins

Before starting the discussion of membrane protein modeling, a brief overview of membrane protein will be projected. Membrane proteins are classified into integral, peripheral and lipid-anchored proteins. Integral membrane proteins stay permanently attached to the lipid bilayer of plasma membrane (i.e. GS, GPCRs). Peripheral membrane proteins (e.g. phospholipase A2 or C, ) are temporarily attached to the lipid bilayer by non-covalent interactions. Lipid anchored proteins are covalently attached to the fatty acids like myristate/palmitate or glycosylphosphatidylinositol (GPI) which is a glycolipid that attaches to the C-termini during posttranslational modifications [218] (e.g. G proteins, kinases).

Integral membrane proteins can further be classified into 3 groups depending on their localization:

- type I single-pass transmembrane domain with cytoplasmic C-termini (like amyloid precursor protein, CD44, cadherins, Notch and other substrates of GS )
- type II single-pass transmembrane with extracellular C-termini
- multipass transmembrane (e.g. transporters: ABC transporters, ATPases, enzymes (GS, MraY), receptors (GPCRs), channels: porins, Na<sup>+</sup>/K<sup>+</sup> channels).

In order to adapt to the hydrophobic nature of the lipid bilayer, the integral membrane proteins have characteristic ~ 11-25 residues long patches of hydrophobic amino acids in the transmembrane domain. It is thermodynamically favorable as it minimizes the insertion and interaction energy between the protein and plasma membrane. Structurally the multipass transmembrane proteins can be subdivided into two categories- transmembrane helical and beta barrels. To collect the information of membrane proteins, various databases have been created. For instance, the orientation of proteins in membrane (OPM) database provides the calculated location of the membrane proteins with crystal structures in the hydrophobic core of the lipid bilayer. As of Nov, 2012, it consists of 660 transmembrane proteins and 1086 peripheral proteins (<http://opm.phar.umich.edu/> [219]) It is based on minimization of free energy for transferring the protein from water to the lipid bilayer. Apart from that, several databases on GPCRs, have been developed (<http://www.gpcr.org/7tm>)



## 6.5 Protein-Protein and Protein-ligand Docking

# 7 Methods

## 7.1 Structure building and simulations of GS components

N-terminal Fragment of PS 1 (NTF), Aph-1, Nicastrin and PEN-2 were modeled using Rosetta Membrane Ab-initio method [220–222] due to lack of homologies. The long (82 amino acid) N-termini of PS 1 (NTF) was modeled separately using Rosetta Ab-initio tool. Prediction of membrane topology was obtained from OCTOPUS server which uses hidden Markov models and artificial neural networks [223].

Threading in HHpred was used to find out distant homologies and models were built using Modeller with/ without secondary structure constrains [224,225]. Also models were generated from automated server I-TASSER [226,227]. Then all models were compared, clustered in various topologies and scored using Prosa 2003, ProQ [228,229], Rosetta Scores [230]. 4-8 good scoring models from various scoring programs were selected for MD simulation in implicit and explicit membrane environment using IMM1[215] in CHARMM [231] and G53A6 forcefield [232]in GROMACS [233,234] respectively.

### 7.1.1 Modeling of Pen-2

Models of Pen-2 were constructed using the following 3 methods:

#### Rosetta ab-initio membrane

The following options were used to generate 1000 models by rosetta ab-initio membrane program [220,221] available at:

[http://www.rosettacommons.org/manuals/archive/rosetta3.1\\_user\\_guide/app\\_membrane\\_abinitio.htm](http://www.rosettacommons.org/manuals/archive/rosetta3.1_user_guide/app_membrane_abinitio.html)

l

3 and 9 amino acids long fragment databases were obtained from Robetta server [230]. Location of helices found by octopus server is used here: 19-39 and 57-77.

Options:

-in:file:fasta ../pen2.fasta

-in:file:frag3 ../aat000\_03\_05.200\_v1\_3

-in:file:frag9 ../aat000\_09\_05.200\_v1\_3

```
-in:file:spanfile pen2.span
-in:file:lipofile pen2.lips4
-abinitio:membrane
-membrane:no_interpolate_Mpair
-membrane:Meny_penalties
-out:nstruct 1000
-database /opt/gromacs/rosetta/rosetta_database/
```

### HHSearch and Modeller

The following sequence alignment obtained from HHSearch [225] for distant homologs was feed into Modeller [235] to build homology models.

>P1;pen2

sequence:pen2: 1: : 101: :: : 0.00: 0.00

```
MNLERSVN--EKLNLCKRYLGGFAFLPFLWLVNIFWFFREAFVLPAYTEQSQIKGYVWRS--
-AVGFLFWVIVLTSWITIFQIYRPRWGALGDYLSFTIPLGTP*
```

>P1;1qzmA

structureX:1qzm.pdb: :A: :A:ATP-dependent protease LA; oligomerizationdomain, AAA+ protein, hydrolase; 1.90A {Escherichia coli} SCOP c.37.1.20:Escherichia coli:1.90:0.18

```
-----TE--DEKLNIAKRHLL-----
-----*
```

>P1;1zrjA

structureX:1zrj.pdb: :A: :A:a.140.2.1 (A1-37) Heterogeneous nuclear ribonucleoprotein U-like protein 1 {Human (Homo sapiens) [TaxId9606]}:Human (Homo sapiens) [TaxId 9606]:-1.00:-1.00

```
MDVRRLLKV--NELREELQR-----
-----*
```

>P1;3emlA

structureX:3eml.pdb: :A: :A:Human adenosine A2A receptor/T4 lysozyme chimera; caffeine, GPCR, membrane protein, LCP, mesophase, structural genomics, PSI-2; HET ZMA STE; 2.60A {Homo sapiens}:Homo sapiens:2.60:0.20

```
-----VSLAAADIAVGVLAIPFAITISTGFCAACHGC-----LFIACFVLV---LTQSSIFSLLAIA-----*
```

>P1;1ppjD

structureX:1ppj.pdb: :D: :D:f.23.11.1 (D196-241) Cytochrome c1 subunit of cytochrome bc1 complex (Ubiquinol-cytochrome c reductase), transmembrane anchor {Cow (Bos taurus) [TaxId

9913]}:Cow (Bos taurus) [TaxId 9913]:2.10:0.33

-----RK---RMGLKMLLMMGLLLPLVYAMKRHKW  
SVL-----\*

>P1;1di1A

structureX:1di1.pdb: :A: :A:Aristolochene synthase; sesquiterpene cyclase, isoprenoid biosynthesis, lyase; 2.50A {Penicillium roqueforti} SCOP a.128.1.4 PDB 1dgp\_A:Penicillium roqueforti:2.50:0.25  
AEESKLG--PATKRVLWSMTRE-----WETVHDEIVAEEKIASPDGCSEAAKAYMKGLEQMS  
GNEQWSK-----\*

>P1;2fwkA

structureX:2fwk.pdb: :A: :A:b.38.1.1 (A24-115) U6 snRNA-associated sm-like protein LSM5  
{Cryptosporidium parvum [TaxId 5807]}:Cryptosporidium parvum [TaxId 5807]:2.14:0.31

-----GNNVAML  
VPGGDP\*

### I-Tasser Server

3 models of Pen-2 were obtained from the fully automated I-Tasser server [227].

The models obtained from Rosetta, distant homology modeling, I-Tasser etc look pretty similar with the only difference in the location of transmembrane helices and orientation of N-terminal and C-terminal loops. These models are classified into two topologies with customized script which take into account the span of the TM helices. Topology 1 gives TM region: 21-40 and 60-80 which is similar to the predictions of TM region made by servers (HMMTOP, TM-HMM, Octopus etc). However topology 2 gives TM region: 17-41 and 54-80 which is different from servers. Final models of Pen-2 were chosen by rosetta energy, proQ and prosa 2003 scores, statistics for rosetta results and some database search performed [236] namely: inter helical distance (7-10 Angs.); crossing inter helical angle (from -130 to -160 and from 148 to 168); type of contacting residues at the helical interface (small amino acids, rather polar: Gly, Ser, Leu, Ile, Ala) and by visual inspection. We chose 3 models for topology 1 and 5 for topology 3. The models are sorted by rosetta energy. The remaining scores which could be useful for membrane proteins: pair potential and lennard-jones attraction potential. The N and C-termini were a bit unfolded in some of the models. Therefore MD Simulations were performed.

All the models were subjected to both 20 ns implicit and 80 explicit solvent simulation in IMM1 and g53a6 forcefield respectively. The disadvantage of explicit solvent is slower convergence to final structure whereas in implicit solvent it is achieved quite fast in absence of friction since the effect of solvent is represented by free energy of solvation. However, the atomistic interactions between the

solvent the protein can not be probed. It is possible in explicit solvent.

### **MD Simulation:**

8 best models were optimized by molecular dynamics simulations in all atom POPC bilayer in Gromacs (Gromos96 53a6 force-field) (van der Spoel 2005) and implicit membrane model 1 (IMM1) (Lazaridis 2003) in CHARMM. Formation of lipid bilayer from random orientation of lipid and Pen-2 was simulated in coarse-grained forcefield Martini (implemented in Gromacs).

The following procedure was followed for simulation in POPC bilayer: 0. Starting with minimized Pen2 in g53a6 FF with frozen backbone, 1. Inflategro [237] was used for inserting protein into membrane with strong position restraints so that RMSD change is close to zero. 2. adding water (Single Point Charge (SPC) model [238]) , ions to neutralize the system with gromacs tools (increasing the vdwradii of C to 0.4 so that water is not inserted inside membrane) 3. l-bfgs and steepest descent minimization for geometry optimization with g53a6 FF and modified Berger parameters for POPC from Prof. Tielmann's website ([http://moose.bio.ucalgary.ca/index.php?page=Structures\\_and\\_Topologies](http://moose.bio.ucalgary.ca/index.php?page=Structures_and_Topologies) ). 5. 100 ps MD with frozen backbone atoms of the protein, and with no pressure coupling (isochoric-isothermal (NVT) ensemble with temperature controlled by the Berendsen weak coupling algorithm [239]). After the simulation, checking the system to ensure no water molecules are present in membrane part of the system. If it's necessary, remove them 6. once again, l-bfgs + steep for geometry optimization. Following annealing, 1 ns of NPT equilibration was performed using the Nosé–Hoover thermostat [240,241] and Parrinello–Rahman barostat [242,243]. Secondary structure were assigned using DSSP algorithm [216]. From this point, all the MD simulations are conducted with pressure coupling (npt), 7. 2 ns MD with position restraints only on transmembrane helices. 8. 80 ns production run

### **7.2 Modeling of Bilyer Formation with Pen-2**

Pen-2 models were converted to coarse grain representation. Secondary structure constraints were derived from DSSP and topology was generated with elastic restraints. Pen-2 was centered in a  $6.4 \times 6.4 \times 10 \text{ nm}^3$  box and ions were added to make the system neutral. The system was minimized with steepest descent algorithm, 128 POPE molecules were added and the system was minimized again, 2000 water molecules were added the system was minimized again. Then 50 ns of MD simulation in NVT ensemble followed by another 50 ns NPT was carried out during which bilyer was formed in few simulations.

## 7.3 MD Simulations of PS1-CTF

### 7.3.1 Coarse-grain (CG) simulations

CG simulations were conducted using Gromacs software [233,234] and MARTINI force field [205,206]. Two types of phospholipids were used DPPC (dipalmitoylphosphatidylcholine) containing 4 hydrophobic grains in each chain and DLPC (dilauroylphosphatidylcholine) containing 3 hydrophobic grains. The starting systems contained NMR structure of protein, randomly distributed phospho lipids (ca. 460 molecules) and water (ca. 10000 grains). The necessary number of counterions (sodium cations) were added to ensure the electric neutrality of the system. Prior to the simulation the system was minimized, shortly equilibrated (about 1 ns) and minimized again. During the equilibration the protein structure was frozen. The productive simulations were run at a constant temperature 300 K and pressure 1 bar. The initial dimensions of the periodic box were 12 nm x 12 nm x 10 nm. The isotropic pressure coupling was used with Berendsen coupling constant 0.2 ps. Timestep 20 fs was applied. Ten simulations with DPPC and ten with DLPC phospholipids were run for 1  $\mu$ s each. Selected simulations where proper bilayer was formed were extended for another 1  $\mu$ s. For micelle generation one CG simulation with 200 DPC (dodecylphosphocholine) detergent molecules was conducted for 1  $\mu$ s with initial random distribution of detergent.

### 7.3.2 Implicit Membrane Simulations

Replica Exchange and Molecular Dynamics simulations in continuous environments were performed using the program CHARMM [231](Brooks 1983). A united-atom forcefield, CHARMM19 [244] was used with a method IMM1 (Implicit Membrane Model) [215]. Two different values of hydrophobic core thickness were chosen 2.6 nm and 3.0 nm. Twenty Replica Exchange (REx) simulations were conducted starting from ensemble of twenty NMR structures. Each Replica Exchange simulation was running in eight temperature windows with a temperature range 300 K and the lowest window had temperature 300 K. Selected structures were simulated using Molecular Dynamics method at a constant temperature 300 K and pressure 1 bar. In both methods the temperature was controlled with Langevine thermostat (friction coefficient was set to 5 ps<sup>-1</sup>). The bonds involving hydrogen atoms were constrained using the SHAKE algorithm, allowing an integration time step of 2 fs.

## 7.4 Simulations in Amber FF

Prior to MD simulations, the complex was centered in an 8 nm<sup>3</sup> cubic periodic box, the system was solvated using TIP3P water model [245]. Counter ions were added to neutralize the system (150 mM NaCl). The system was minimized in Gromacs 4.5.5 using the Amber99sb Force Field [246] and a steepest descent algorithm [247] followed by L-BFGS [248,249] with harmonic restraints (force

constant:  $1000 \text{ kJmol}^{-1}\text{K}^{-1}$ ) applied to the backbone atoms of the protein. The same restraints were applied in the following 100 ps NVT and 100 ps NPT equilibration. The final MD runs were carried out for 80 ns with the NPT ensemble without any restraints.

During all simulations a 2 fs integration time step was used, the temperature was maintained at 300 K with the V-rescale algorithm and the pressure was coupled at 1 bar by isotropic pressure coupling utilizing the Parrinello-Rahman algorithm (time constant 10 ps, isothermal compressibility of water:  $4.5 \times 10^{-5} \text{ bar}^{-1}$ ). Long-range electrostatics was calculated by fourth order particle mesh Ewald (PME) [250,251] algorithm with a grid spacing of 0.16 nm. All bonds were constrained using a fourth order P-LINCS algorithm [202,203]. Electrostatic and van der Waal's interactions were cut-off at 1 nm and a dispersion correction was applied to account for it. Periodic boundary conditions were implemented in three dimensions. Initial atomic velocities, prior to NVT equilibration, were obtained from Maxwell's distribution at 300 K. Water molecules were constrained by SETTLE [252]. Neighbor lists for nonbonded interactions were updated at every 5th step. Images were prepared using PyMOL [253] and Visual Molecular Dynamics (VMD) program [254].

## 7.5 Modeling of GPCRs (CB<sub>1</sub>, CB<sub>2</sub> and $\beta_1$ AR)

Molecular models of human cannabinoid receptors CB<sub>1</sub> and CB<sub>2</sub> were prepared by a combined strategy joining Modeller [235], Rosetta-*ab-initio* [230] and Rosetta-loop-modeling [255]. In the first stage, we performed multiple sequence alignments of sequence derived from cannabinoid receptors and available from the GPCR template structures: bovine rhodopsin, human  $\beta_2$ -adrenergic receptor, turkey  $\beta_1$ -adrenergic, human A<sub>2A</sub> adenosine receptor (protein codes from Protein Data Bank: 1U19, 2RH1, 2VT4, 3EML, respectively). Multiple sequence alignments were prepared by two commonly used tools: MUSCLE [256] and CLUSTALW (version 2.0) [257]. Both methods gave similar results. The highest pairwise sequence score according to CLUSTALW (based on normalized identity) was achieved for: CB<sub>1</sub>-A<sub>2A</sub>R pair (24%), CB<sub>2</sub>- $\beta_1$ AR (22%) and also for CB<sub>1</sub>- $\beta_1$ AR and CB<sub>2</sub>-A<sub>2A</sub>R (21%) pairs. Pairwise sequence score with rhodopsin was very low, 15% (CB<sub>2</sub>) and 13% (CB<sub>1</sub>), the same with  $\beta_2$ AR: 18% (CB<sub>2</sub>) and 14% (CB<sub>1</sub>). Consequently, we decided to chose these two receptors structures: A<sub>2A</sub>R and  $\beta_1$ AR, both bound to antagonists in their crystal structures, as templates for homology modeling of CB<sub>1</sub> and CB<sub>2</sub>. Alignment input for Modeller was prepared automatically by MUSCLE and adjusted manually in the Seaview editor [258] to preserve important functional motifs and disulfide bridges. Both CB<sub>1</sub> and CB<sub>2</sub> receptors have single disulfide bridge located in the loop EC2 (both cysteins are within this loop: 257-264 in CB<sub>1</sub> and 174-179 in CB<sub>2</sub>) [259] so the sequences of cannabinoid receptors were manually aligned with templates A<sub>2A</sub>R and  $\beta_1$ AR within the EC2 area to put cysteine residues in proximity. The final alignments for both

templates are shown in Fig. 7.1. For transmembrane parts of modeled receptors we chose the DOPE (Discrete Optimized Protein Energy) option [260] in Modeller with subsequent MD-slow refinement of short loops. The mean DOPE score for 10 generated models was: -43281.0 (CB<sub>1</sub> based on A<sub>2A</sub>), -42440.2 (CB<sub>1</sub> based on  $\beta_1$ AR), -38421.2 (CB<sub>2</sub> based on A<sub>2A</sub>) and -37946.3 (CB<sub>2</sub> based on  $\beta_1$ AR). These scores reflect primarily the interactions within the bundle of transmembrane helices which are responsible for stabilization of the receptor. As the models based on A<sub>2A</sub>R were scored much better, we chose them for further docking studies. Subsequently, the Rosetta-loop-modeling protocol was used for remodeling the longest loop (IC3 – between helices TH5 and TM6): residues 300-335 in CB<sub>1</sub> and residues 217-234 in CB<sub>2</sub>. Additionally, the N- and C-terminal parts of receptors, which are located outside the membrane, were modeled separately by the Rosetta-*ab-initio* protocol and joined with the rest of the protein by Modeller. The two final models of each receptors characterized with the highest scores according to the DOPE measure were subjected to further analysis. We used these two conformations per each CB receptor to assess whether the choice of the model from the Modeller generated ensemble influences the docking results to any extent.

To refine the CB1 and CB2 structures obtained from homology modeling we used Implicit Membrane Model (IMM1) method [215] in the program CHARMM [231](2). It gives a possibility for fast convergence to final structure. The implicit solvent or continuous environment (CE) method reduces the number of degrees of freedom which are necessary for the evaluation of energy and force allowing more efficient sampling of phase space compared to explicit solvent methods. The water/membrane/water system in IMM1 is achieved by changing the parameters of the system along the perpendicular axis of the membrane. In the layer 0.6 nm thick between water and membrane environments the properties are changing smoothly according to a sigmoidal function so at the hydrophobic border of this layer 90% of the environment derives from the hydrophobic core while at the water border 90% of the environment derives from bulk water properties. The value of hydrophobic core thickness chosen for molecular dynamics simulation was 3.2 nm. The CB1 and CB2 structures were simulated for 1ns with an integration time step of 2 fs at a constant temperature of 300 K and pressure of 1 bar. Constant temperature was achieved with Langevine thermostat by setting friction coefficient to 5 ps<sup>-1</sup>. The bonds involving hydrogen atoms were constrained using the SHAKE algorithm.

We conducted five simulations for CB1 and six for CB2 each time using different structures being the best scored receptor models based on the A<sub>2A</sub>R template from Modeller. The RMSD on C-alpha atoms was 0.27 nm for the transmembrane part of CB1 and 0.18 nm for CB2 on average. The representative plots of  $\chi_1$  angle for residues in the rotamer toggle switch were found to be stable in most cases. In some cases, the changes reflecting the spontaneous action of the switch were observed. This indicates that the models of CB1 and CB2 receptors we used were functional at least in the area of the binding site which is enough for ligand docking.

For the CB<sub>1</sub> and CB<sub>2</sub> receptor models, which were based on A<sub>2A</sub>R template, we performed further studies, involving flexible docking of two antagonists (AM630 and NESS-0327) and two agonists (anandamide (AEA) and (-)- $\Delta^9$ -THC). Input conformations of ligands were prepared using the LigPrep protocol from the Schrodinger Suite [261]. To sample different protonation states of ligands in physiological pH we used the Epik module [262]. From our set of ligands of CB<sub>1</sub> and CB<sub>2</sub> receptors, only AM630 was used in the protonated state (protonated nitrogen atom in morpholine ring) based on pK<sub>a</sub> calculations (6.2±0.6). However, literature data [263] provide unequivocal evidence on this protonation, therefore we decided to dock both forms of AM630. The obtained poses were similar, however, the protonated AM630 poses were characterized with higher energy values. Docking of all ligands were performed by Autodock 4.2 [264] using the genetic algorithm (GA) procedure. The following parameters for GA were used: 1.9 nm-large (50 grid points) search box, 150 individuals (poses - ligand/receptor conformations) in each population, 20 independent populations per each analyzed system (ligand-receptor complex), 2,500,000 energy evaluations per single evolution run, post-docking cluster analysis and other default settings. Some amino acids were set flexible during the docking. For this we chose amino acids with bulky residues close to the potential binding site (based on literature data): L3.29<sup>(193)</sup>, V3.32<sup>(196)</sup>, F3.36<sup>(200)</sup>, F268 (EC2 loop), E5.37<sup>(273)</sup>, F5.42<sup>(278)</sup>, T5.47<sup>(283)</sup>, W6.48<sup>(356)</sup>, L6.51<sup>(359)</sup>, L6.52<sup>(360)</sup>, C7.42<sup>(386)</sup> for CB<sub>1</sub>, and T3.33<sup>(114)</sup>, F3.36<sup>(117)</sup>, F183 (EC2 loop), D5.38<sup>(189)</sup>, S5.42<sup>(193)</sup>, I5.47<sup>(198)</sup>, F5.51<sup>(202)</sup>, W6.48<sup>(258)</sup>, V6.51<sup>(261)</sup>, L6.52<sup>(262)</sup>, M6.55<sup>(265)</sup> for CB<sub>2</sub> receptor. Numbers of residues according to Ballesteros-Weinstein numbering scheme [265] in which x.50 denotes the most conserved residue in each helix).

Building of the complete structure of  $\beta_2$ AR was done on the basis of the crystal structure of human  $\beta_2$ ART4 lysozyme fusion protein with bounded carazolol (PDB ID: 2RH1) [266]. Modeling of N- and C-terminal domains of the receptor (residues Met:1 to Glu:30 and Cys:341 to Leu:413 respectively) was conducted using ITASSER server [226,227]. The structure of the longest second intercellular loop of the receptor (residues Leu:230 to Leu:266) was predicted using CABS program [267]. Single palmitoyl chain was added to Cys:341 at the end of the cytoplasmic helix H8. Obtained  $\beta_2$ AR model was inserted into equilibrated palmitoyl-oleoyl-phosphatidylcholine (POPC) cell membrane model by means of Inflategro procedure [237]. Model of  $\beta_2$ AR embedded in POPC lipid bilayer was then solvated with water molecules and ions were added. Energy minimization was conducted applying 2000 steps of the steepest decent algorithm followed by 2000 steps of the L-BFGS algorithm. Then, the molecular dynamic (MD) simulation lasting 40 ns was performed using GROMACS (v. 3.3) program [233]. All calculations were conducted using modified GROMOS96 force field (ffG53a6 parameters set) [232] with additional parameters for POPC molecules [268]. SPC water model [269] was used and the PME procedure [250] was applied for treatment of the long-range electrostatic interactions. All bonds with hydrogen atoms were constrained by the LINCS algorithm [202]. MD was performed at the temperature of 310 K, pressure



of 1013 hPa, and simulation time step was set 1 fs.

To obtain (*R,R*)- and (*S,S*)-fenoterol isomer structures and force field parameters for MD simulation the PRODRG server [270] [41] was used. The ligands were inserted in the middle of the binding site of the  $\beta_2$ AR model to preserve the interaction between D3.32 and the protonated amine nitrogen of ligands. To investigate differences in binding of (*R,R*)- and (*S,S*)-fenoterol isomers similar starting structure of two receptor-ligand complexes were generated during restrained MD simulation lasting 200 ps. Protein backbone atoms were constrained to their initial positions using “freeze” option and weak harmonic distance restraints (the distance was 0.3 nm) were imposed on three receptor-ligand atom pairs (pair 1: C<sub>1</sub> atom of D3.32 residue and protonated nitrogen atom of ligand; pair 2: oxygen atom of hydroxyl group of S5.42 and oxygen atom of first hydroxyl group of 1,3-benzenediol moiety; pair 3: oxygen atom of hydroxyl group of S5.46 and oxygen atom of second hydroxyl group of 1,3-benzenediol moiety). Finally, two step MD simulation of receptor-ligand complexes was performed. During the first step, lasting 2 ns, weak harmonic position restraints were imposed on backbone atoms of transmembrane helices of the receptor only and ligand-receptor distance restraints were released. In the second step the production run was conducted lasting 5 ns with no restraints. The described above two step MD simulation scheme was repeated 22 times applying random starting velocities for every atom, 11 times for receptor-(*S,S*)-fenoterol complex and also 11 times for receptor-(*R,R*)-fenoterol complex. Simulation parameters were identical to those used for MD simulation of unliganded  $\beta_2$ AR model.

a)

```

1
CB1 MKSILDGLAD TFRRTITDDL LYVGSNDIQY EDIKGDMASK LGYFPQKFPL TSFRGSPFQE KMTAGDNPQL
CB2 -----
A2AR -----

71
CB1 VPADQVNITE FYNKSLSSFK ENEENIQCGE NFMIDIECFMV LNPSQQLAIA VLSLTGLGFT VLENLLVLCV
CB2 ----- ---MEECWVT EIANGSKDGL DSNPMKDYMI LSGPQKTAVA VLCTLLGLLS ALENVAVLYL
A2AR ----- ---IMGSSVYI TVELAIAVLA ILGNVLCWA

141
CB1 ILHSRSLRCR PSYHFIGSLA VADLLGSVIF VYSFIDFHVF H-RKDSRNVF LFKLGGVTAS FTASVGSFLF
CB2 ILSSHQLRRK PSYLFIGSLA GADFLASVVF ACSFVNFHVF HGVD-SKAVF LLKIGSVTMT FTASVGSLLL
A2AR VWLNSNLQ-N VTNYFVVS LAADIAVGVLA IPFAITISTG F-C-AACHGC LFIACFVLVL TQSSIFSLLA

211
CB1 TAIDRYISIH RPLAYKRIVT RPKAVVAFCL MWTIAIVIAV LPLLGW---- ---CEKLS VCS-----
CB2 TAIDRYLCLR YPPSYKALLT RGRALVTLGI MWVLSALVSY LPLMGW---- ---CCPRP -CS---ELFP
A2AR IAIDRYIAIR IPLRYNGLVT GTRAKGIIAI CWVLSFAIGL TPMLGWNNCG QSQGCGEGQV ACLFED--VV

281
CB1 DIFPHIDETY LMFWIGVTSV LLLFIVYAYM YLLWKAHSHA VRMIQRGTQK SIIHTSEDG KVQVTRPDQA
CB2 ---LIPNDY LLSWLLFI AF LFSGIIYTYG HVLWKAHQHV ASLSGHQDRQ ----- ---VPGMARM
A2AR ---FMNYMV YFNFFACVLV PLLLLLGVYL RDLAAR--- ----- ---RQLRSTL

351
CB1 RMDIRLAKTL VLILVLIIC WGPLLAIMVY DVFG-KMNKL IKTVFAFCSM LCLLNSTVNP IYIALRSKDL
CB2 RLDVRLAKTL GLVLA VLLIC WFPVLALMAH SLAT-TLSDQ VKKAFAFCSM LCLINSMVNP VIYALRBEI
A2AR QKEVHAASKL AIIVGLFALC WLPLHIINCF TFFCPDCSHA ELWLMYLAIV LSHTNSVVNP FIYAYRIREF

421
CB1 RHAFRSMFPS CEGTAQPLDN SMGDSCLHK HANNAASVHR AAESCIKSTV KIAKV TMSVS TDTSAEAL
CB2 RSSAHHCLAH WKKCVRGLGS EAKEEAPRSS VTETEADGKI TPWPDSRDL DLSDC-----
A2AR RQTFRKIIRS HVLRQ-----
    
```

b)

```

1
CB1 MKSILDGLAD TFRRTITDDL LYVGSNDIQY EDIKGDMASK LGYFPQKFPL TSFRGSPFQE KMTAGDNPQL
CB2 -----
β1AR -----

71
CB1 VPADQVNITE FYNKSLSSFK ENEENIQCGE NFMIDIECFMV LNPSQQLAIA VLSLTGLGFT VLENLLVLCV
CB2 ----- ---MEECWVT EIANGSKDGL DSNPMKDYMI LSGPQKTAVA VLCTLLGLLS ALENVAVLYL
β1AR ----- ---QWEAGMS LLMALVLLI VAGNVLVIAA

141
CB1 ILHSRSLRCR PSYHFIGSLA VADLLGSVIF VYSFIDFHVF HRKD-SRNVF LFKLGGVTAS FTASVGSFLF
CB2 ILSSHQLRRK PSYLFIGSLA GADFLASVVF ACSFVNFHVF HGVD-SKAVF LLKIGSVTMT FTASVGSLLL
β1AR IGSTQRL-QT LTNLFITSLA CADLVVGLLV VFPGATLVVR GTWLWGSFLC ELWTSLDVLC VTASIETLCV

211
CB1 TAIDRYISIH RPLAYKRIVT RPKAVVAFCL MWTIAIVIAV LPLLGW---- ---NCEK LQSVCSDFIP
CB2 TAIDRYLCLR YPPSYKALLT RGRALVTLGI MWVLSALVSY LPLMGW---- ---TCP RP--CSELF
β1AR IAIDRYLAIT SPFRYQSLMT RARAKVIIC VVAISALVSF LPIMMHWRD EDPQALCYQ DPGCCDFV--

281
CB1 HDDETYLMFW IGVTSVLLLF IVYAYMYILW KAHSHAVRMI QRGTQKSIII HTSEDGKVQV TRPDQARM DI
CB2 LIPNDYLLSW LLFIAFLFSG IITYTYGHVW KAHQHVASLS GHQDRQ---- ---V PGMARMRLDV
β1AR -TNRAYAIAS SIISFYIPLL IMFVALRVY REAK----- ---E QIR-----EH

351
CB1 RLAKTLVLLI VVLIICWGPL LAIMVYDVFG KMNKLKTVF AFCSMLCLLN STVNP IYIAL RSKDLRHAFR
CB2 RLAKTLGLVL AVLLICWFPV LALMAHSLAT TLSDQVKKAF AFCSMLCLIN SMVNPVIYAL RSGEIRSSAH
β1AR KALKTLGIIM GVFTLCWLPF FLVNIVNVFN R-DLVEDWLF VAFNWLGYAN SAMNP IYICR -SPDFRKAFK

421
CB1 SMFPSCGTA QPLDNMGS DCLHKHANNA ASVHRAAESC IKSTVKIAKV TMSVSTDTSA EAL
CB2 HCLAHWKKCV RGLGSEAKEE APRSSVTETE ADGKITPWP DSRDL DLSDC-----
β1AR RLLAF-----
    
```

Fig 7.1 The alignment of human CB1 and CB2 receptor sequences with A2AR (a) and β1AR (b) templates. Transmembrane helices of templates are encircled (red dashed line), the conserved residues (x.50) in each helix are in blue and cysteine residues forming disulfide bridge are in yellow. Conserved sequence motifs in cannabinoid receptors are underlined.

## 7.6 Molecular modeling of FPR1

### 7.6.1 Homology modeling and refinement of FPR1

The homology models of FPR1 were obtained by Modeller 9v8 using the crystal structure of chemokine receptor type 4 (CXCR4, PDB code: 3OE0) which shares the highest homology (31.0% identity, 53.8% similarity) with FPR1 according to Discovery Studio Visualizer. They are located in the same the  $\gamma$  branch of phylogenetic tree of GPCRs (gpcr.scripps.edu). Since the region corresponding to helix H8 at cytoplasmic side of CXCR4 is unfolded in the crystal, the crystal structure of human  $\beta_2$ -adrenergic receptor (PDB Code: 2RH1) was used as the second template for the H8 regions of FPR1. The sequence alignments were performed automatically in MUSCLE and adjusted manually in Discovery Studio Visualizer for proper aligning of conserved motifs and disulfide bridge. The 1500 models of initial FPR1 receptor were generated in Modeller with fully annealed protocol, and the optimal model was chosen according to DOPE (Discrete Optimized Protein Energy) score. Low homology regions of loops between transmembrane helices were constructed with loop refinement protocol in Modeller and the lowest DOPE score model from 1000 generated models was selected for further study. To obtain the proper orientation of the receptor in the membrane the refined model of FPR1 was aligned with CXCR4 crystal structure (PDB code 3OE0) taken from OPM (Orientations of Proteins in Membranes) database. The hydrogen atoms were added to the FPR1 structure according to the physiology pH environment. To remove unfavorable steric contacts and to release strain among amino acid residues the model was submitted to Prime (Schrödinger 2011 suite) for backbone constrained truncated-Newton minimization refinement, using the OPLS\_2005 force field and implicit membrane model.

### 7.6.2 Receptor model equilibration in explicit membrane

Using the builder tool for Desmond in Maestro 9.2 program the FPR1 model was embedded into pre-equilibrated POPE (1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphoethanolamine) lipid bilayer solvated with water and NaCl to make the system neutral and set ionic strength 0.15 M. The total number of atoms was approximately 54,000 including 28 Na<sup>+</sup> and 40 Cl<sup>-</sup> ions, about 10,000 water molecules, and 161 POPE molecules. The periodic box dimensions were about 6.8 nm  $\times$  7.2 nm  $\times$  9.4 nm. Equilibration of the system was performed at constant pressure and temperature (NPT ensemble; 310 K, 1 bar) and Berendsen coupling scheme with one temperature group. All bond lengths to hydrogen atoms were constrained using M-SHAKE. Van der Waals and short-range electrostatic interactions were cut off at 1.0 nm. Long-range electrostatic interactions were computed by the particle mesh Ewald (PME) summation scheme. A RESPA (time-reversible reference system

propagator algorithm) integrator was used with a time step of 1.6 fs. Long-range electrostatic interactions were computed every 4.8 fs. Harmonic positional restraints on the protein were tapered off linearly from 10 to 0 kcal/mol<sup>-1</sup>Å<sup>-2</sup> over 16 ns.

### 7.6.3 Ligand preparation and docking

Both ligands fMLF and tBoc-MLF were built in Maestro program. Ligand preparation utility was used to optimize the geometry of initial structures. Systematic conformational search was performed in MacroModel and the top five conformers with the lowest potential energy were kept for docking. The docking procedure was performed using Glide (Schrödinger 2011 suite). Ligand molecules were initially placed in the binding pocket with a random pose. Cubic boxes centered on the ligand mass center with a radius 1.5 nm for both fMLF and tBocMLF defined the docking binding regions. Flexible ligand docking was executed in all cases. Twenty poses per ligand out of 2000 were included in post-docking energy minimization. Top three scored poses were similar to each other, thus only one the best scored pose per each ligand was chosen as the initial structure for MD simulations.

### 7.6.4 Molecular Dynamics

To obtain the non-standard residues (-CHO and tBoc-) the force field parameters for MD simulation, the partial atomic charges for the ligands were obtained in GAUSSIAN 09 program via obtained Hartree-Fock 6-31G\* electrostatic potential (ESP) and then using the fitting procedure performed by the R.E.D. tool . The membranous system was built and equilibrated as mentioned above. Three 40 ns MD simulations with no restraints were conducted employing CHARMM36 full-atom force field for Apo-FPR1 as well as its complexes with agonist fMLF and antagonist tBocMLF. Data analysis was done using Desmond utilities and the molecular figures were made in VMD and Pymol .

## 7.7 Homology Modeling of Human CXCR4 and Dopamine D3 Bound to Ligands

### 7.7.1 Homology Modeling

The CXCR4 and DRD3 protein sequences were aligned with: bovine rhodopsin, human  $\beta$ 2-adrenergic receptor, turkey  $\beta$ 1-adrenergic, human A2A adenosine receptor (protein codes from Protein Data Bank: 1U19, 2RH1, 2VT4, 3EML, respectively) by MUSCLE [256] and CLUSTALW2 [257] software. CLUSTALW2 scores were used to choose the most appropriate template (2VT4).

Additionally, MUSCLE-derived multiple sequence alignments were prepared from BLAST[271] hits for each target protein sequence. The protonation state of the ligands and placement of disulfide bridges were confirmed by literature search [272–275]

Automatic alignments were manually adjusted to remove gaps inside the TM helices and to preserve disulfide bridges detected experimentally. The homology modeling was performed using the DOPE modeling option in Modeller, followed by slow MD refinement of loops [235]. The best scoring models (according to the DOPE score [260]), three per each target, were subjected to further analysis. The ligands conformations were prepared by Ligprep [261] and Epik [262] protocols from the Schrodinger Suite. The docking was performed in two ways. First, using the Glide approach [276], and second, using the Autodock 4.2 [264] with Gasteiger charges assignment. The top scoring poses (according to the glide score and the Autodock free energy of binding) were refined in Glide and scored again. The five top scored complexes per each target were submitted. All of them were originally prepared in Glide docking procedure, not in Autodock.

### **7.7.2 Criteria for prediction analysis, scoring and ranking:**

The DOPE (Discrete Optimized Protein Energy) and the DOPE scoring profiles, Glide score, Autodock free energy of binding and visual inspection were used in the assessment.

### **7.7.3 PS1 CTF: APP simulations in implicit membrane**

The proteins are treated in atomic detail while the effect of water and membrane was represented by adding an extra term in energy function. Implicit membrane model (IMM1) energy function can be expressed as  $W_{imm1} = E + \Delta G_{sol}$  where  $E$  is the intramolecular energy from CHARMM19 polar hydrogen forcefield and  $\Delta G_{sol}$  is the solvation free energy which varies depending on the position of the atom relative to the membrane. Simulations were conducted in membranes containing 30% anionic lipid with area per lipid of  $0.7 \text{ nm}^2$ , salt concentration of 0.1 M, the valence of the electrolyte +1, and the position of the plane of smeared charge according to Gouy-Chapman theory relative to the nonpolar/polar interface was set to 0.3 nm.

## 8 Results and Discussion

### 8.1 MD Simulation of NMR Structure of PS1-CTF

PS1 belongs to intramembrane proteases (I-CLIPs) which process its substrates inside the lipid bilayer [6,9]. Many of the mutations linked to the Early-onset Alzheimer's Disease is linked to PS1. During activation PS1 is proteolytically cleaved into N-terminal (NTF) and C-terminal fragments (CTF) each containing one catalytic aspartate residue [21]. Although the topology of the NTF is well accepted, it is somewhat controversial for CTF although the 3 transmembrane topology is mostly accepted [52,277,278]. Here we have validated the first structure of CTF obtained from NMR studies [53] in SDS micelles by performing molecular dynamics (MD) simulations in detergent micelles and lipid bilayer.

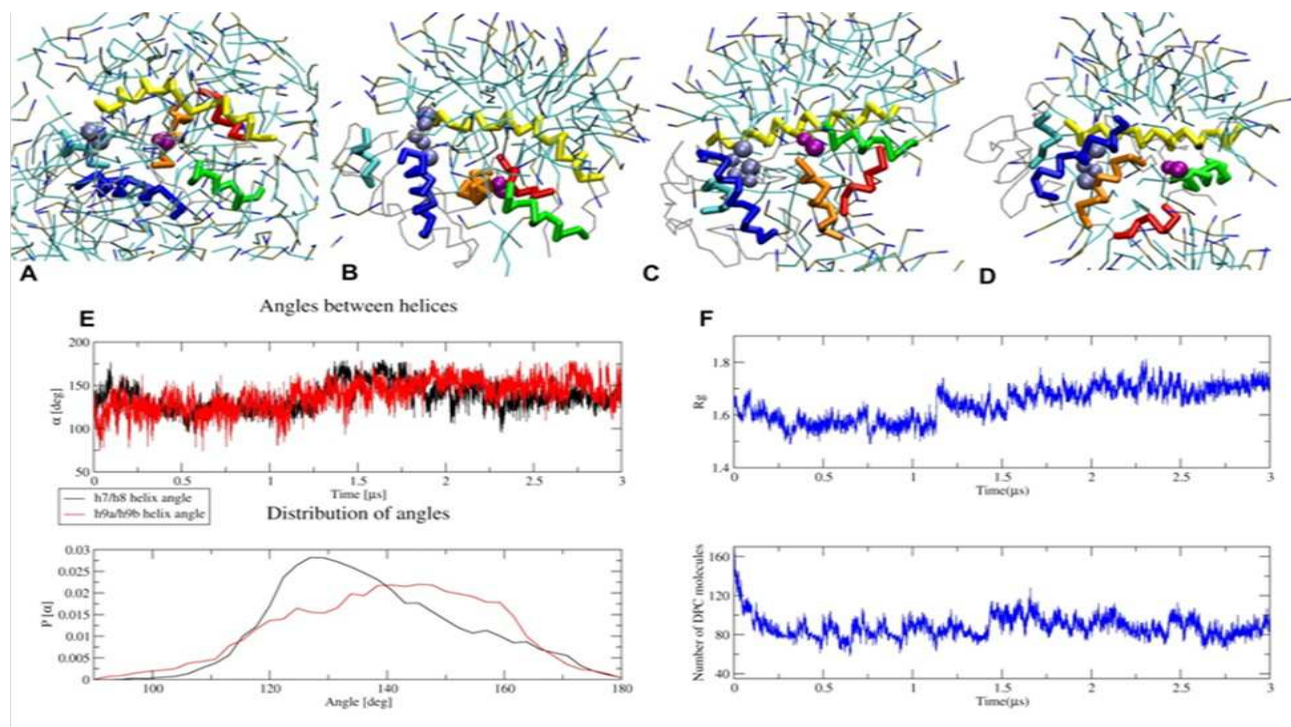
#### 8.1.1 PS1-CTF MD Simulations in Detergent

The structure of CTF in SDS micelles determined by NMR Spectroscopy was characterized by lack of typical transmembrane part. To investigate further the 3D structure of CTF in lipid bilayer, we used molecular modelling methods, and especially simulations in water/lipid environments. To clarify which regions of CTF structure divide into hydrophobic and hydrophilic parts, we conducted MD simulation of CTF immersed in random mixture of detergent and water molecules. To achieve a reliable simulation time needed for micelle formation, and extensive sampling of conformational space we chose a coarse-grain approach.

##### Simulations in DPC micelles

For micelle simulations, we used 2 00 DPC (dodecylphosphocholine) coarse-grain molecules in periodic box 12 nm x 12 nm x 10 nm filled with water grains. The coarse-grain CTF was placed centrally, and it was frozen for the first part of simulation lasting 50 ns. Several micelles were created and the central one formed around the protein contained about 60 molecules of detergent. Its representative structure is shown in Fig 8.1. It is striking that detergent molecules are grouped mainly around helix 8 and other regions of proteins are in contact with water. For the following 1  $\mu$ s of simulation the CTF was unfrozen and adopted a more compact structure mainly by rearrangement of its N-terminal part and changing the position of helices 9a and 9b. During such process, a number of detergent molecules in this micelle diminished to about 50. They were still grouped around helix 8 but also h7. A representative structure of this micelle is shown in Fig. 8.1D. Placement of helical elements is similar in both structures although a range of variability of angles between helices is high what can be seen in Fig.8.1E. By the end of 1  $\mu$ s simulation with movable CTF, the angles between helices h7/h8 and h9a/h9b were both close to 125°. They are similar to analogous angles in ensemble of NMR structures where they are 120°  $\pm$ 3° for h7/h8 pair and 108°  $\pm$ 6° for h9a/h9b pair. During

coarse-grain simulations, all helical segments were restrained what is an internal feature of this method because otherwise the secondary structure elements have tendency to unfold. However, a number and position of helical segments is determined by NMR experiments so employing of such restrains is justified.



**Fig 8.1 MD Simulation in DPC detergent:** A-D. Evolution of randomized detergent molecules and initially frozen CTF structure in course of time: A (initial), B (50 ns), C (1  $\mu$ s), D (3  $\mu$ s). E. Evolution of the angles between helices 7 and 8 (black) and helices 9a and 9b (red) and their distributions. F. Radius of gyration of the protein (Upper) and number of detergent molecules close to the protein (Lower). Incorporation of more DPC molecules from adjacent micelles at about 1.5  $\mu$ s results in increasing angles between helices (they become more anti-parallel) and an increasing radius of gyration (protein becomes more elongated). Coloring scheme: helix  $\alpha$  in cyan, helix  $\beta$  in blue, helix 7 in green, helix 8 in yellow, helix 9a in orange, and helix 9b in red. The catalytic residue D385 is shown as purple spheres and the PAL (proline, alanine, leucine) motif [36,66] as blue spheres. Structures of CTF – they differed mainly in shape and location of N-terminal part. Starting from random distribution of lipids was necessary because lack of explicit transmembrane segment of CTF precluded placement of CTF into already equilibrated bilayer. Selected simulations in which a proper bilayer was formed were extended for another 1 or 2  $\mu$ s when necessary to achieve stability.

### 8.1.2 MD Simulations in Lipid Bilayers

We also investigated a dynamics of CTF in membrane bilayers. Two types of membrane were used differing with their thickness: one was composed of DLPC phospholipids (dilauroylphosphatidylcholine) containing 3 hydrophobic grains in each chain and the second membrane was composed of DPPC phospholipids (dipalmitoylphosphatidylcholine) containing 4 hydrophobic grains. The distance between phosphorous grains (and also between nitrogen grains) from both layers of the membrane was 3.6 nm for DLPC and 4.0 nm for DPPC. The starting systems contained NMR structure of protein, randomly distributed phospholipids (ca. 460 molecules) and water (ca. 10000

grains). The necessary number of sodium cations were added to ensure a neutral charge of the system. The periodic box was the same as in micelle dynamics. Ten simulations with DPPC and ten with DLPC phospholipids were run for 1  $\mu$ s each starting from different NMR

In some simulations, the lipid discs were created, and in such cases CTF resided primarily close to an edge of the disc what altered the structure of protein, and had an influence on values of angles between helices. A representative structure of CTF in DPPC membrane is shown in Fig 8.2. There is a hydrophobic mismatch between the longest helix 8 and the membrane resulting in large angle about  $50^\circ$  to the membrane normal. The CTF structure was stable during the 2  $\mu$ s extension of CG simulation. The value of h7/h8 angle was focused around  $130^\circ \pm 10^\circ$  whereas h9a/h9b angle stabilized around  $90^\circ$  with much broader range of variability similar to that in a micelle simulation. Initially the latter value was about  $60^\circ$  but it was shortly changed to  $90^\circ$  after about 0.5  $\mu$ s of simulation. The whole N-terminal part of CTF was located out of the hydrophobic part of the membrane and both helices h $\alpha$  and h $\beta$  resided in hydrophilic part of the membrane. The catalytic residue Asp385 and PAL motif on the loop between helix 8 and helix 9a were located just beneath the hydrophilic part of the membrane and close to one another. A short loop between helices 7 and 8 as well as a short segment beyond helix 9b were not crossing the membrane/water border but they preferred to reside close to internal membrane border between its hydrophobic core and hydrophilic layer.

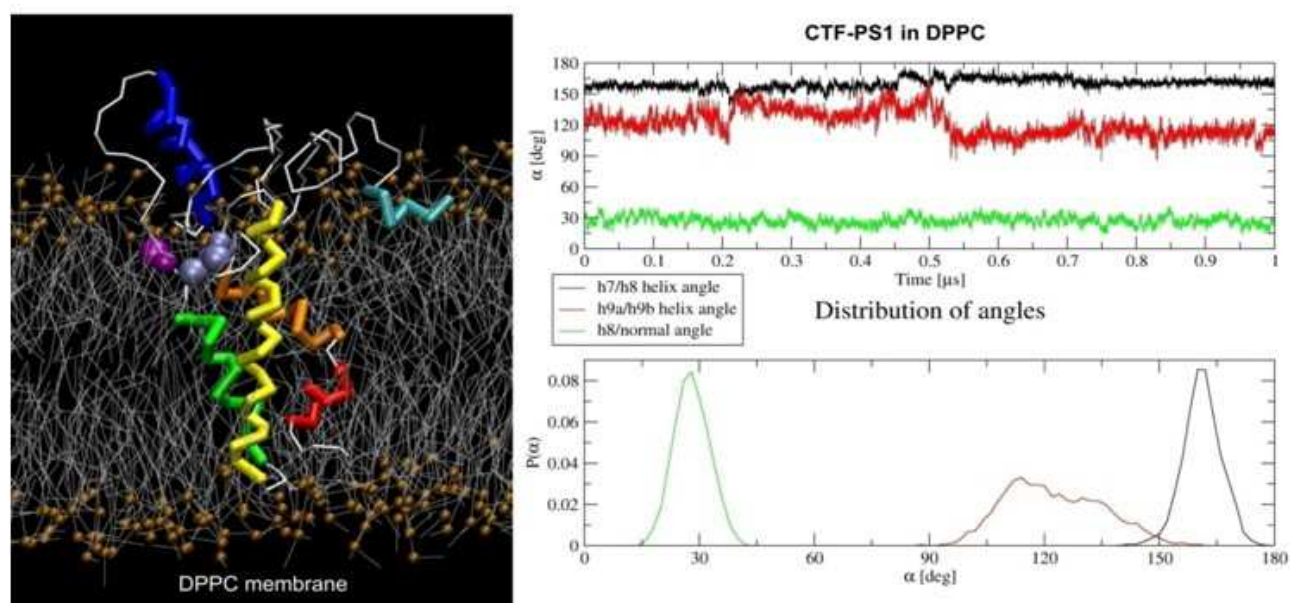


Fig 8.2 Structure of Coarse Grained CTF in DPPC membrane after 3  $\mu$ s simulation, following a 1  $\mu$ s simulation with random distribution of lipids. Evolution of angles between helices 7 and 8 (black), helices 9a and 9b (red), and helix 8 with a normal to the plane of the membrane and their respective distributions .

This is against the hydrophobic mismatch of CTF in this membrane and a tendency of helix 8 to adopt a position perpendicular to the membrane. However, a tryptophan residue Trp404 in this loop



prevents it from crossing the membrane/water border regardless of presence of adjacent Asp and Asn residues. Additionally, the segment beyond helix 9b is ending with a hydrophobic motif FYI which efficiently prevent this part of CTF structure from going into bulk water and attaches it into internal membrane border. In fact these both segments resided close together in nearly all CG simulations. The similar situation is for the thicker membrane composed of DPPC lipids.

Although there is nearly no hydrophobic mismatch of helix 8 (30° of the angle between helix 8 and a normal to the membrane plane) both segments are residing between hydrophobic core of the membrane and its hydrophilic part (Fig. 8.2). The catalytic Asp residue is close in space to the PAL motif and they both are residing in the opposite hydrophobic/hydrophilic internal border of DPPC membrane. The helices 7 and 8 are more antiparallel (160° ±10°) than in a case of DLPC membrane (130°) whereas a distribution of an angle between helices 9a and 9b is similarly broad as in simulation in DLPC although mean value was changed from 90° to about 120° toward more antiparallel orientation of helices 9a and 9b.

### 8.1.3 MD Simulations in Implicit Lipid Bilayer

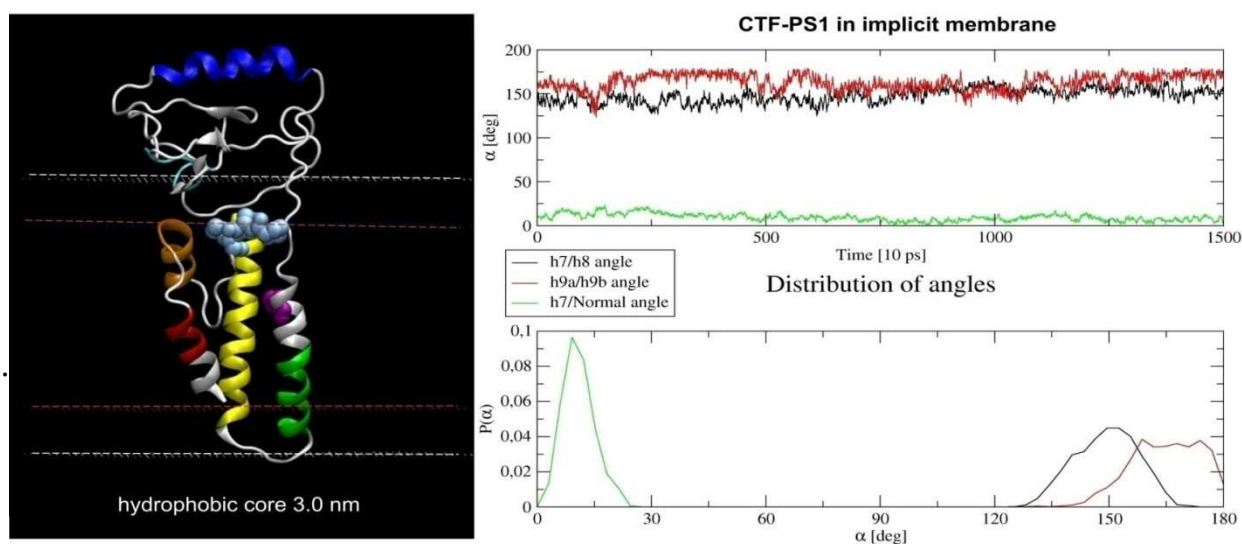
To perform simulations with atomistic representation of protein, but retaining a possibility for fast convergence to final structure, we employed Implicit Membrane Model (IMM1) method [215] in the program CHARMM [231]. The implicit solvent (or in other words continuous environments – CE) method reduces the number of degrees of freedom that are necessary for the evaluation of energy and force, which allows for more efficient sampling of phase space than explicit solvent methods. The water/membrane/water system in IMM1 is achieved by changing the parameters of the system along the line perpendicular to the membrane. In the layer 0.6 nm thick between water and membrane environments the properties are changing smoothly according to a sigmoidal function so at the hydrophobic border of this layer 90% of the environment derives from the hydrophobic core while at the water border 90% of the environment derives from bulk water properties. Two different values of hydrophobic core thickness were chosen 2.6 nm and 3.0 nm. For the former we performed twenty Replica Exchange (REx) simulations (each REx was run in eight temperature windows ranging 300 K) starting from twenty different NMR structures. Such parallel simulations of copies of the system (replicas) in different temperatures allowing exchanges of replicas at regular intervals of time with acceptance ratio dependent on probability based on Metropolis-Hastings criteria [279]:

$$p = \min \left( 1, \frac{\exp \left( -\frac{E_j}{kT_i} - \frac{E_i}{kT_j} \right)}{\exp \left( -\frac{E_i}{kT_i} - \frac{E_j}{kT_j} \right)} \right) = \min \left( 1, e^{(E_i - E_j) \left( \frac{1}{kT_i} - \frac{1}{kT_j} \right)} \right)$$

improves convergence of the protein structure to the near global optimum solution. Because the

bilayer is already organized in this method the orientation of the initial structure of CTF had an influence on the final state. We oriented the NMR structures in such a way that helix 8 was entirely immersed in the membrane and parallel to the membrane normal. However, in many REx simulations the N-terminal part of CTF stayed divided between two parts of the membrane. Such structures were removed from subsequent MD simulations.

The representative structures taken from simulations in 2.6 nm and 3.0 nm hydrophobic cores membrane are shown in Fig. 8.3. Similarly to CG simulations the helices  $h\alpha$  and  $h\beta$  are outside the membrane but close to water/membrane interface with exception of  $h\beta$  in case of thicker membrane. The angle of helix 8 with normal to the membrane plane is about  $15^\circ$  for thinner membrane and  $10^\circ$  for thicker one. In both membranes the helices 7 and 8 stay antiparallel and an angle between them is  $160^\circ$  (thinner membrane) and  $150^\circ$  (thicker). The big difference is for helices 9a and 9b because they are nearly parallel ( $30^\circ$ ) in thinner membrane and anti-parallel ( $165^\circ$ ) in thicker one. However, in both membranes they are residing inside the membrane. Such different orientation of helix 9a (helix 9b is oriented in the same way in both structures) denotes how easily the helices may adopt different conformations in the absence of explicit solvent if there is enough space in the thicker membrane for flipping the helix. A short C-terminal segment beyond helix 9b is close to the loop between helices 7 and 8 as in was in CG simulations.



*Fig 8.3 Structure of CTF in implicit membrane after MD simulation with hydrophobic core thickness of 3.0 nm. Red dashed lines indicate the hydrophobic core of the membrane and white dashed lines denote the border of the bulk water. The evolution of angles between helices 7 and 8 (black), helices 9a and 9b (red), and helix 8 with a normal to the plane of the membrane and their respective distributions.*

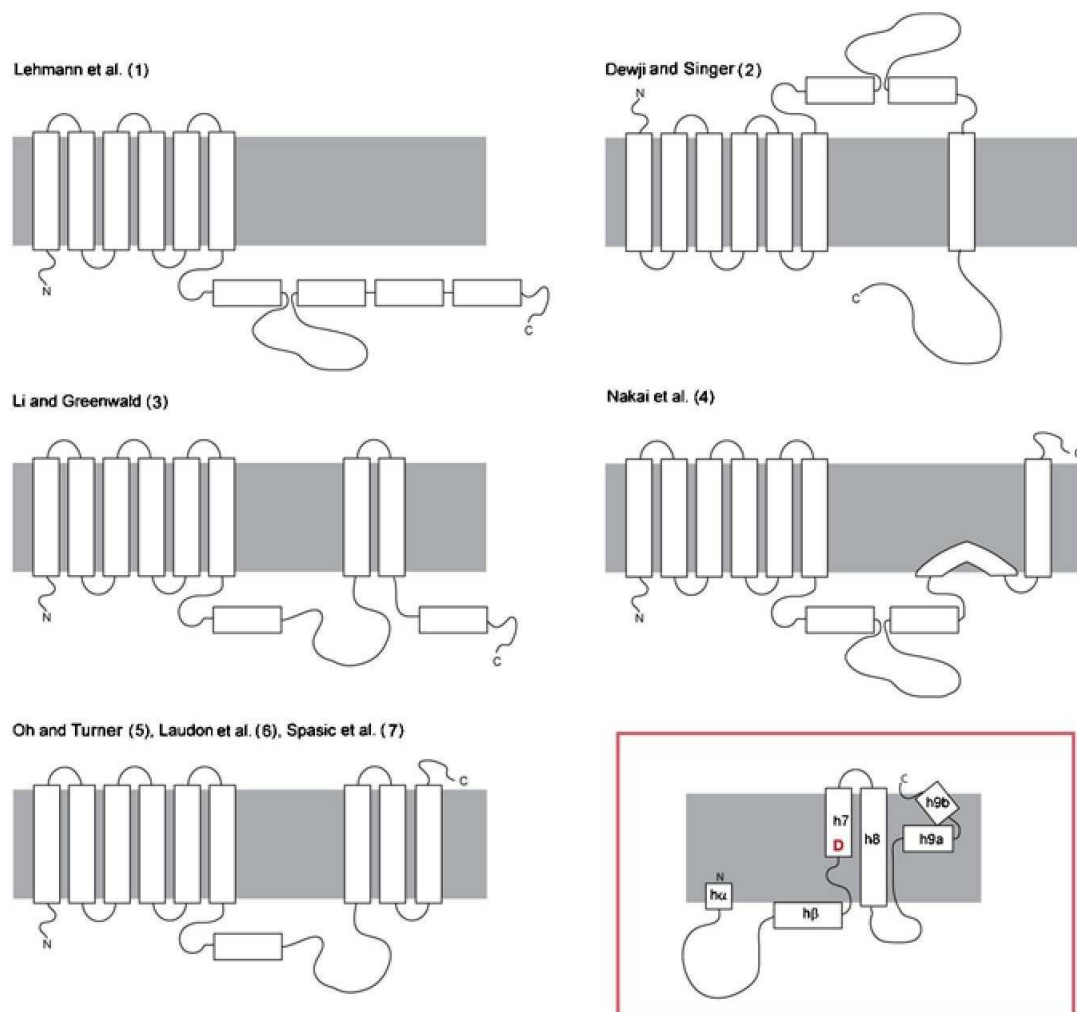
However, a catalytic residue Asp385 is located in the center in the membrane and not close to the internal hydrophobic/hydrophilic border. The PAL motif resides close to this border so its distance to Asp385 is larger than in CG simulations. The interesting feature of CE simulations compared to CG

is that secondary structures are not restrained and may unfold and refold. In case of CTF simulation in 3.0 nm membrane the helices 7 and 9b were elongated so the catalytic residue Asp385 became a part of a helix. However, much longer unconstrained atomistic simulations are required to confirm this result

As shown in the molecular dynamics simulations, the structure obtained after simulation in micelles is structurally close to that of NMR structure in SDS micelles. However, to get insight into the whole structure of  $\gamma$ -secretase complex the additional investigations are required for structure of other parts of the complex and also to reveal molecular role of AD mutations and substrate recognition.

The CTF structures obtained in bilayer simulations (both in coarse-grain (Fig. 8.2) and implicit membrane (Fig. 8.3) showed larger difference to that of NMR structure. The angle between helices 7 and 8 remained relatively stable but dependent on membrane width, whereas that between partial helices 9a and 9b were changing. The bilayer environment in the simulations promoted also longer membrane helices so additional investigations are needed to validate such effects.

### 8.1.4 Topology of PS1-CTF



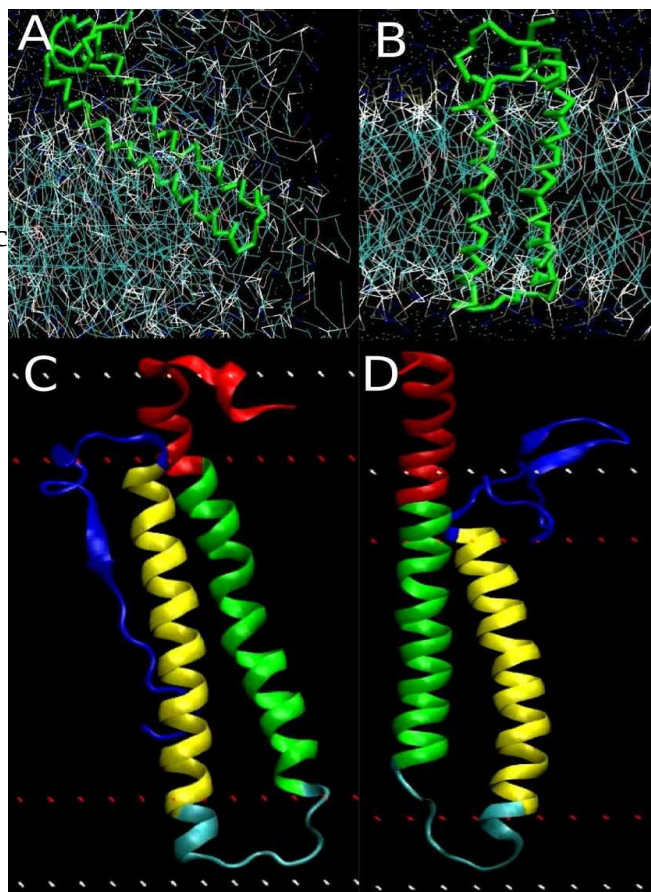
*Fig 8.4 Proposed topologies of CTF. Topologies, as suggested by several groups based on biochemical evidence (mutation and immunofluorescence studies) are shown. The first six helices belongs to NTF whereas the latter helices belong to CTF. Proposed topology of CTF topology based on NMR evidence is bordered in red.*

CTF structure revealed by NMR studies has three membrane spanning regions which is in agreement with the nine transmembrane domain model of presenilin 1 (Fig. 8.4). However, it has novel characteristics in order to facilitate intramembrane catalysis. It contains a putative half-membrane-spanning helix N-terminally harboring the catalytic aspartate, a severely kinked helical structure toward the C terminus as well as a soluble helix in the assumed-to-be unstructured N-terminal loop.

### 8.1.5 Modeling of Pen-2

Pen-2 is

- Required for Notch Signalling
- Processing of APP by  $\gamma$ -secretase
- Along with APH-1 it regulates proteolytic processing of presenilin
- It stabilizes CTF and NTF of presenilin
- Zebrafish lacking Alzheimer presenilin enhancer 2 (Pen-2) demonstrate excessive p53-dependent apoptosis and neuronal loss



*Fig 8.5: Initial (A) and final (B) position of Pen-2 during the simulation of DPPC lipid bilayer formation starting from random orientation of Pen-2 and DPPC molecules. Initial(C) and final (D) structures of Pen-2 during 20ns simulation in IMM1 with 4 nm hydrophobic core. The hydrophilic N-terminal loop which is inside the membrane in the model, projects outward of the bilayer after the simulation.*

Biochemical cysteine cross-linking experiments have indicated the interactions between PS1 NTF and Pen-2. It was proposed that TMD4 of PS1 interacts with TMD1 of Pen-2 [31,56]. GS modulators were shown to bind to Pen-2.

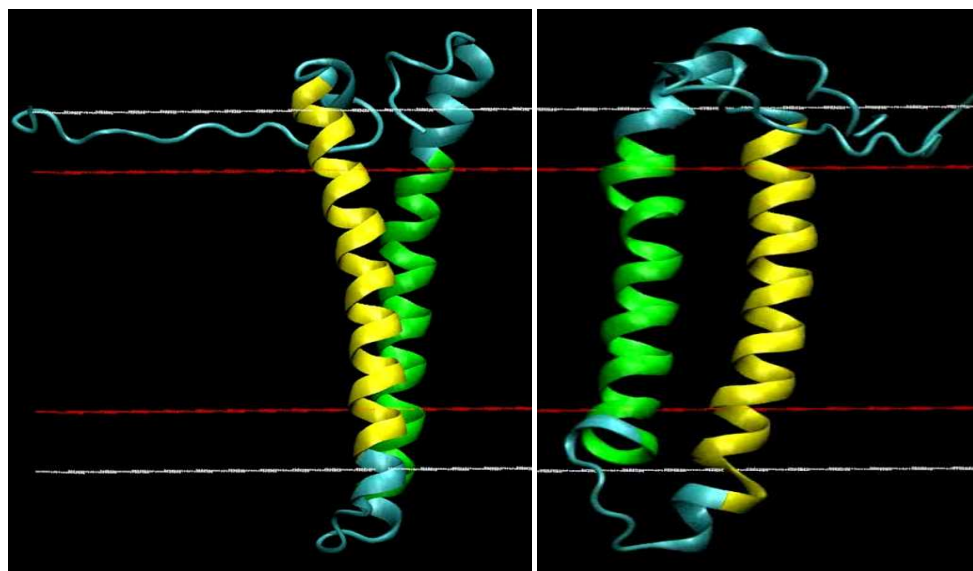
During the simulation in IMM1, if the width of the membrane is increased, the protein tends to form longer N terminal helix and the C-terminal comes out to the ECF as one should expect. Otherwise, in few models the C-termini was buried in the membrane. For those models, where C-termini was outside the membrane, during both explicit and implicit membrane simulation, they form compact structure (often a beta sheet and comes close to the helix). Therefore the conclusion is that most of the models converge to similar structures during MD simulation Fig 8.5.

From the models of PEN-2, its chemical shift was predicted using the software SPARTA. As found in the size exclusion chromatography, Pen-2 forms monomer or dimer depending upon its

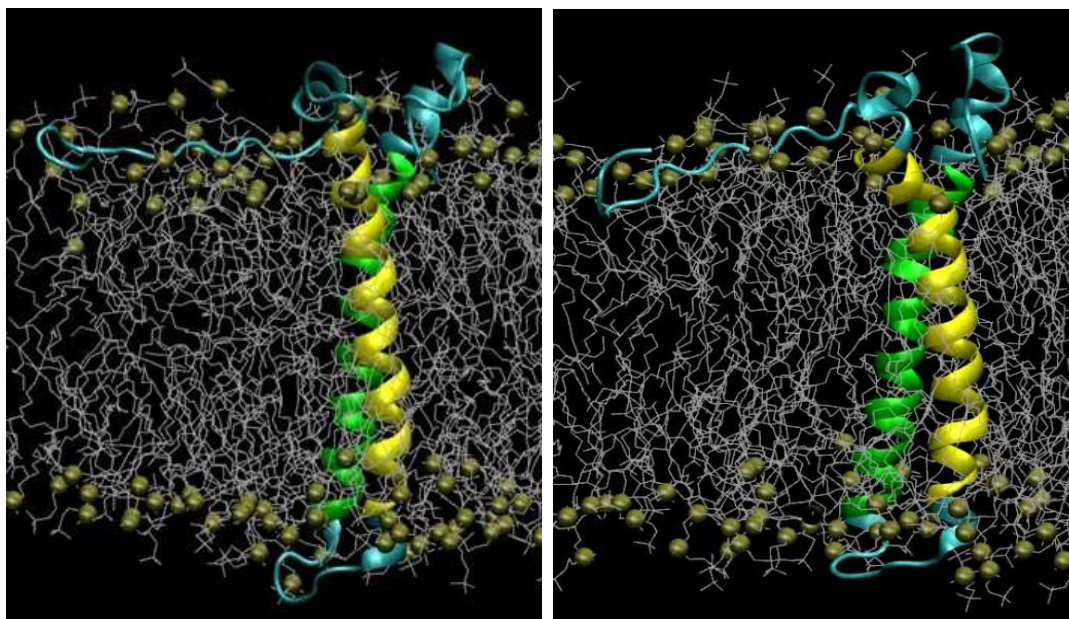
concentration. Pen-2 could dimerize in plasma membrane. There is only one cysteine at position 15 at N-terminal helix of Pen-2 which mediating interaction.

The RMSD change after all atom 80ns simulation in POPC membrane remains close to 3-4 Å for backbone atoms. The change in RMSD is mostly in the hydrophilic N-terminal and C-terminal loop region of Pen-2 which keeps moving throughout the simulation. Moreover, the loops become more compact by the end of the simulation.

During 20ns simulation in IMM1, the structure of Pen-2 change quite rapidly especially in the loop region which become more and more compact in shape and come close to the TM helices. Experimental evidence show that the C-terminal loop bind to the Presenilin 1 CTF and is associated with the assembly of the  $\gamma$ -secretase complex. In some models where the N-terminal loop is located inside the membrane ejects towards the outside during the simulation. This supports the experimental results. In addition, formation of Pen-2 embedded in lipid bilayer can be simulated starting from random orientation of Pen-2 and lipid in Martini coarse-grained forcefield. During 50 ns equilibration in Martini force field, bilayer formation was observed starting from random mixture of Pen-2 and DPPC lipids. Further during the simulation of Pen-2 of topology 1, the length of the TM helices increase depending on the hydrophobic core thickness used in the simulation and the resulting structure becomes close to that of topology 2.



*Fig 8.6: Starting (A) and final (B) structure for Pen-2 after 20 ns Simulation in implicit IMM1 (2.7 nm hydrophobic core thickness ) implemented in CHARMM*



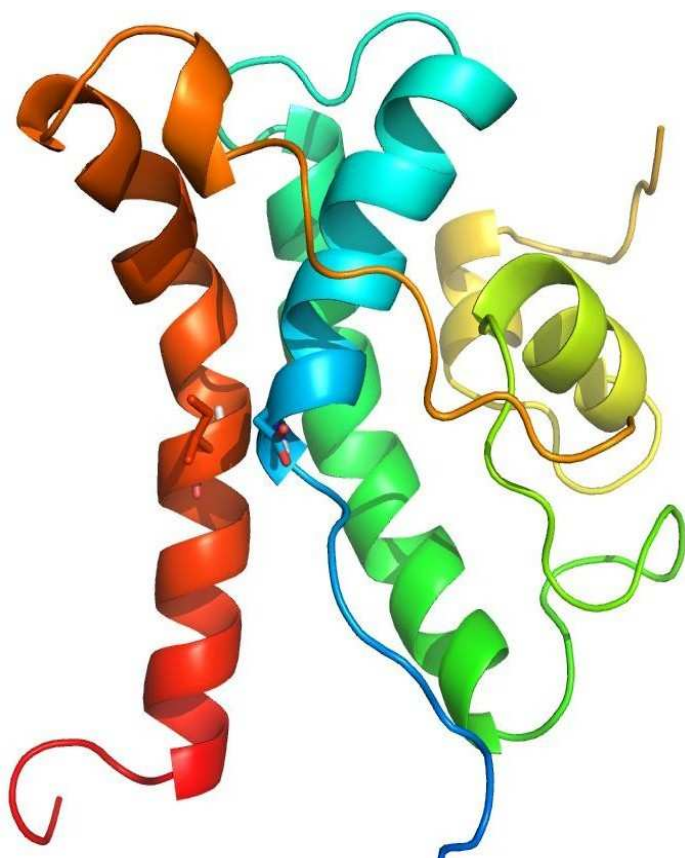
*Fig 8.7: Starting (left) and final (right) structure for Pen-2 after 80 ns simulation in explicit POPC membrane surrounded by water and ions (not shown). The structure remains almost unchanged except the change in the C-termini which becomes more compact and interacts with Presenilin.*

In absence of experimental data, the models of Pen-2 can be unreliable since there is no significant homology of Pen-2 compared to already solved structures like in GPCRs. So further experimental evidence is required in order to verify the models.

## 8.2 Modeling of PS1 CTF-APP Interface and CTF L383 mutations

By using protein protein docking programs like ClusPro [280], GRAMM [281] and HADDOCK [282] which does 6 dimensional conformational search by rotation and translation, models of PS1 CTF – APP were obtained Fig 8.8.

After mutating the GLGD motif of CTF to WLWD or WLGD or GLWD etc (big tryptophans should cause steric hindrance to decrease the stability we observe higher interaction energy (i.e. less stability) in mutants. Another interesting fact is that binding site of the APP changes and the new site has good interaction energy. Therefore it might be possible that mutating one or two residues leads to alteration of binding site of CTF and APP which is away from the catalytic Asp. That's why there is



*Fig 8.8: Docking of APP (PDB ID:2lp1) C-termini (red) to PS1 CTF without long N-termini. The catalytic aspartate of CTF and Val of APP which is cleaved off during A $\beta$ 40 formation has been shown. sometimes no A $\beta$  production or less production.*

We performed MD in IMM1 with implicit solvent to find out the distance distribution between catalytic Asp of CTF and the peptide bond it cleaves for forming Ab40 and Ab42 respectively. So MD simulations in IMM1 starting from the docked conformations of the mutants (Fig 8.9 and Fig 8.10).

It appears that the distances depend on the starting conformation of the CTF-APP interface, it is



changing during MD simulation and one might get more stable distance distribution from prolonged MD simulation. However in these short simulations the distance for Ab40 formation decreases a bit and that of Ab42 slightly increases which supports the formation of higher Ab42 in these mutants. For WLGD (G382W) where is no activity the Ab40 distance slightly increase and Ab42 distance decrease a lot. The tryptophan is somehow interfering the catalysis.

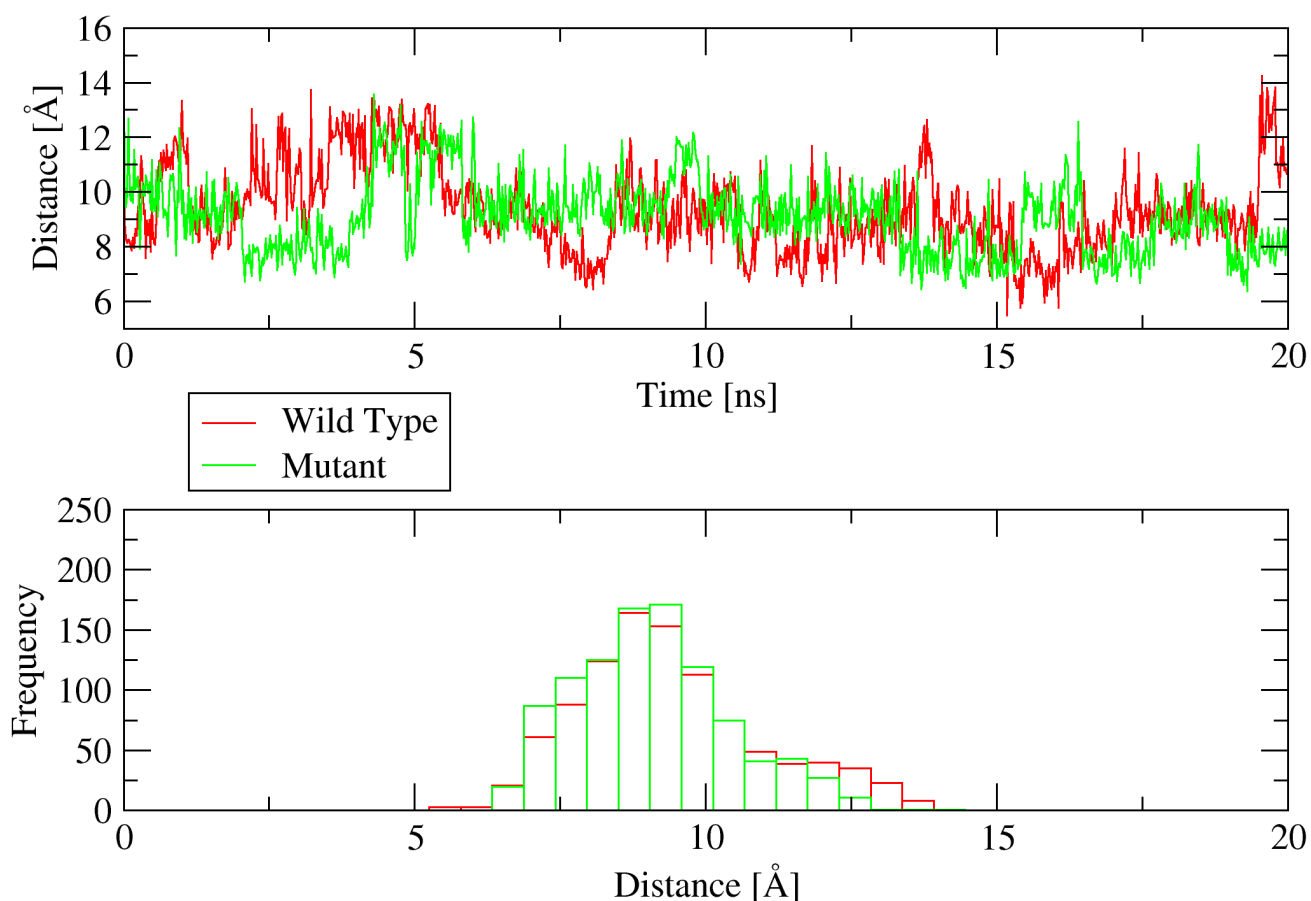
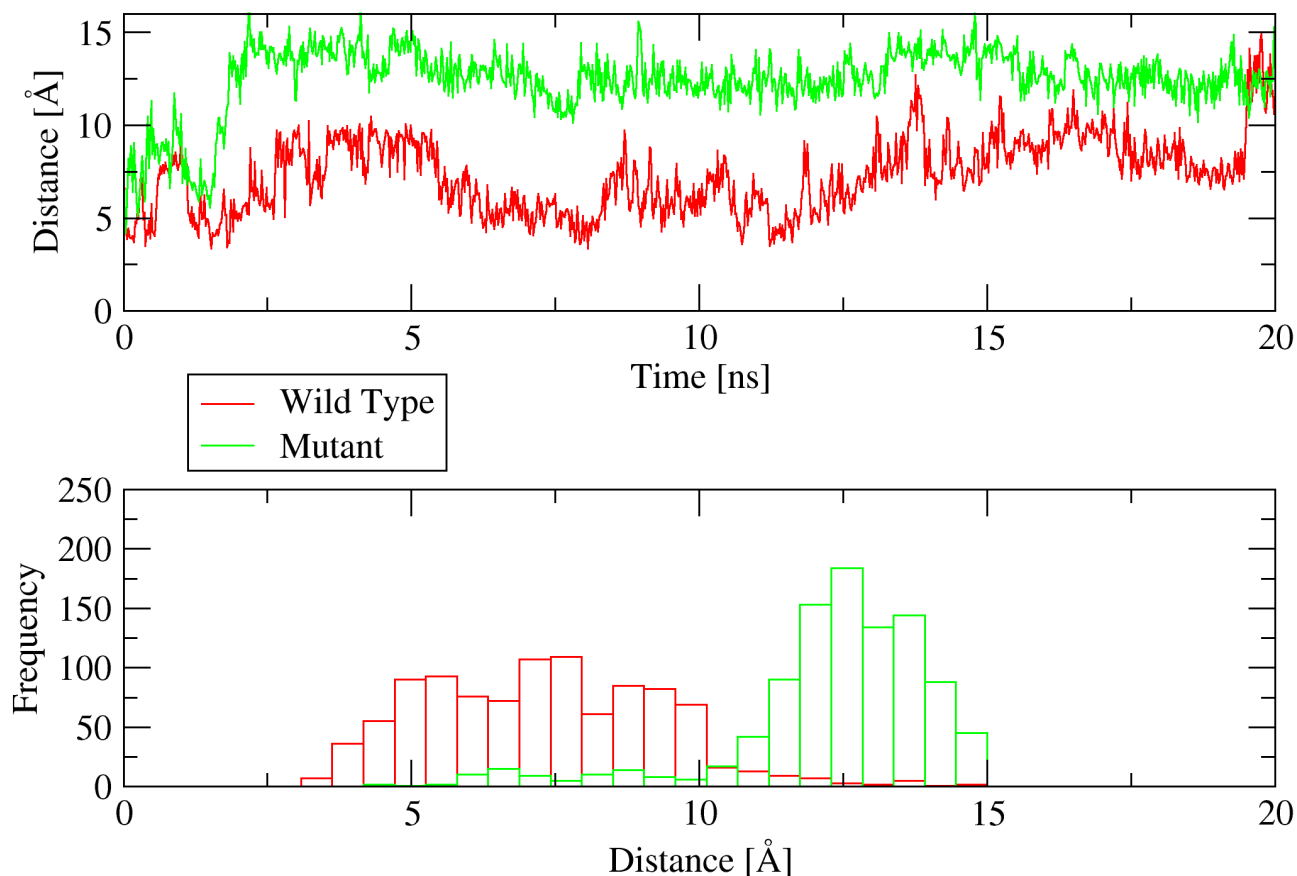


Fig 8.9: In case of L383A mutation, the distance between catalytic aspartate to the cleavage site for Ab42 formation decreases resulting in higher Ab42 production compared to WT.

During all atom simulations in POPC membrane, water molecules interacting with catalytic Asp was found. In case of the CTF structure, half of it is in hydrophilic environment. APP is probably in the water cavity as found in the cryoEM structure. Another interesting fact he pointed out is that in order to cleave the peptide bond in APP, the helix must be unfolded before the cleavage which is typical for



*Fig 8.10: In case of L383W mutation, the distance for cleavage to produce Ab42 decreases, resulting in higher Ab42 levels*

intermembrane cleaving proteases. Aspartates can't attack the alpha helix directly. There are many possibilities of a contact CTF-APP because the APP is probably unfolded in the region of cut. H7 is probably unfolded to form a pore but degree of unfolding is not known. Probably D385 is part of a helix like in NMR CTF structure. The helix of a substrate is probably also slightly unfolded, however,

there is no structure of protease-substrate complex in such reaction step. The APP must be a helix before a contact with H7 (probably binding H8 or even APH-1 with the same GxxxG motif). So we can trace two structures: (1) with helical APP not in contact with D385, and (2) with slightly unfolded APP (possibly only one helix turn or even a half) ready for a cleavage.

We started with a hypothesis that H7 of CTF is longer than that of the reported in NMR structure. It would add another GxxxG interactions. However that part of helix unwound more or less in all simulations, which supports the most probable hypothesis of water pore.

It was found in the experiments that the Ab42 production increases maximum in L383W, then L383P, L383Y and then L383F. In L383P, there is much more Ab43 produced compared to Ab42 and almost

half of Ab40. And in G382W, there is no activity.

Based on the paper by [35] the substrate binding site of presenilin should lie near the active site. During 40 ns simulations in g53a6 ff of CTF (h7 and h8), APP interface in POPC membrane for native, L383P and L383W. The Ab40 distance is around 8-10 Å and is always less than Ab42 and Ab43 distance which is reasonable. But sometimes there is initial velocity effect and the CTF and APP tend to move away from each other (It is not reproducible by repeating the simulation). Similar things were noticed also in IMM1. From DFT studies [283] found that the cleavage of Val-Ile bond (Ab 40) is 9 times more favourable than Ala-Thr (ab40) which is similar to the data we have. Mutations probably change the charge distribution or cause steric hindrance among these residues, resulting in alteration of cleavage specificity due to different energy barriers for cleavage. Therefore in the mutations L383P, L383W and G384A: reduced activity and higher Ab42 production because of the population of APP conformations close to Asp with Ab42 cleavage site increases. G382W: no activity probably due to the presence of Trp close to Asp causing steric hindrance and failure of catalysis.

I performed a 20 ns simulation of the native ctf-app complex in Implicit Membrane (2.6 nm hydrophobic core and ) in CHARMM. Then I measured the distance between OD1 of Asp and N of Ile of APP (the peptide bond it cleaves for Ab40 generation during catalysis in presence of water) over the trajectory (as shown in the plot, x axis is represents 20 ns of simulation expressed in 1000 frames and y axis represents the distance in Angstrom. The second plot is the histogram of distance distribution over the trajectory) It is worth keeping in mind that since there is no friction in implicit membrane, the original length of simulation is much higher than 20 ns. Since water is present during catalysis, this interaction could further be probed by Molecular Dynamics simulations in explicit membrane.

Due to the topology problem GxxxG motif interactions of APP TMD with the GVKLG motif in PS TM7 cannot take place. So the correct situation would be like this:

GVKLGLGDFIFYSVLV GKASATASG  
 KKKLMVLTIVIVTAIVVGGVMLGIIAGKNSG

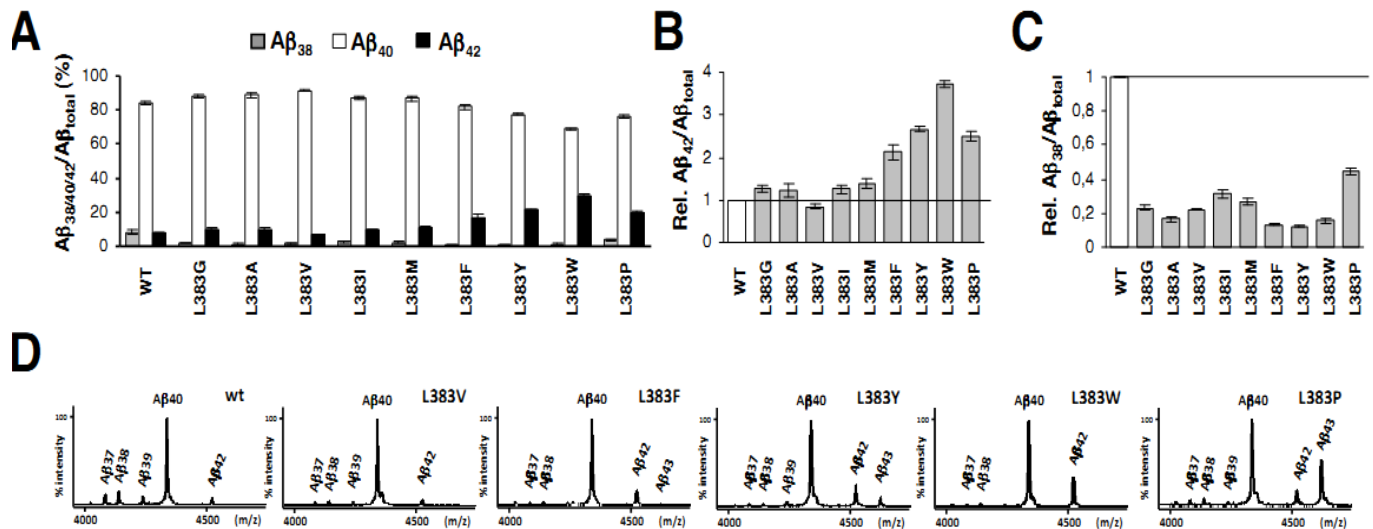
Potentially PS TM8 (shown below in the correct orientation relative to substrate) could aid in the interaction via helix-helix interaction motifs:

AKKFIALLLLTLCL GI LIAVFCAITTN  
 KKKLMVLTIVIVTAIVVGGVMLGIIAGKNSG

### Outlook:

We have made several attempts to help you explain the effects of mutations within GLGD motif in PS1 CTF. After thorough review of literature, we have come to a conclusion that we should try modeling a situation which is closer to reality. Here is a sequence alignment of gamma-secretase substrates, Notch1, CD44 and APP with cleavage sites added. We noted striking similarities between

the substrates and strongly believe that we should change our approach and model CTF-NTF-substrate complex, where the substrate is a helix heavily tilted so that the cleavage site and catalytic center are near. Next, the top three hydrophobic residues would be unwound and the basic and polar residues that follow provide the unwinding force by interacting with water pore. After the first cleavage, we should model Ab43 being cut to Ab40, but this time charged C-terminal carboxyl group would interact with water to play the same role as the basic residues during the first cleavage. To do our task the best possible way, we have several questions regarding the complex we



**Fig 8.11:** (A) Levels of secreted A<sub>38</sub>, A<sub>40</sub> and A<sub>42</sub> species in conditioned media of HEK293/sw cells stably expressing H<sub>6</sub>X-tagged PS1 wt or the indicated PS1 L383 hydrophobic mutants were quantified by a highly specific A sandwich immunoassay and plotted as a percentage of the total A measured. Bars represent the mean of three independent experiments ± S.E. (B, C) Data of (A) were plotted such that A<sub>42</sub>/A<sub>total</sub> ratios (B) and A<sub>38</sub>/A<sub>total</sub> ratios (C) produced by the PS1 L383 mutants were expressed relative to those of PS1 wt that were set 1.00. (D) Secreted A in conditioned media of HEK293/sw cells stably expressing H<sub>6</sub>X-tagged PS1 wt or the indicated representative PS1 L383 mutants was immunoprecipitated with antibody 4G8 and subjected to MALDI-TOF MS analysis. Note the relative peak changes of individual A species. Note that the spectrum of L383V is representative for the L383G, L383A and L383I mutants, which show a similar profile of A species. Source: Prof. Harald Steiner

want to model.

## 8.2.1 SVM Predictions

Support vector machine (SVM) is a powerful way of machine learning successfully used in various fields of Bioinformatics. Here SVM implementation SVMlight was used to train the MOLGEN mutation dataset (<http://goo.gl/I7cb7>) of Presenilin 1 mutations which results in alteration of Ab40/Ab42 ratio. Increase in the ratio was indicated as positive outcome (+1) and decrease with (-1).

The amino acids are classified into hydrophobic (1), polar (2) and charged (3) ones, there are  $3^2=9$  types of mutation possible. The SVM was trained like increase in Ab40/Ab42: +1 135:12 , decrease in Ab40/Ab42: -1 25:23 where second column is the residue number and third column is the change due to mutation (hydrophobic to polar in first example). Further all the mutations were encoded by sparse encoding.

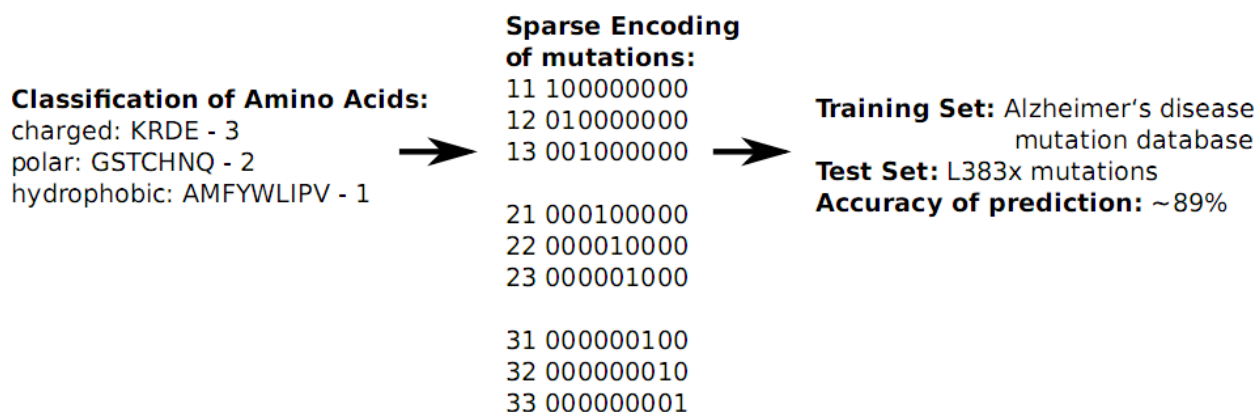


Fig 8.12 SVM prediction methodology

After filtering when the dataset was used the training set for classification in svmLight [284], it gives 88% accuracy and with the experimental mutation set of 17 mutations of PS1 L383, it gives 89 % accuracy.

The amino acids are classified into hydrophilic(1), hydrophobic (2) and polar (3) ones, there are 9 types of mutation. The machine can be trained like increase in Ab40/Ab42:

1 135:12

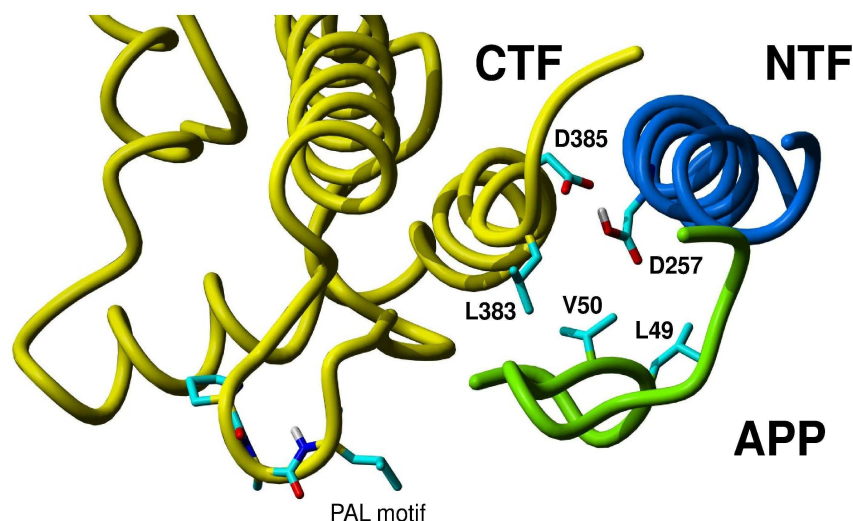
decrease in Ab40/Ab42

-1 25:23

where second column is the residue number and third column is the change due to mutation (hydrophilic to polar in first example). So there are  $3 \times 3 = 9$  possibilities of mutation i.e. 1 → 1 i.e. 11, 12, 13, 21, 22, 23, 31, 32, 33. But since 22 is double of 11, it can result in false training. So sparse encoding approach was used and the following mutations are encoded like in the figure. After filtering, there was 57 reliable mutations to train and many of the mutations result in negative outcome (-1). When I used the training. Due to lack of structural information of the complex modeling did not lead to any convincing results so far.

Currently the method is being improved by adding information from position specific scoring matrices (PSSM) derived from multiple sequence alignments. Further neural network is being used to improve the quality of the predictions.

## 8.2.2 Modeling APP, PS1 CTF and NTF Interface:



*Fig 8.13 Proposed interaction sites of APP with PS1 CTF and NTF showing catalytic residues D385 and D257 also the first cleavage site L49 and Ab40 cleavage site V50.*

We tried many different arrangements of APP, CTF and NTF fragments including that based on recent FlaK protease structure [88]. Inclusion of some restraints will be necessary to keep pieces of the active site together because of lack of the rest of the gamma-secretase complex. However we still did not obtain a model which can describe properly the experimental findings at least qualitatively. We added more cleavage sites (for instance at 49) on the figure to get a more clear picture. We used one predicted helix from NTF that contains the second catalytic aspartate. This used a piece encompassing aminoacids from 244 to 264  $\pm$  3: WTAWLILAVISVYDLVAVL. 2006 SCAM data from Iwatsubo and De Strooper it is reasonable to assume a helix, which may be kinked according to the Sato data. The helix could also be like this: WTAWLILAVISVYDLVAVL: CP are not in any more as shown in most models of PS topology. Fig 8.13 shows a model of one of arrangements with fragment of APP approaching the active site for the first cleavage at L49. APP is in unfolded conformation. Other cleavages at 42 and 40 were also shown.

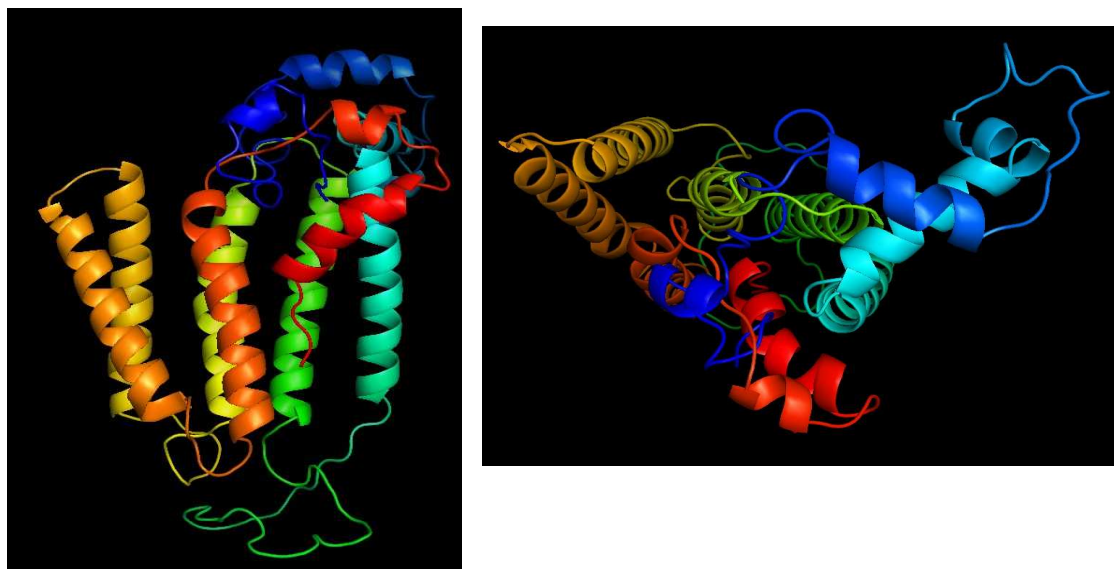
### **Modeling based on density functional theory (DFT):**

I performed geometry optimizations and transition state search of the APP-PS1 complex with the co-ordinated given in that DFT paper usin Gaussian09 [283]. However, the results could not be reproduced (i.e. no cleavage of the peptide bond was observed). Further calculations were performed

in PM6-DH2 using Mopac2009 without fruitful results.

## 9 Modeling of PS1-NTF

Models of NTF were obtained in the similar way to Pen-2 and were scored by a consensus Rosetta, ProQ and Prosa2003 scored. Following is the top scoring model of PS1-NTF without the N-terminal loop (82 amino acids).



*Fig 9.1: Model of PS1 NTF. left: side view showing six transmembrane helices. right: top view*

## 10 Activation mechanism of GPCRs

### 10.1 Introduction

GPCRs (also called 7TM receptors) form a large superfamily of membrane proteins, which can be activated by small molecules, lipids, hormones, peptides, light, pain, taste and smell etc. Although 50% of the drugs in market target GPCRs, only few are targeted therapeutically [285]. Such wide range of targets is due to involvement of GPCRs in signaling pathways related to many diseases i.e. dementia (like Alzheimer's disease [87]), metabolic (like diabetes [286]) including endocrinological disorders [287], immunological [288] including viral infections [289], cardiovascular [288], inflammatory [290], senses disorders [291], pain [292] and cancer [293]. Upon activation by extracellular agonists, GPCRs pass the signal to the cell interior. Ligands can bind either to extracellular N-terminus and loops (e.g. glutamate receptors) or to the binding site within transmembrane helices (Rhodopsin-like family) (Fig 10.1). They are all activated by agonists although a spontaneous auto-activation of an empty receptor can also be observed.

Biochemical and crystallographic methods together with MD simulations, and other theoretical techniques provided models of receptor activation based on the action of so-called “molecular switches” buried in the receptor structure. They are changed by agonists but also by inverse agonists evoking an ensemble of activation states leading toward different activation pathways. Switches discovered so far include the ionic lock switch, the 3-7 lock switch, the tyrosine toggle switch linked with the nPxy motif in TM7, and the transmission switch proposed *ad hoc* based on a flood of recent crystal structures (instead of the tryptophan rotamer toggle switch which seems to work in another way than it was thought before). The suggested global toggle switch consisting of vertical rigid motion of TM6 seems to be implausible based on the recent crystal structures of GPCRs with agonists. Because of the intrinsic instability of GPCRs resulting in their multiple functionality, the investigations of the action of molecular switches are extremely difficult but may provide highly selective drugs acting not even on a single receptor subtype but a single pharmacological subprofile [294–296]. The detailed knowledge of the GPCR activation mechanism could be very useful in designing specific drugs. We proposed novel activation mechanism of cannabinoid receptors CB<sub>1</sub> and CB<sub>2</sub> [294], and elucidated the role of water in the activation of formyl peptide receptor (FPR1) [296]

GPCRs interact with very diverse sets of ligands which bind to the transmembrane (TM) segments and sometimes also to the receptor extracellular domains. Each receptor subfamily undergoes a series of conformational rearrangements leading to the binding of a G protein during the activation process. All GPCRs preserved the 7-TM scaffold during evolution but adapted it to different sets of ligands by structure customization. Binding of structurally different agonists requires the disruption of



different intramolecular interactions, leading to different receptor conformations and differential effects on downstream signaling proteins. The dynamic character of GPCRs is likely to be essential for their physiological functions, and a better understanding of this molecular plasticity could be important for drug discovery.

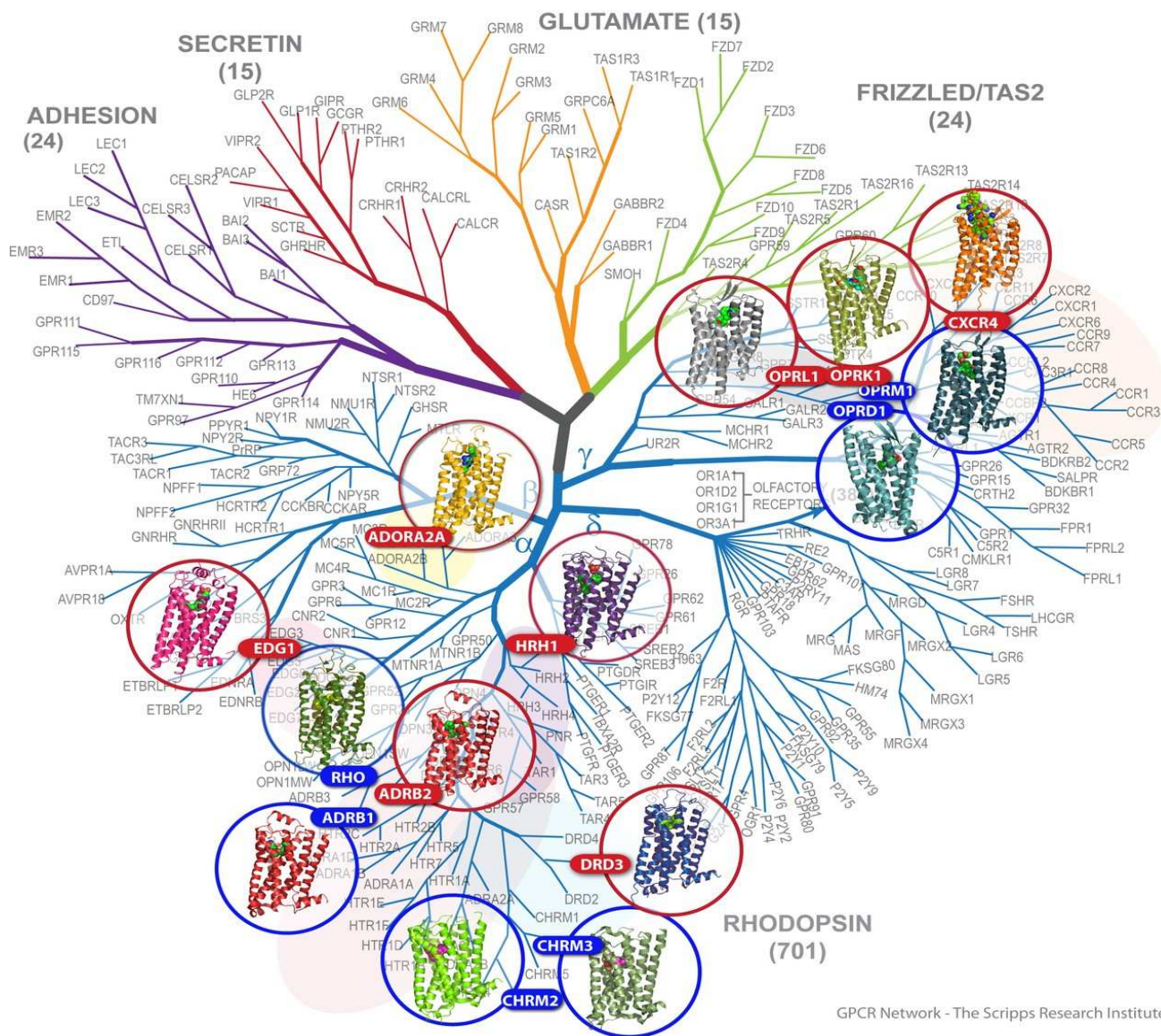
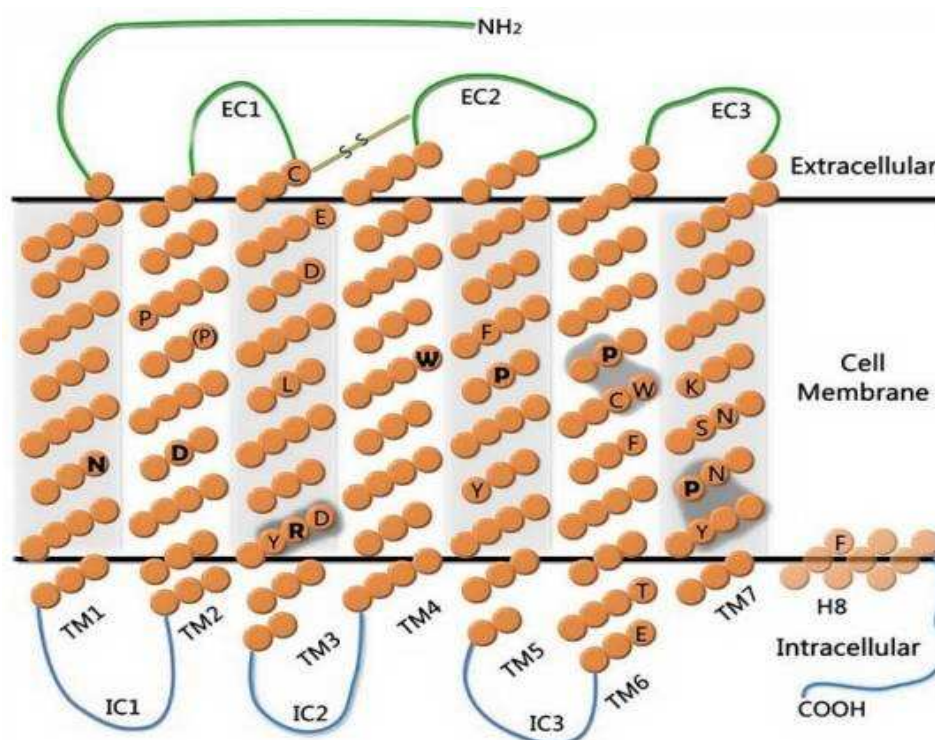


Fig 10.1 Phylogenetic tree of GPCRs showing solved structures in cartoon diagrams. (Source: <http://gpcr.scripps.edu>)

Experiments suggest that agonist binding and receptor activation occur through a series of conformational intermediates. Transition between these intermediate states involves the disruption of intramolecular interactions that stabilize the basal state of a receptor. Such profound changes are evoked by the action of molecular switches. The switches proposed so far for different GPCRs include the “rotamer toggle switch” involving the CWxPxP sequence on TM6, the switch based on the Np<sub>x</sub>xY(x)F sequence linking TM7 and H8, the “3-7 lock” interaction connecting TM3 and TM7 (involving Schiff base-counterion interaction in rhodopsin), and the “ionic lock” linking transmembrane helices TM3 and TM6 and employing the E/DRY motif on TM3. In the rhodopsin structure all these switches are closed (inactive state) [297], however, in the recent crystal structures

of  $\beta$  1- and  $\beta$  2-adrenergic receptor complexes with antagonists and inverse agonists the “ionic lock” is open while the “rotamer toggle switch” remains closed [298].



*Fig 10.2 General scheme of topology and location of conserved residues in Rhodopsin-like GPCRs. Number of residues and their locations in each TM is based on chemokine receptor CXCR4 (H8 is not present in the crystal structure so it is shown transparent). Residues in bold are the most conserved in each TM. Sequence motifs are shown as gray areas. An alternative position of proline residue in TM2 is denoted by (P).*

Highly conserved residues among G-protein-coupled receptors indicate not only molecular switches but they are also located at the interior of individual structural segments, suggesting a dual role for these segments. Firstly, structural segments stabilize secondary structure elements of the native protein, and secondly, they position and hold the highly conserved residues at functionally important environments. Two main classes of force curves were observed in SMFS experiments (Single Molecule Force Spectroscopy) in rhodopsin [299]. One class corresponded to the unfolding of rhodopsin with the highly conserved Cys110–Cys187 disulfide bond remaining intact and the other class corresponded to the unfolding of the entire rhodopsin polypeptide chain. In the absence of the Cys110–Cys187 bond, the nature of certain molecular interactions within folded rhodopsin was altered. These changes highlight the structural importance of this disulfide bond and may form the basis of dysfunctions associated with its absence.

With the determination of the first structure of the complex between a G-protein coupled receptor

(GPCR) and a G-protein trimer ( $G\alpha\beta\gamma$  in 2011 [300,301] a era of GPCR research was opened for structural investigations of global switches with more than one protein being investigated. The previous breakthroughs involved determination of the crystal structure of the first GPCR, rhodopsin, in 2000 [297] and the crystal structure of the first GPCR with a diffusible ligand ( $\beta$ 2AR) in 2007 [266]. How the seven transmembrane helices of a GPCR are arranged into a bundle was suspected based on the low-resolution model of frog rhodopsin from cryo-electron microscopy studies of the two-dimensional crystals [302]. The crystal structure of rhodopsin, that came up three years later, was not a surprise apart from the presence of an additional cytoplasmic helix H8 and a precise location of a loop covering retinal binding site. However, it provided a scaffold which was hoped to be a universal template for homology modeling and drug design for other GPCRs – a notion that proved to be too optimistic [297]. Seven years later, the crystallization of  $\beta$ 2-adrenergic receptor ( $\beta$ 2AR) with a diffusible ligand brought surprising results because it revealed quite a different shape of the receptor extracellular side than that of rhodopsin. This area is important because it is responsible for the ligand binding and is targeted by many drugs. Moreover, the ligand binding site was much more spacious than in the rhodopsin structure and was open to the exterior. In the other receptors crystallized shortly afterwards the binding side was even more easily accessible to the ligand. New structures complemented with biochemical investigations uncovered mechanisms of action of molecular switches which modulate the structure of the receptor leading to activation states for agonists or to complete or partial inactivation states for inverse agonists.

Here I will describe the proposed activation mechanisms together with molecular switches, compare them and try to generalize the findings with respect to the other GPCRs not only from family A (the most populated Rhodopsin-like family) but also other families of these mysterious receptors. The action of molecular switches was most extensively investigated in the case of two types of receptors: rhodopsin and the  $\beta$ -adrenergic receptors. The recent reviews on activation and action of molecular switches in the Rhodopsin family of GPCRs were published in 2009 by Ahuja and Smith [303], Nygaard et al. [304] and also by Hofmann et al. [305]. Some other reviews on activation mechanisms were published earlier by Strange [306] and lately by Deupi and Standfuss [307]. The three year period since 2009 has been very fruitful for the GPCR research and provided detailed explanations on how some of the switches work as well as redefined some hypotheses in this field. In a very recent review [308] Unal and Karnik tried to generalize the concept of molecular switches and came to the idea of a coordinated domain coupling in GPCRs which could be a consequence of the dynamic nature of these receptors. According to this hypothesis when a ligand is bound to a receptor extracellular domain a decrease in the intrinsic disorder of this domain cooperatively changes the conformation of the neighboring receptor domain. Certainly, some other original concepts will be emerging based on the still growing number of crystal structures and other data associated with

### 10.1.1 Superfamily of GPCRs

The superfamily of G-protein coupled receptors (GPCRs) can be divided into five main families: Glutamate, Rhodopsin, Adhesion, Frizzled/Taste2 (consisting of frizzled, smoothed and taste2 receptors), and Secretin, according to the GRAFS classification system [309] which displaced the previous A-F system [310,311]. The GRAFS system was formed using the Hidden Markov Model approach to analysis of multiple sequence alignments of all GPCRs from 13 eukaryotic genomes. All five families were formed in the early stage of metazoan evolution and the number of GPCRs in each family increased during evolution. At present, sequence diversity of GPCRs and their abundance is enormous, giving organisms more ways to adapt to various environmental conditions [312]. Additionally, the Rhodopsin family, the largest and the best described of all, is divided into four groups:  $\alpha$ ,  $\beta$ ,  $\gamma$  and  $\delta$ , out of which only the  $\delta$  group does not have any representative in the PDB database. The above internal classification of the Rhodopsin family is still under discussion as other methods such as NJ (Neighbor-joining) or UPGMA (Unweighted Pair Group Method with Arithmetic mean) provided phylogenetic trees of a different fan-like shape [313,314]. Lately, using the multidimensional scaling (MDS), a non-phylogenetic statistical method adapted to evolutionary distant sequences, Chabbert and co-workers [315] showed that the Rhodopsin family should be divided into 4 groups. The central group, G0, is formed by peptide receptors, opsins and melatonin receptors. The second group, G1, includes somatostatin and opioid receptors, chemokine and purinergic receptors, proteinase activated receptors and acid receptors. The G2 group is formed by biogenic amine receptors and adenosine receptors. Finally, the G3 group consists of receptors for melanocortin, phospholipids and cannabinoids, glycoprotein hormone receptors and leucine-rich repeat (LRR) containing receptors, prostaglandin receptors and Mas-related receptors. This classification of Rhodopsin GPCRs emphasized the role of proline residues patterns in TM2 (transmembrane helix 2) and TM5 (observed in correlated mutations) which was confirmed by the recently solved CXCR4 structure [316].

Despite the large sequence diversity, all GPCRs most probably share the same fold: seven transmembrane helices joined by extracellular and intracellular loops of varied length (see Fig 10.2). A 7TM core is well preserved in all known to date protein structures of GPCRs despite the high degree of sequence variability within this region. It is worth noticing that the seven helix bundle is not a unique feature of G-protein coupled receptors, since there are other proteins in eukaryotic and prokaryotic organisms which share this fold. For example in eukaryotes, a 7TM fold appears in high-conductance  $\text{Ca}^{2+}$ -activated potassium channels (BKCa) [34] and in ligand-gated ion channels [317] which do not have any evolutionary relationship with GPCRs. Particularly, in prokaryotes,

light-induced structural changes in the 7TM core of bacteriorhodopsin [318,319] or halorhodopsin [320] provoke pumping of protons (hydrogen ions) or halide ions, respectively, through the membrane, or induce protein-protein interactions which initiate a signaling cascade associated with phototaxis, as in the case of the sensory rhodopsin I-transducer complex [319].

While the 7TM core is a typical common feature of GPCRs, the extracellular and intracellular regions differ in structure, sequence and length allowing interactions with various signaling molecules and ligands: ions, organic odorants, amines, peptides, proteins, lipids, nucleotides and photons [309]. Moreover, many GPCRs, so-called orphan receptors, still lack a reliable assignment of interacting ligands and some of them may not even need ligands for activation but, most probably, are self-activated through heterodimerization [321]. In general, ligands bind to the extracellular loops and the N-terminus, while the intracellular part of GPCRs is involved in protein-protein interactions with G proteins, arrestin or other subunits.

Depending on the GPCR family different regions of receptors are involved in the activation process (see Table 1). The common role of GPCRs is a signal transmission to the cell interior through interactions with molecules, such as the G protein or arrestin, by changing the structure of their transmembrane domains and/or extracellular and intracellular parts after the ligand binding. G protein-coupled signal transduction involves dissociation of G protein into  $G\alpha$  and  $G\beta\gamma$  subunits which modulate enzymes or membrane channels leading to a highly amplified signaling cascade. In the absence of any ligand a G-protein coupled receptor is believed to be in a dynamical equilibrium between the inactive (R) and the less populated active (R\*) state. Binding of an agonist molecule (full or partial) is thought to increase the probability of the receptor converting to R\* [322]. Such a scheme is preferred in the case of GPCRs with diffusible ligands, however, in the case of rhodopsin, endowed with a very tight ligand binding site, probably the induced fit mechanism is employed [323]. It is possible that in most GPCRs both mechanisms are operational but in different proportions. These two types of activation paths, the dynamic equilibrium of receptor states and the agonist-induced conformational change, will be described in detail later on. Antagonists prevent the binding of both agonists and inverse agonists into the orthosteric site (a binding site for endogenous agonists) but they can also change the receptor structure (or choose a particular state of the receptor according to ensemble of conformational receptor states) which can even induce receptor internalization in some cases. GPCRs activation and signal transmission can also be influenced by allosteric or ago-allosteric modulations induced not only by several ligands known to date [324] but also through either negative or positive cooperation between protomers within a dimer. Many functional studies proved dimerization of GPCRs [325] although even a monomer is able to activate its G protein [326], to bind arrestin or to undergo phosphorylation catalyzed by GRKs (GPCR

kinases) [327]. The role of oligomerization in activation and signal transduction by GPCRs is still not clear although some experimental and theoretical studies involving not only Rhodopsin but also the Glutamate and Secretin families proved its relevance [328,329]. Moreover, many GPCRs, so-called orphan receptors, still lack reliable assignment of interacting ligands and some of them are probably self-activated i.a. through heterodimerization [321].

GPCRs were traditionally considered to be monomeric and recent studies of GPCRs reconstituted in high-density lipoprotein particles have confirmed that these receptors can exist and function as monomers [326,330]. However, as evidenced by biochemical measurements of cooperativity, biophysical determinations of fluorescence or by bioluminescence resonance energy transfer between protomers, co-immunoprecipitation and other methods GPCRs from various families can assemble as dimers or higher-order oligomers [331–333]. Currently, dimerization was proposed to play a role in processes ranging from ligand binding to receptor signaling, maturation, trafficking and regulation. For the glutamate receptor family of GPCRs activation involves a movement of the N-terminal Venus flytrap domain (VFTD) within a dimeric GPCR entity to activate the membranous domains, which suggests that dimerization is mandatory for agonist-induced activation [334]. It is likely that various allosteric interactions between monomers in an oligomeric complex represent those that occur between distinct sites within a given GPCR monomer [335]. To date, experimental data suggest that GPCR dimers and oligomers are functionally asymmetric which was characterized especially for GABABR and the mGluR receptors from the glutamate family [336,337]. Because of asymmetric functionality the subunits in a GPCR dimer possibly adopt different conformations in a particular receptor state (inactive or active). Recent studies on dopamine D2 receptor dimers have also demonstrated asymmetric communication between an agonist-bound and an antagonist-bound protomers within the D2 dimer [333].

Homo- and hetero-dimerization can modulate the signaling properties of receptors and mediate cross-talk between GPCR pathways [338]. Crystal structures can also directly suggest novel allosteric sites with specific properties and selectivity. For example, a cholesterol binding site located in the interface between protomers consisting of helices TM1 and H8, has been observed in many dimeric  $\beta$ 2-adrenergic receptor crystal structures [339]. Cholesterol can modulate receptor thermostability and ligand affinities. However, even for the most studied dimers the identification of the functionally relevant interface is still very difficult. This is partly due to the transient mode of the interactions and the technical problems of differentiating between specific and nonspecific binding in membrane environment [340]. On the other hand, it was observed in crystallization studies that nonspecific or partial dimerization of GPCRs can prevent crystal formation because it introduces heterogeneity to the system, so it is usually avoided. Therefore, in most crystal structures of GPCRs

analyzed so far the receptor molecules have been found in non-functional (antiparallel or tilting) orientations. The recent crystal structures of CXCR4 [316] are rather exceptions revealing a parallel dimer arrangement involving helices TM5 and TM6. The dimer interface is virtually identical in five different crystal packing forms of CXCR4 with both peptide and small-molecule antagonist, which can suggest that it is functionally relevant [341].

Using a fluorescence resonance energy transfer (FRET) to characterize the oligomerization of  $\beta$ 2AR Fung et al. [332] proposed that  $\beta$ 2AR forms predominantly tetramers when reconstituted in phospholipid vesicles. Agonists and antagonists had little effect on the relative orientation of protomers in the oligomeric complexes so it was suggested that the tetramer structure is loose enough to accommodate a large, outward movement of the cytoplasmic part of TM6. In contrast, binding of inverse agonists led to a significant increase in FRET efficiencies for most labeled amino acid pairs. This could suggest that inverse agonists can induce tighter packing of protomers and/or stimulate the formation of larger oligomers (possibly octamers or larger structures) by employing the additional interface at the receptor surface. The results provide new insights into  $\beta$ 2AR oligomerization and reveal a possible mechanism for the functional effects of inverse agonists. The interface involving helices TM1 and H8 was proposed for a symmetrical dimer (as it was found in the crystal structures of this receptor type) so the tetramer would be a dimer of dimers. Upon binding of an inverse agonist the dimers could form a tighter structure and additionally the tetramers could stick together to form octamer engaging an interface involving helices TM4 and TM5. In this way a larger oligomer can be formed using more tetramers. Similar interfaces were proposed for rhodopsin oligomers based on the AFM measurement [342,343], however in this case the interface in a rhodopsin dimer involved helices TM4 and TM5 while contacts between rows of dimers were maintained by helices TM1 from different protomers.

**Table 1.** Characterization of GPCRs families. Important motifs are bolded and residues involved in molecular switches are underlined. Uppercase letters in the motifs indicate completely conserved positions, lowercase indicate well-conserved positions (>50%) and x indicates any amino acid. The residues are numbered according to the Ballersteros-Weinstein numbering scheme [90]

Family	Orthosteric binding site	Overall sequence diversity	Well-conserved motifs/residues and molecular switches based on [21, 69, 127]										C-terminus & Helix 8		
			No. of conserved res. in TM1 region	N-terminus	TM1	TM2	EC1	TM3	TM4	EC2	TM5	TM6		TM7	
Rhodopsin	TM region, EC loops, less frequently: N-terminal region (LRR, LDLa, PAR)	high	25	Cys	Asn1.50	Asp2.50 Pro2.59 (Pro2.58) <sup>a</sup>	-	Cys3.25 Glu3.28 <sup>b</sup> Asp3.32 Glu/Gln3.37 Leu3.40 Asp3.49 (Glu3.49) Arg3.50 Tyr3.51 Tyr3.60 (d/e)Ry	Trp4.50	Cys	Phe5.47 (Asn/Asp5.47) Pro5.50 Tyr5.58	Glu6.30 Thr6.34 Phe6.44 Cys6.47 Trp6.48 Pro6.50 CwxP	Lys7.43 (Tyr7.43) Asn7.45 Ser7.46 Asn7.49 Pro7.50 Tyr7.53 nPxxy	Phe7.60	
Glutamate	N-terminal region (VFTM, SUSHI)	low	94	Cys	Aliphatic, aromatic, polar	Aliphatic, aromatic, polar	-	Cys	Cys	Aliphatic, aromatic, polar, charged(+)	Cys	Aliphatic, aromatic, polar, charged(+)	Cys	Aliphatic, aromatic, polar, charged(+)	-
Adhesion	N-terminal region (i.a. GPS, HBD)	high	6	Cys exCxHl/s	-	Polar, charged(+)	Cys	Cys	Aliphatic	Cys	-	Aliphatic, polar	Cys	-	-
Secretin	N-terminal region, EC loops, TM6	high	33	Cys CxxxDxxxx CW rxCxxxGxw	Aliphatic, polar	Aromatic, aliphatic, charged(+)	Cys	Aliphatic, aromatic, charged(-)	Aromatic, aliphatic, charged(+)	Cys	Aliphatic, polar	Aliphatic	Aromatic, aliphatic, polar	-	-
Frizzled/Taste2	N-terminal region (Wnt binding domain <sup>c</sup> ), EC loops	low/ high <sup>e</sup>	81 in Taste2: 103	Cys	Aliphatic, aromatic, charged(-) <sup>f</sup>	Aliphatic, aromatic, charged(-) <sup>f</sup>	Cys T/M, I/V <sup>f</sup> NxWaVtnH <sup>f</sup>	Aliphatic, aromatic, charged(-)	Aliphatic, aromatic, charged(+)	Cys	Cys	Aliphatic, aromatic, charged(+)	Aliphatic, aromatic, polar	Aliphatic, aromatic, charged(+)	L/R

<sup>a</sup>Only in the G1 group according to Chabbert's classification [32].

<sup>b</sup>Only in the Rhodopsin PDB structure.

<sup>c</sup>Only in LGR receptors (Leucine-rich repeat-containing GPCRs). LGR receptors are members of the G3 group according to [32].

<sup>d</sup>Not in Taste2 receptors.

<sup>e</sup>Low – Frizzled, high – only in the N-terminal region of Taste2 receptors.

<sup>f</sup>Only in Taste2.



## 10.2 Switches in rhodopsin-like receptors

The activation of GPCRs occurs most probably through series of conformational changes called molecular switches. The crystal structures enables the researchers to almost see them in action by comparing structures of the same receptors with agonists and antagonists (Fig 10.3). Based on the crystal structures we describe those molecular switches that are well characterized and proposed to work in most of GPCRs.

### 10.2.1 The Ionic Lock Switch

The presence of the first switch, the ionic lock, has been shown in the first GPCR X-ray structure obtained by Palczewski et al. [297]. The inactive state of bovine rhodopsin shows a strong intramolecular interaction between residues Glu3.49/Arg3.50 of the conserved (d/e)Ry motif in TM3 and residues Glu6.30/Thr6.34 in TM6 (Fig 10.3G). The authors of that paper concluded that "it could be one of the critical constraints keeping rhodopsin in the inactive occupation", but also noted that this region has high crystallographic B-values, meaning that the side chains may assume different conformations. Following that work as well as mutagenesis studies, which showed the importance of the (d/e)Ry motif [344], the activation mechanism of GPCRs has been described as a cascade of altering molecular switches in conserved microdomains [303,345]. In this mechanism, ligand binding triggers a series of molecular switches (including the TM3-TM6 ionic lock) to unlock the G protein-binding site in the intracellular face of the receptor, leading to G protein activation.

Apart from inactive rhodopsin there are only few crystal structures in which this particular ionic lock is observed. The dopamine D3 receptor [346] and adenosine A2A receptor [347] (but only with selected antagonists) are the only other GPCRs that show the Arg3.50-Glu6.30 ionic lock in the crystal structure. In addition, the residues Asp3.49 and Arg3.50 are forming hydrogen bonds with Tyr3.60 (located in IC2), possibly stabilizing the ionic lock and restraining a helical conformation of IC2. The turkey  $\beta$ 1-AR structure has the ionic lock open but because of the close proximity of helices TM3 and TM6, the different rotamers of these residues would yield the switch closed. The AA2R-T4L chimera has a similar structure in this region and the (d/e)Ry motif is forming a Asp3.49-Tyr3.60 hydrogen bond which restrains the conformation of intracellular loop 2 (IC2). In the human  $\beta$ 2AR structure the ionic lock between Arg3.50 and Glu6.30 is absent; instead, a hydrogen bond between the highly conserved Tyr3.60 and His6.31 is present. CXCR4 is lacking the Glu6.30 residue and no ionic lock is present between TM3 and TM6. In the histamine H1 receptor the ionic lock is also absent; instead, Arg3.50 adopts a different conformer and forms a hydrogen bond with Gln6.36 in TM6 which can also bridge helices TM3 and TM6 to some extent. The lack of the ionic

lock despite the presence of residues capable of forming strong interactions has been intriguing; some attribute it to the inclusion of the T4L fusion protein in crystal structures, which may affect the interactions in TM6.

### 10.2.2 The 3-7 Lock Switch

In rhodopsin the key restraint which is broken first upon retinal isomerization is a salt bridge between a protonated Schiff-base of retinal-Lys7.43 and a counterion, Glu3.28 (Fig 10.3J). This switch is called the 3-7 lock because a link between TM3 and TM7 is broken during activation. A similar mechanism probably exists in other receptors, especially with amine-type ligands (aminergic receptors) e.g. histamine H1 [348] or dopamine D3 [346], which were crystallized with antagonists bound and also in opioid receptors (OPR) for which an extensive modeling was done [349–351]. In these receptors the switch is composed of different residues: Tyr7.43 (which is more conserved than lysine present in rhodopsin) and Asp3.32 which substitutes for the rhodopsin's counterion, Glu3.28. Asp3.32 is located on the same face of TM3 and deeper in the receptor interior which compensates for a shorter length of its side chain. In  $\beta$ 1- and  $\beta$ 2-adrenergic receptors there is also a hydrogen bond, Asp3.32-Tyr7.43, but additionally, the Asn7.39 residue, positioned one turn of helix away of Tyr7.43, is linked to Asp3.32 by a protonated nitrogen atom of aminergic receptor ligands. This is why disruption of the Asp3.32-Tyr7.43 hydrogen bond does not break the link between TM3 and TM7 so the 3-7 lock switch is not functioning in adrenergic receptors. Opening of the 3-7 lock was suggested by Khorana [352] to be the first switch activated in rhodopsin and possibly it is one of the first switches that can be activated upon ligand binding in some other GPCRs. In (Fig 10.3) it is represented by one panel with rhodopsin structures.

### 10.2.3 Transmission Switch (Former Trp Rotamer Toggle Switch)

In all crystal structures with agonists there are movements of TM5 and TM6 but they vary considerably. Several similarities can be observed including a relocation of conserved residues Trp6.48 and Phe6.44 towards Pro5.50 (Fig 10.3A-C). Such movements were called “a transmission switch” by Deupi and Standfuss [307] instead of the previous name “rotamer toggle switch”. This novel and larger switch links the agonist binding site with the movement of TM5 and TM6 through rearrangement of the TM3-5-6 interface. This is possibly the most common switch among GPCRs. After movement of Trp6.48 in rhodopsin the Phe6.44 residue situated one helix turn away toward the cytoplasmic side of TM6, is displaced toward Leu5.51 as the whole TM6 is rotating horizontally. In  $\beta$ -adrenergic receptors a little contraction of the binding site is observed while in rhodopsin isomerization of retinal makes the binding site much larger. The interaction of Ser5.42 and Ser5.46 with agonists stabilizes the receptor conformation which leads to a 2.1 Å movement of TM5 and

about a 1.4 Å movement of Pro5.50 whereas, unlike in rhodopsin, there is no movement of Trp6.48. However, there is a rotation and movement but only of the cytoplasmic part of TM6. This is translated to a relocation of Ile3.40 from its position at Pro5.50 and a motion of Phe6.44 due to rotation of TM6. Activation of this switch seems to be limited to some classes of GPCRs. Apart from these differences the activation mechanism of A2AR is similar to that of rhodopsin because Trp 6.48 is also moved and TM6 rotated. There are also similar rearrangements in β2AR and rhodopsin in TM5 and TM6, by movements of Phe6.44 towards Pro5.50 and Leu5.51 together with the movement of Ile3.40 away from Pro5.50 – such translocations are part of the transmission switch. Agonist binding in A2AR results in the relocation of Ser7.42 and His7.43 which, together with Thr3.36 in TM3, coordinate a part of the agonist. These interactions resemble the 3-7 lock between the protonated Schiff base of the retinal-Lys7.43 and Glu3.28, which is critical for rhodopsin activation. Although the TM3-agonist-TM7 interactions in the adenosine receptor are formed, rather than broken, upon activation, they could fulfill a similar role in arranging TM3 and TM7 in the active and inactive conformations [307].

The switches together with the hydrogen bond network between conserved residues, motifs and structural water molecules constitute an extended interface between different areas in GPCRs which facilitate the large movements linking ligand binding to cell signaling. Based on the recent crystal structures of inactive and activated, as well as constitutively active rhodopsin, one can elucidate the activation scheme of this protein and the role of particular switches as it was done by Standfuss et al. [307]. The structure of retinal in all-trans conformation, but unbound from Lys7.43, represents the active structure of rhodopsin nearly identical to the Meta-II state. This structure was published nearly simultaneously with the Meta-II rhodopsin structure with covalently bound retinal with and without GαCT (C-terminus of Gα subunit) [353]. The structures agree with each other in location of the main conserved amino acids. A covalently bound all-trans retinal behaves as a full agonist, whereas when unbound, it behaves as a partial agonist but maintains the critical interactions between the β-ionone ring and helices TM5 and TM6. The structure of a constitutively active mutant Glu3.28Gln [354] represents probably a trapped intermediate when retinal is either entering or exiting the binding site. The transition from an inactive to active state of GPCR includes large rigid motion of TM6. In the case of rhodopsin this is not a vertical hinge movement (named a global toggle switch) but a horizontal (in plane of the membrane) rotation of TM6 that leaves the shape of the helix intact. The characteristic bend of TM6 is imposed by Pro6.50 which is a part of the CwxP motif. The other highly conserved amino acid, Trp6.48, is tightly packed against retinal in the ground state of rhodopsin as it has a central role in the transmission switch model (previously called a rotamer toggle switch) of activation of these receptor. The structure of the Glu3.28Gln mutant places this residue 3.6 Å from its ground state position. However, no rotamer change is observed as it was proposed based

on biochemical experiments and also computer simulations. Instead, Trp6.48 follows the retinal (its  $\beta$ -ionone ring) maintaining contact with the C18 methyl group.

#### 10.2.4 Tyrosine Toggle Switch (nPxxxy Motif)

A region called the hydrophobic barrier (Fig 10.4A) separates the water mediated hydrogen bond network from the (d/e)Ry motif which is critical for G protein activation (Fig 10.3D-F). In the active Glu3.28Gln-G $\alpha$ CT structure, a rotation of TM6 disrupts the water mediated link between Trp6.48 and Ser7.45 and reorganizes the ground-state hydrogen bond network. The hydrophobic barrier opens and Tyr7.53 of the nPxxxy motif, together with Tyr5.58, rearrange to fill the hydrophobic gap and to extend the hydrogen bond network towards the (d/e)Ry motif and G $\alpha$ CT peptide (Fig 10.4B). The role of this barrier in molecular switching was explained by Standfuss et al. [354] based on studies involving the crystal structure of rhodopsin with all-trans retinal in the binding site. The hydrophobic barrier was described earlier also by Schertler's group [355] upon crystallization of inactive rhodopsin in a trigonal crystal form. A more extended motif nPxxxy(x)5,6F was proposed by Fritze et al. [356] to explain the presence of the interaction between Tyr7.53 in TM7 and Phe7.60 in helix H8 in the inactive structure of rhodopsin (Fig 10.3D). Such an interaction can additionally stabilize the inactive state of the receptor. However, in crystal structures of other GPCRs, such as the adrenergic and adenosine receptors (Fig 10.3E-F) such interaction is not seen despite the fact that Phe7.60 is present. It probably indicates that these receptors could be partially activated.

The hydrophobic barrier consists of six residues between helices TM2, TM3 and TM6 (Leu2.43, Leu2.46, Leu3.43, Leu3.46, Met6.36 and Met6.40) and many of them are conserved in the rhodopsin-class of GPCRs. The rearrangement of hydrogen bonds is relatively minor but they directly link changes in the CwxP motif in the retinal binding pocket with the most conserved residues in TM1 (Asn1.50) and TM2 (Asp2.50) and nPxxxy in TM7. On the cytoplasmic side, the rotation of TM6 opens the hydrophobic barrier allowing Tyr5.58 and Tyr5.53 to swing into the protein interior. They provide additional interactions with water molecules extending the hydrogen bond network toward the hydrophobic barrier to the (d/e)Ry motif at the cytoplasmic surface of TM3. The ionic lock involving residues in this motif, Glu3.49-Arg3.50 and Glu6.30, is broken and allows binding of G $\alpha$ CT peptide in a position that is occupied by TM6 in a ground state. Thus, rotation of TM6 and displacement of Trp6.48 results in a hydrogen bond network connecting residues from the retinal binding site to those at a cytoplasmic surface critical for activation of G protein.

Also the recently obtained A2AR-T4-lysosyme structure exhibits features of agonist induced rearrangements of a cluster of hydrophobic residues in TM3-5-6 near the binding site (Ile3.40,

Leu5.51, Phe6.44 and Trp6.48) similarly to the active structures of  $\beta$ 2AR [301,357] and rhodopsin [353,354]). However, although Tyr7.53 from the nPxy motif is relocated towards the receptor center the relocation of TM6 is only 3Å which is much smaller than in active rhodopsin (6 Å), nanobody  $\beta$ 2AR (8 Å) and in the complex with trimeric G protein (14 Å). These changes may be blocked by the presence of the fused lysozyme structure. However, the changes of residues close to the binding site suggest that this conformation resembles the Meta-I structure of rhodopsin which does not allow binding to G protein. Possibly in some GPCRs the full adaptation to G protein binding may be achieved in the presence of a G protein or other interaction partners that stabilize the cytoplasmic domain.

### 10.2.5 The Elusive “Global Toggle Switch”

The number of conserved motifs found in transmembrane helices of the Rhodopsin family receptors is significantly higher than in the other GPCR families (see Table 1 and Fig 10.2) indicating their potential role in receptor stabilization and activation. It was proposed that receptors of the Rhodopsin family most probably share the common mechanism of activation - the so-called "global toggle switch" [358]. According to this model the TM6 helix performs a vertical see-saw move around the central Pro6.50 residue during receptor activation induced by binding of an agonist. The upper part of TM6 is closing around the ligand, while the lower (near the intracellular surface) is opening to prepare for the G-protein binding. It was suggested that during activation by an agonist the rearrangement of TM3 and TM5 also takes place, though to a minor extent than in the case of the TM6 movement [307]. An accompanying kink in TM7 is induced by changes in the hydrogen-bond network between TM7 and TM1, TM2 and TM6 [304].

Using the metal ion site engineering techniques and based on the obtained distance constraints for  $\beta$ 2-adrenergic receptor Elling et al. [359] developed the so-called “global toggle switch” mechanistic model. In this model Asp3.32 was an anchoring point for monoamine binding in TM3 helix. The authors engineered metal ion sites, which activated the receptor, between the extracellular parts of TMs. Copper and zinc ions alone and in complex with aromatic chelators acted as potent agonists in sites constructed between position 3.32 (Asp - known to bind ligand directly - or its mutation to His) and the Cys or His mutations of specific amino acids at TM6 and TM7 close to the binding site. To fulfill the distance constraints the residues involved in the orthosteric ligand binding pocket had to move closer to each other during receptor activation. In this model an inward movement of the extracellular segments, especially those of TM6 and, to some extent, TM7, was coupled to the well-established outward movement of the cytoplasmic segments of these helices. The authors suggested that the pivot points for these vertical seesaw movements are the highly conserved proline bends of the involved helices. Based on the present crystal structures of  $\beta$ 2AR the global toggle switch must be modified because only a slight inward motion of the extracellular part of TM6 was

detected. In rhodopsin there is even an increase of the retinal binding site and the same is in the case of A2AR – the binding site is smaller for antagonists regardless if they are smaller (caffeine) or bigger (ZM241385) than the agonist (adenosine). TM3, and not extracellular part of TM6 which is not moving, is responsible for this shrinkage of the binding site. There is an unusual bend on TM3 (close to Val3.32) in receptor structures with bound antagonists. This residue (located in the same position as extremely important Asp3.32 in receptors for monoamine ligands) may be now regarded as a part of the agonist/antagonist sensor.

### 10.2.6 Role of Conserved Residues

Rhodopsin-like GPCRs lack a long N-termini except for PARs (protease-activated receptors which do not need agonist-binding to be activated) with an N-terminal thrombin-cleaved part releasing a tethered ligand, LGRs (GPCRs containing LRRs – leucine-rich repeats) interacting with glycoproteins and LDLa (a low-density lipoprotein receptor class A). In most Rhodopsin-like GPCRs an agonist interacts with extracellular loops and the TM region. Although sequence diversity in the TM region is quite high even within the Rhodopsin family the motifs involved in the activation mechanism are well conserved, i.e. (d/e)Ry, CwxP and nPxxxy. In Table 1 we indicated all conserved residues in the Rhodopsin family of receptors and underlined these which are believed to be involved in molecular switches. The residues are also visualized on the topological scheme of GPCR (7TM receptor) in (Fig 10.2).

In TM1 the most conserved residue is Asn1.50, involved in a structural water-mediated hydrogen-bonding network between TM1, TM2 (Asp2.50), TM6 (Trp6.48) and TM7 (Asn7.45, Ser7.46, Asn7.49, Tyr7.53). The Asn 1.50 residue is an arguable element of the receptor activation process, namely the TM3, TM5 and TM7 movements [304]. The conserved proline residue in TM2 [314], either in position 2.58 (e.g. in a recently solved CXCR4 structure) or 2.59 (rhodopsin and adenosine receptors), which induces a helix kink in the first case or a helix bulge in the latter, is crucial for the ligand binding, but does not play a significant role in the receptor activation. As in most of GPCRs cysteine residues are highly conserved in the Rhodopsin family and form disulphide bridges stabilizing the receptor structure. The most important cysteine pair is Cys3.25 connected with Cys in EC2. Glu3.28, which is present only in the Rhodopsin PDB structure, serves as a counterion with the protonated Schiff base in 11-cis-retinal and possibly stabilizes an inactive state of opsin [360]. Asp3.32 with Trp7.40 and Tyr7.43 (instead of Lys7.43, more frequent in the Rhodopsin family – see Table1) are a unique fingerprint only for biogenic and trace amine receptors (a G2 group) not shared by any other Rhodopsin-like GPCR. Asp3.32 with Tyr7.43 were proved to form the TM3 – TM7 ionic lock stabilizing the unbound, inactive state of the receptor [361]. Asp3.32 is believed to serve as a counterion for the amine groups of native ligands [362]. A residue in the 3.36 position is not well-conserved in Rhodopsin-like GPCRs, however, it was confirmed experimentally by

site-directed mutagenesis, that this residue interacts with Trp6.48 and stabilizes the inactive state not only in the  $\beta$ 2-adrenergic receptor (Val3.36, van der Waals interactions) [266] but also in serotonin receptors (Ser/Cys/Thr3.36, hydrogen-bonding) [361], the opsin subclass (Gly3.36) [363] and cannabinoid receptors (Phe3.36 aromatic stacking with Trp6.48 – a rotamer toggle switch) [294,364]. Glu/Gln3.37 is a key residue in agonist binding to LH and TSH receptors [365]. A well-conserved Leu3.40 residue which is close to Pro5.50 before activation and becomes distant after, plays a key role in the TM3 – TM5 movement [307].

A well-conserved Trp4.50 is a cholesterol binding-site which is visible in the structure of human  $\beta$ 2-adrenergic receptor [339]. Phe5.47 stacks against Phe6.52 and possibly interacts with agonists [366] but still little is known about its function. Phe6.44 together with Phe6.52, Leu3.40 and Leu5.51 is forming a hydrophobic and aromatic cluster around Trp6.48 involved in conformational rearrangements of TM5 [307,361]. Pro6.50, like Pro7.50, produces a helix kink around which TM6 performs movements during activation. Tyr7.53 in the nPxy motif interacts with Phe7.60 in helix H8 and forms a molecular switch between active and inactive conformation.

### 10.2.7 Role of Extracellular Loops in Ligand Binding and Switching

The extracellular loops also have an influence, although sometimes transiently, on ligand binding and could participate in some types of molecular switches. The recent crystal structures of GPCRs revealed that the part of the receptor extending from the orthosteric ligand-binding site in the transmembrane domain to the cytoplasmic side is highly structurally conserved. In contrast, the extracellular surface of GPCRs is substantially diverse and, therefore, could be a target of highly selective drugs. However, still little is known about the coupling of the extracellular surface to the ligand-binding compartment. Bokoch et al. [367] used NMR spectroscopy to investigate ligand-specific conformational changes around a salt bridge linking extracellular loops EC2 and EC3 (Asp192-Lys305) in  $\beta$ 2AR. It was demonstrated that small-molecule hydrophilic drugs that bind within the transmembrane core and exhibit different efficacies towards G-protein activation (agonist – formoterol, neutral antagonist – alprenolol or unliganded receptor, and inverse agonist - carazolol) also stabilize distinct conformations of the extracellular surface. Such conformational coupling supports the possibility of an efficient allosteric action of specific drugs targeting this diverse surface with high subtype selectivity. Although the specific salt bridge used to monitor these conformations may not be present in other GPCRs it is likely that ligand-induced changes at the extracellular surface are relevant for other family A GPCRs.

In adrenergic receptors only one residue in the EC2 loop can interact with ligands in the binding site: this is Phe201 in  $\beta$ 1AR and an equivalent residue, Phe193, in  $\beta$ 2AR. A disulphide bridge located two residues away from that phenylalanine residue keeps the proper conformation of EC2 and assures

that such interactions with the ligand will be preserved. In the recent crystal structures of these receptors with agonists the ligands do not appear to interact with this residue, however, if we examine the possible entrance way of the ligand into the receptor binding site we can notice that the ligand may interact transiently with Phe193/201. Therefore, it is possible that Phe 193/201, together with other residues from the extracellular loops, could participate in the action of molecular switches. The binding sites of other receptors with diffusible ligands are more spacious so the binding of a ligand is more straightforward and could be done without a transient binding to the EC2 loop. However, even in those receptors the ligands can interact with the EC2 loop. An interesting case is a recent crystal structure of a chemokine receptor CXCR4 with a peptide ligand CVX15 consisting of 16 amino acids [316]. Binding of that ligand involves a large number of residues from the EC2 loop and also from the N-terminus. However, because of the lack of structures of CXCR4 with agonists, there is no direct data on the involvement of extracellular surface residues in molecular switching. Even in the case of rhodopsin the EC2 loop, which tightly covers the retinal binding site, is moving upon retinal isomerization and this movement, from the retinal-binding site, is coupled to the rotation of TM5 and to the inward motion of the TM6-EC3-TM7 segment [368].

The hydrophobic ligands, like retinal in the case of rhodopsin, are probably entering the receptor binding site directly from the membrane. There are two openings of the retinal-binding site in the crystal structure of opsin (ligand-free rhodopsin) [369] one between the extracellular ends of TM5 and TM6, and another between TM1 and TM7. It was suggested that the opening between TM5 and TM6 could be selective for the uptake of 11-cis-retinal. The smaller opening between TM1 and TM7 could be a site for the release of all-trans-retinal. A putative external binding site for retinal is possibly located in the kink region between TM7 and H8 closely to palmitoyl chains [370]. The mechanism of retinal movement is potentially significant for vision in the regeneration pathway, the disorders of which have been associated with different forms of blindness. In the recent structure of CXCR4 there is also a gap between EC ends of helices TM5 and TM6 which is filled up by lipids. The hydrophobic ligands of this receptor could potentially enter the receptor binding site through this hole. However, the open question remains which residues could be responsible for sensing the ligand type and which ones participate in switching mechanisms.

### 10.3 Activation Schemes

The recent period proved to be very fruitful in GPCR research – many new structures were crystallized and, what is even more important, first time with agonists ( $\beta$ 1- and  $\beta$ 2-adrenergic and adenosine receptors and recently also rhodopsin with all-trans retinal). This greatly facilitated elucidation of the activation scheme of these receptors. Now, another breakthrough has been made



i.e., the long awaited crystal structure of the complex of GPCR with the whole G protein is available and one can expect that similar structures of other GPCRs will be also available. Possibly, a new and exciting mystery to be solved is the allosteric influence of dimers on the process of activation. In a very interesting review compiled by Deupi and Kobilka about the energy landscapes of GPCR activation it is shown how structural changes of GPCRs during their activation can be visualized on energy landscapes. Because of high structural similarity of all crystallized GPCRs, the activation scheme is probably similar for all GPCRs so it is suggested that the receptors are passing through the same stages during the activations process. This similarity is much higher in the cytoplasmic side of the transmembrane bundle. This region contains residues involved in receptor activation and binding of a G protein. Similar conformational changes underlying activation of GPCRs are also deduced from numerous biochemical and biophysical experiments. Probably also a sequence of events is nearly identical and involves the following steps: first, small changes in TM5 and TM7, then a large change in TM6, and then neutralization of Asp3.49 in the (d/e)Ry motif (Fig 10.4). Two-dimensional energy landscapes seem to be more advantageous over one-dimensional energy plots but currently too little is known for precise construction of such plots. Possibly new crystal structures supplemented by long molecular dynamics simulations will help in designing so useful but also elegant and eye-catching charts. 2D or even 3D energy plots make possible dissection of the reaction pathway into discrete non-sequential conformational changes providing alternate routes of activation through the energy landscape. In this way some events may be skipped for some ligands and a full or partial activation state can still be achieved.

### 10.3.1 Two Types of Activation Paths

The substantial amount of data obtained from rhodopsin and also adrenergic receptor activation can serve as a framework to reveal activation of other GPCRs. There are some variations, though. It is suggested [42] that the  $\beta$ 2AR is not trapped in a fully inactive conformation in the absence of agonist but its internal flexibility allows the receptor to explore different conformations. This may suggest a shallow energy landscape with several conformational states separated by relatively low energy barriers. On the contrary, for rhodopsin (and similarly activated GPCRs) it is supposed that binding of agonist is invoking an induced fit of the receptor structure. Therefore, agonists have to bind to the receptor with high affinity and this high binding energy is used to initiate conformational changes (“jump”) over the highest initial barrier of energy. Retinal isomerization in rhodopsin provides such high energy. In case of other receptors (although being classified in the Rhodopsin-like family because of their sequence) the ligands have relatively low affinity and rapid dissociation rates; these features may indicate a conformational selection procedure of activation. After ligand binding the sequence of events during receptor activation is regarded as being similar in all these receptors. Any differences are rather not in a number of steps required for full activation but rather in the size of

energy minima depths. The well-known example of differences in the activation scheme is the existence of an open ionic lock in crystal structures of  $\beta$ 1AR and  $\beta$ 2AR adrenergic receptors with antagonists and inverse agonists bound that may suggest a very low energy barrier for opening of this switch. The late stages of  $\beta$ 2AR activation, which are supposed to be analogous to achieving the Meta-II stage in rhodopsin activation, involve a similar set of conformational changes, i.e. rearrangement of TM6 and neutralization of Asp3.49 in the (d/e)Ry motif of TM3. According to the above two schemes of activation, the partial agonism can be also explained in two ways. For those GPCRs from which partial agonists dissociate faster than full agonists, not all binding events last long enough to promote activation of the G protein. Another possibility is that the partial agonists stabilize different intermediate conformations that lead to alternate activation pathways and to non-optimal G protein activation. Particular steps can be achieved either by induced fit upon binding of a ligand or by conformational selection but the achieved equilibrium states would be completely indistinguishable. It is suggested that the induced fit mechanism is present in rhodopsin and angiotensin AT1 receptor whereas  $\beta$ 2AR may function by selecting specific receptor substates by the ligand.

## 10.4 Theoretical Studies on the Action of Molecular Switches

### 10.4.1 Single TM Studies

One of the first computational studies aimed at GPCRs switches was done in 2001 by Ballesteros *et al.* [344] who simulated the disruption of the TM3-TM6 ionic lock. Simulations presented in that work concerned only TM6 and short MD runs to simulate the bending of TM6 at the Pro6.50 position were performed. From those computational studies combined with experimental mutations of the Glu6.30 residue it was concluded that a conformational rearrangement of TM6 is highly correlated with the extent of constitutive activity of different mutants. A similar approach was used later to study the conformational switch in the 5-HT<sub>2C</sub> receptor [371]. Again, a combined computational-experimental study showed that a conserved Tyr7.53 residue is interacting with the conserved Tyr7.60 (in helix 8) contributing to the switching of the receptor among multiple active and inactive conformations. Although the 'ionic lock' is still regarded as an important switch it can be open in crystal structures of GPCRs even with antagonists and inverse agonists. Currently, only in the inactive and partially active rhodopsin structure (batho and lumi intermediates) that switch is closed [372].

A similar approach was presented in the 2002 paper by Shi *et al.* [373]. In this work Monte Carlo techniques were used to sample rotamer changes among the X6.47- Trp6.48-Phe6.52 residues of the

human  $\beta_2$ AR model of TM6. The results show a high correlation between the conformation of side chains of these residues and the helix kink at the Pro6.50 position, which was consistent with the experimental data for rhodopsin [374]. While it was clear that simulations on isolated helices could not predict the global interaction and changes in GPCRs, the results showed the usefulness of computational methods for studying ionic locks.

The same ionic lock has been studied in the 5-HT<sub>2A</sub> system in the 2002 paper by Visiers *et al.* [345]. Here, I focus on the electrostatic properties of the conserved residues. By solving the Poisson-Boltzmann equation to obtain electrostatic potentials of the different conformers of important residues on the TM3-TM6 model, it was found that Glu3.49 may undergo protonation upon activation of the GPCR. The activation of the protein has been also explained as a change in the kink at Pro6.50 which allows the ends of TM3 and TM6 to move away from each other. Based on the computational results it was suggested that selected, single-point mutations (Glu6.30Asn, Glu6.30Gln, Glu6.30Leu) would disrupt the electrostatic interactions of the (d/e)Ry motif with this residue. This prediction was confirmed later by the results of site directed mutagenesis, where it was shown that a neutral residue at the 6.30 position increases the activity of 5-HT<sub>2A</sub> in the absence of the ligand, similarly to the human  $\beta_2$ AR case.

## 10.4.2 Studies on a Complete GPCR Model

### Investigations of the Ionic Lock Switch

The most prominent method to study the dynamics of GPCRs is nowadays Molecular Dynamics (MD) of the complete GPCR model. One of the first MD simulations focusing on TM3-TM6 ionic locks was performed in 2002 by Greasley *et al.* [375], who simulated the  $\beta_1$ AR model based on the rhodopsin structure of Palczewski. The methodology used the united atom model and included a large number of short (150 ps) MD runs of the protein only (without environment), using NOE constraints to preserve the  $\alpha$ -helix structure of TMs. The short time of simulations was due to the limited computational resources available at that time. The results showed a very high stability of the Arg3.50-Glu6.30 salt bridge. Combined with experimental mutational data (Glu6.30 mutations that weakened this ionic lock constitutively activated the receptor) the results showed that the transition from the inactive to active state of  $\alpha_{1b}$ AR involves a rearrangement of helices TM3 and TM6. The structure of  $\alpha_{1b}$ AR is still not available but predicted movements were confirmed by crystal structures of activated  $\beta_1$ - and  $\beta_2$ -adrenergic receptors.

In the same year Rohrig *et al.* presented their work in which a full-atom model of the bovine rhodopsin has been immersed in a layer of a n-octane mimicking lipid bilayer [376]. The goal of these simulations was to show the effect of retinal *cis-to-trans* isomerization on the dynamics of the protein. Indeed, it was found that retinal isomerization serves as a trigger for propagation of the signal to the surrounding helices. Despite the limited simulation time authors were able to see some rearrangements in TM6 and to a lesser extent also in TM4 and TM5. They have also noticed cleavage of selected hydrogen bonds in these helices, however the ionic lock remained stable over the course of simulations. In 2003 a team led by Thomas B. Woolf applied MD simulations to the full atom bovine rhodopsin model including DOPC lipid membrane and surrounding by water molecules [377]. Their 40 ns simulation (a huge amount in those days) became a basis for a number of similar MD simulations on GPCRs in the following years. Similarly to the Rohrig studies this work concentrated on the rhodopsin vicinity and the changes in residues interacting directly with the rhodopsin ligand. The authors noted a strong interaction between residues forming the ionic lock (Arg3.50-Glu6.30) throughout the whole simulation time ( Fig 10.3G).

A similar approach as in the Crozier work has been applied to a large number of GPCR MD studies, although the advances in computational powers allowed to lengthen the simulation scale to tens and hundreds of nanoseconds. One of the interesting works showing the stability of various locks was the MD investigation of the opioid receptors models. In the  $\mu$ -OPR opioid receptor there is no glutamic acid residue in the 6.30 position, but the TM3-TM6 lock is still present in the modeling studies due to the hydrogen bond between Arg3.50 and Thr6.34 [349,378]. In the most recent simulations this lock remains unbroken even in the presence of selected antagonists or agonists [350,351], most possibly because of a rather short length of simulations (nanosecond time scale) compared to the time needed for activation of the receptor (milliseconds). Binding of antagonists has also no effect on the Trp6.48 rotamer switch which remains in the initial, rhodopsin-like vertical position. Interestingly,  $\mu$ -OPR agonists toggle the Trp6.48 position to a horizontal one which, during simulations, forces the change in positions of aromatic residues around the highly conserved Pro6.50. It is also worth noticing that the mutation of Leu6.30Glu in the  $\mu$ -OPR system inactivates this receptor [379]. Simple calculations of the electrostatic interactions for two helices (TM3 and TM6) have shown that the Leu6.30Glu mutation enhances the hydrogen bond network around the mutated residue and stabilizes the inactive state.

### **Breaking of the 3-7 Lock Switch**

During the MD simulations of  $\mu$ -OPR it was noticed that another switch, the 3-7 lock, a link between TM3 and TM7 (a hydrogen bond Asp3.32-Tyr7.43), remains stable for the apoprotein simulations and in the presence of antagonists. The presence of agonists, however, forced the hydrogen bond to

break [349]. For one of the agonists it was also shown that rotation of Trp6.48 was linked to the cleavage of the Asp3.32-Tyr7.43 hydrogen bond and occurred within 1 ns after its breaking. Later, MD studies have shown that a similar cascade of events is likely to occur also upon agonist binding and activation of  $\delta$ -OPR and  $\kappa$ -OPR [350]. Recently, a structurally similar  $\kappa$ -OPR agonist and antagonist pair, guanidinonaltrindole (GNTI) compounds, were investigated by molecular dynamics simulations [351]. 5'-GNTI is an antagonist while 6'-GNTI is an agonist. Ligands were relaxed in the receptor binding site by the simulated annealing routine. During a series of MD simulations of the ligand-receptor complexes in DPPC membrane, the 3-7 lock was broken when the agonist was bound, but remained unbroken upon binding of the antagonist. The hypothesis of 3-7 lock breaking on the agonist binding still awaits confirmation, since no experimental structure of any opioid receptor is available at this time.

### **Beyond Classical MD Techniques**

In 2009 Provasi *et al.* studied the dynamics of the  $\delta$ -OPR system using a metadynamics approach [380]. This approach allows for an efficient exploration of multidimensional free energy surfaces of GPCRs (and other biological systems) by adding a history-dependent bias to the interaction potential of the system. The required microsecond-scale well-tempered metadynamics has been achieved using the united-atom model for lipids (DPPC and cholesterol molecules) and the full-atom model for protein and ligand. It allowed the authors to suggest a preferential entry pathway of the NLX antagonist, starting from the  $\delta$ -OPR surface and ending in the proper binding pocket of the GPCR, and to evaluate the bonding constants for the ligand. The observed binding pocket was extremely close to the previously predicted one and the starting structure for their system had all the known locks including the Arg3.50-Thr6.34 of TM3-TM6 lock (or '3-6 lock' instead of 'ionic lock' since no salt bridge can be formed in this case) and the Asp3.32-Tyr7.43 of 3-7 lock.

It is also worth mentioning that other computational techniques have been used to model the activation of rhodopsin and they gave similar results. In 2006 Niv *et al.* used an Elastic Network Model to study the inactive form of rhodopsin [381]. In this model, a harmonic potential with a single force constant accounts for pairwise interactions between all C $\alpha$  atoms within a certain cutoff distance. The analysis of the structural relation of the inactive normal modes to the transition vectors towards the active conformations has been discussed. It has been found that the active form of rhodopsin should be characterized by structural changes in TM5-6-7, while helices TM1-2-3-4 were shown to be the most stable ones which were confirmed later by crystal structures of Meta-II rhodopsin. Niv *et al.* also predicted the rotamer toggle switching of Trp6.48 and they claimed that it was in agreement with the early spectroscopic data [382]. However, this was not confirmed by later crystal structures of opsin and also rhodopsin with all-*trans*-retinal bound [353] and unbound but still

present in the binding site [354]. The spectroscopic properties of Trp6.48 really change during activation of rhodopsin but it is a result of large movement of cytoplasmic part of TM6 and smaller movements of adjacent helices so a local environment of Trp6.48 is altered even if a rotamer of Trp6.48 itself does not change.

Much more detailed approach to the study of ionic-lock induced activation of GPCRs has been proposed in the work of Balamaran *et al.* [383]. In this work the all-atom force field has been used to evaluate the relative stability of various point-mutations in  $\beta_1$ -adrenergic receptor. The results showed good correlation with the experimental data for over 90 single and multiple point mutants of this protein. It was found that Tyr5.58Ala and Val5.61Ala mutations stabilize the Arg3.50–Glu6.30 ionic lock, while Phe7.48Met mutation alters the interaction between the conserved NPxxY motif of TM7 and TM8.

### **Consequences of the Ionic Lock Instability and TM Movements**

The first experimental structure of GPCR, opsin, provided structural data which confirmed the importance of molecular switches and ionic locks [369]. While some of the structures of the early photoproducts of rhodopsin showed only an increased Arg3.50–Glu6.30 distance [372], in the opsin structure these residues are no longer interacting with each other. Arg3.50 is released from the Glu6.30 and Glu3.49 interactions and engages with Tyr5.58 on TM5, an interaction that was not suggested before. At the same time a new interaction between Lys231 and Glu247 is formed to stabilize TM5–TM6 interactions. Also, the Tyr7.53 of the nPxxY motif (TM7) aromatic stacking interaction with Phe7.60 (H8) is broken due to a different orientation of the helices. On the other hand the Trp6.48 toggle switch is in exactly the same position as in the inactive rhodopsin, even though it was indicated by NMR studies that it must change its position and interaction partners during activation [384].

A very similar structure of the active opsin, albeit with a different Trp6.48 rotamer, has been predicted computationally by Bhattachary *et al.* at the same time [385]. The authors started from the inactive (dark) state and have predicted TM conformational changes that are induced by the isomerization of 11-*cis* retinal to all-*trans* retinal with good accuracy. In another study Hornak *et al.* used nanosecond MD guided by NMR distance restraints to simulate the activation of rhodopsin [386]. Results of this work were also in agreement with experimental data and showed the coupling of retinal isomerization to the motions of helices and activation of the receptor which proceeds *via* a series of multiple switches.

To solve this inconsistency in behavior of Trp6.48, a MD simulation of  $\beta_2$ AR and bovine rhodopsin systems combined with mutational analysis of the ghrelin receptor was performed [387]. The 8 ns

simulation of inactive rhodopsin with 11-*cis*-retinal present, showed no change in the Trp6.48 conformation. The removal of retinal, however, allowed this residue to change its rotameric state. After additional 9 ns of simulation Trp6.48 established a moderately strong and not very stable interaction with Phe5.47 on TM5. In the  $\beta_2$ AR case removal of its ligand (carazolol) did not result in large conformational changes of Trp6.48 and no additional interactions have been observed. Additional data obtained from metal ion site engineering confirm the close proximity of these two residue. Mutational data from ghrelin receptors experiments show, on the other hand, that both of these residues are important for constitutive activity and agonist-induced efficacy. Authors suggest that the ionic lock may be one of the several molecular switches that form an allosteric interface between the TMs performing global toggle switch movements that mediate the intramolecular signal which leads to G protein activation. However, since ~30% of GPCRs lack the Trp6.48 residue, it may not be a part of the general activation pathways for 7TM receptors.

The problem of the lack of interaction between Arg3.50 and Glu6.30 in the  $\beta_2$ AR crystal structure, despite their presence in the sequence, has been addressed in a 2009 paper by Dror *et al.* [388]. For this system there was also a variety of biochemical evidence suggesting that an ionic lock between those residues is formed in the inactive state [375]. It was suggested that the broken lock may be a consequence of the techniques used to stabilize  $\beta_2$ AR for crystallization or of the binding of carazolol or timolol, which might reflect the ability of some partial agonists to induce signaling through disruption of this interaction. An all-atom simulation of  $\beta_2$ AR over 10 microseconds showed that an ionic lock forms reproducibly both in apo  $\beta_2$ AR and in the carazolol-bound system. Interestingly, the ionic lock had a tendency to break in simulation every few hundred nanoseconds and the conformation with this lock broken remained stable for tens of nanoseconds. In conclusion the authors suggested that the inactive  $\beta_2$ AR alternates between several major conformations with the ionic lock present and a few minor conformations without the ionic lock. Their long, microsecond simulation time allowed to describe the inactive state of this GPCR as an ensemble of states.

Almost at the same time another MD work on  $\beta_2$ AR and  $\beta_1$ AR appeared [389]. 500 ns MD runs for carazolol-bound  $\beta_2$ AR model and cyanopindol-bound  $\beta_1$ AR model showed the formation of the ionic lock, which remained stable during the stimulation runs, in both systems. In this work no breaking of the ionic lock has been observed; however, the timescale of the simulations might have been too short for such event to occur. Interestingly, water-mediated interactions Trp6.48-Asp2.50 and Asp2.50-Asn7.49-Tyr7.53 have been observed, some of which have been suggested to be important in activation of the thyrotropin receptor. The water-mediated hydrogen bonds also remained stable

during the simulation runs. The results of this study were different than the earlier theoretical investigation of the impact of ligand binding on the conformational state of the protein. Bhattachary *et al.* investigated the perturbations in the helical rotational orientations induced by ligand binding in the TM region of the  $\beta_2$ -adrenergic receptor [390]. They found that norepinephrine (full agonist) and dopamine (a weak partial agonist) break the Arg3.50–Glu6.30 ionic lock and engage the Trp6.48 rotamer toggle switch, while salbutamol (a partial agonist) only breaks the ionic lock and catechol (a very weak agonist) only switches the rotamer toggle.

A longer, 1.02 microsecond simulation of  $\beta_2$ AR presented by Romo *et al.* allowed the researchers to show the interconversion between ionic lock substates [391]. Similarly to previous results authors showed that the Arg3.50-Glu6.30 ionic lock is able to break and reform in the wild-type  $\beta_2$ AR simulations. The lock-breaking event is followed by the reorganization of the cytoplasmic end of TM6 through a clockwise rotation of the helix and movement of the end of TM5 away from TM6. These events were in agreement with the predicted activation models of this system. In 2011 Moukhametzianov *et al.* followed the work of Dror and showed that also  $\beta_1$ AR may have at least two distinct conformations in the inactive state [392]. Their new X-ray structures of this receptor show that TM6 and IC3 loop (connecting TM5 and TM6) may have a bent form, stabilized by previously unseen interactions Met281-Leu282 and His286-Gln237. They also noticed that the new TM6 conformation positions Arg3.50 and Glu6.30 only 3.8 Å apart, which is a distance significantly longer than in the rhodopsin case (with the ionic lock), but also much shorter than in the other conformation of TM6 (where the distance is 6.2 Å).

Experimental structures of active states of rhodopsin and  $\beta_2$ AR have shown that MD predictions of the behavior of ionic locks during activation were generally accurate. The constant development of computational methods and the more and more powerful computers allow us now to reach microsecond simulations [393,394] and obtain information about the general energy landscapes of GPCRs. Such theoretical energy landscapes and activation pathways, even though simplified to include only one or a few coordinates, give a very good means to look inside the GPCR activation mechanisms. The latest experimental results, e.g. highlighting the role of the highly-conserved Tyr5.58 in the stabilization of the active state of rhodopsin [395] are perfect examples of data and events that may be simulated by computational methods to gain more understanding of their mechanism of activation/inactivation and molecular processes involved in them.

## 10.5 Drug Design



Development of selective drugs for GPCRs is challenging because

- there is a high degree of homology among many closely related receptor subtypes that can regulate diverse physiological functions;
- a single receptor may couple to more than one G protein or even signal through G protein independent pathways;
- a receptor can be regulated by multiple allosteric ligands (small molecules but also proteins including GPCRs – what implicate allosteric influence by dimerization or oligomerization either homomers or heteromers);
- predominant signaling behavior of GPCRs may differ in different cells or organs.

Despite the progress which led to insights into the three-dimensional structures of GPCRs in both active and inactive states, the process of developing a drug for a particular GPCR target has become more complex, time-consuming and expensive [396]. Detailed characterization of agonist and antagonist binding behavior provided insight into the allosteric effect of G proteins on receptor structure and agonist binding affinity. The efficacy of ligands for activating the arrestin pathways can differ from those that activate G proteins. Some ligands possess a complex pharmacological behavior acting as agonists and simultaneously antagonists or inverse agonists depending on the pathways they activate and inactivate. Carvedilol is an inverse agonist for  $\beta$ 2AR activation of Gs but a partial agonist for  $\beta$ 2AR activation of arrestin [397]. The complexity of GPCR signaling pathways and ligand efficacy profiles complicate the process of drug discovery. Moreover, specific receptors might exhibit cell type-specific signaling as a consequence of the cell-specific complement of different proteins: signaling, regulatory and scaffolding.

As can be seen from crystal structures the structural diversity of GPCRs is much greater for amino acids lining the pathway into the ligand binding pocket. Although these amino acids do not make direct contact with a bound ligand the size of carazolol, they could contribute to initial transient interactions between the ligand and receptor that affect the ligand association rate. These residues might also play a role in binding larger ligands that extend into the vestibule of the binding pocket, such as the long-acting  $\beta$ 2AR agonist, salmeterol, that is used in the treatment of asthma [396]. The recent crystal structure of the chemokine CXCR4 receptor with an antagonist in the form of a peptide ligand, which extends from the receptor binding side toward the extracellular surface, show such kind of binding for the first time for GPCRs (PDB id 3OE0) [316]. Although many aspects of GPCR function can be explained by a simple two-state model, evidence from biophysical and functional studies support a multistate model in which ligands stabilize a specific conformational state or set of states [323] making the complexity of GPCR signaling similar to the microprocessor work [398].

## 10.6 Conclusions

Investigations of molecular switches in the superfamily of GPCRs is extremely challenging but may be truly rewarding because the detailed mechanism (or mechanisms) of activation of these receptors could help to design highly selective drugs acting not even on a single receptor subtype but on a single pharmacological receptor subprofile. Because of the intrinsic instability of GPCRs resulting in their multiple functionality, the investigations must proceed via elucidation of multiple structures with inverse agonists, antagonists and agonists, possibly also with trimeric G proteins and arrestins. Dimerization is a separate large issue currently unresolved but possibly the allosteric action of GPCRs via dimerization is the most common mechanism influencing receptor functioning. GPCRs are the biggest and one of the most mysterious single group of molecular targets for drugs, therefore, one can be sure that studies on their structures and mechanisms will be continued with an increasing pace.

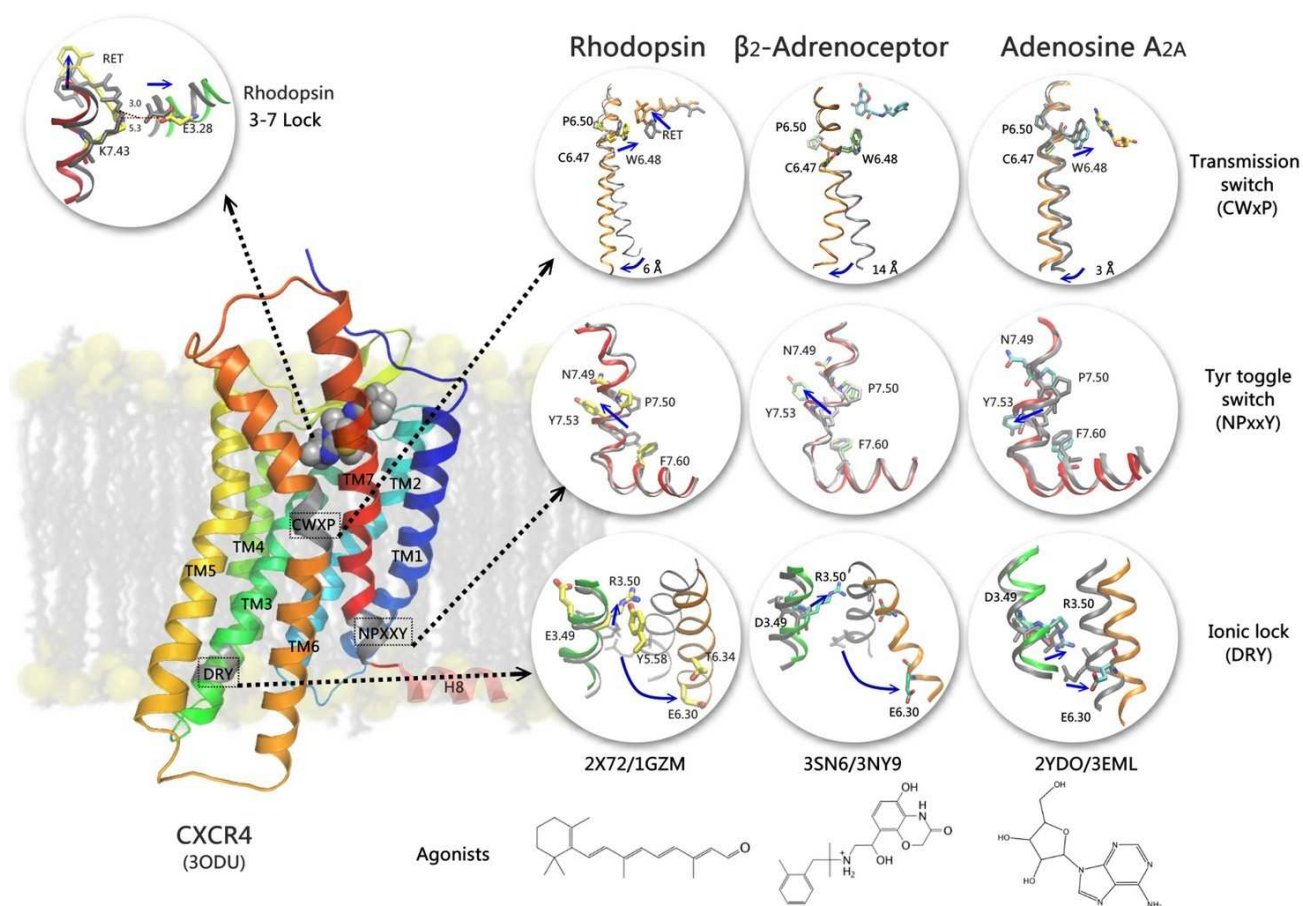


Fig 10.3 The action of molecular switches in GPCRs. Four switches are shown: transmission switch, tyrosine toggle switch, ionic lock, and 3-7 lock. They are shown based on the crystal structures of rhodopsin,  $\beta_2$ AR and  $A_{2A}$ R with agonists and antagonists/inverse agonists. Their id numbers from Protein Data Bank are provided – first number for inactive and second for active receptor. Additionally, the structural formulas of agonists from the crystal structures of active receptors are shown. The general scheme of GPCR structure is shown based on the crystal structure of chemokine receptor CXCR4 with a small ligand. Blue arrows in circular panels indicate motions of receptor structure during action of particular switch. Inactive receptor structure is shown in gray while active one in color. The residues are numbered according to the Ballesteros-Weinstein numbering scheme [265]

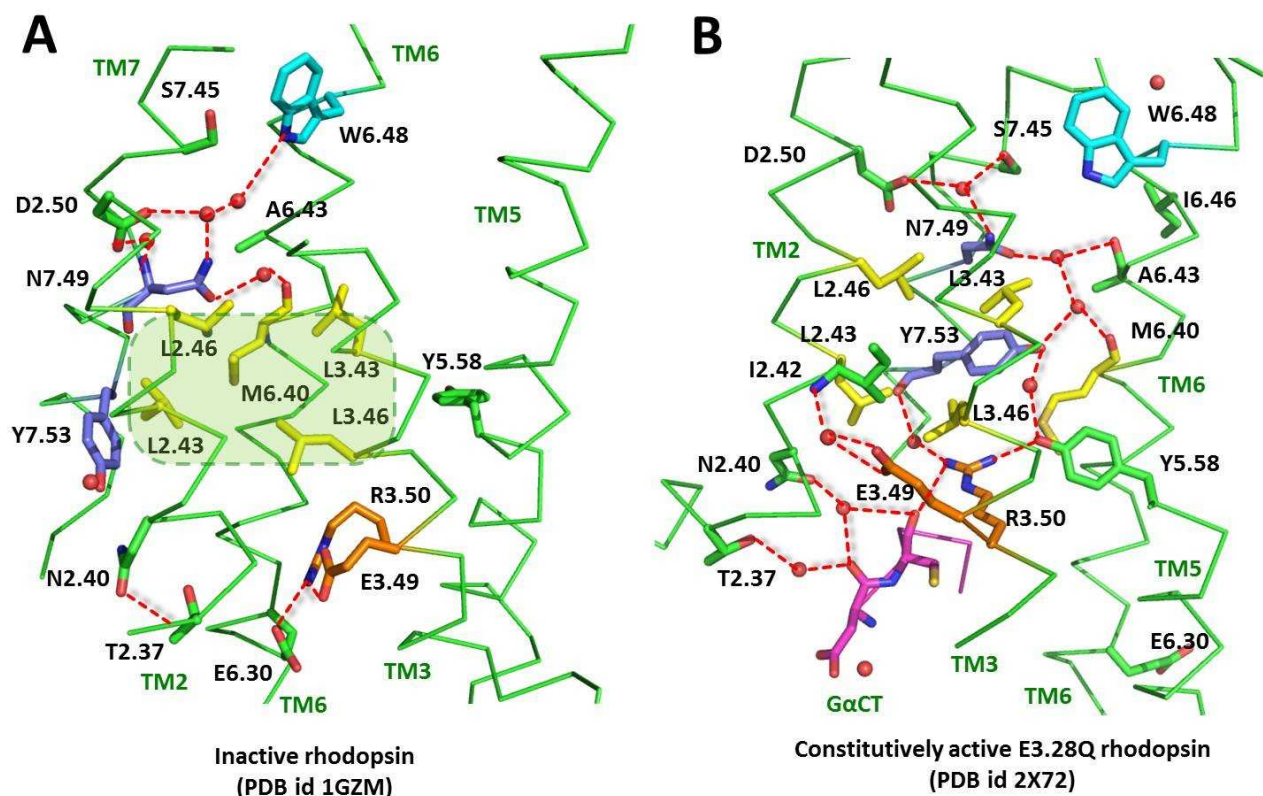
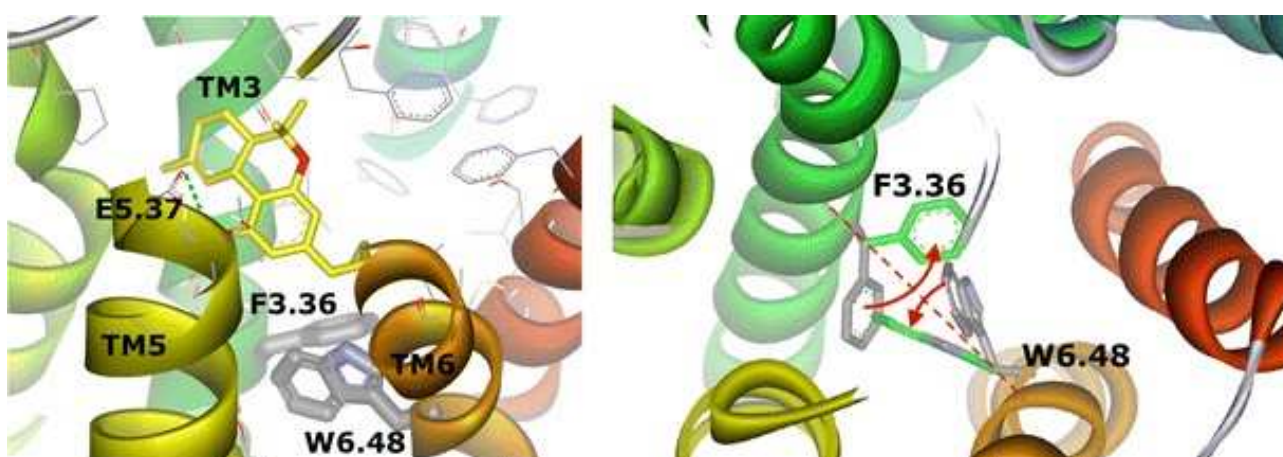


Fig 10.4 Rearrangement of hydrogen bond network in rhodopsin during its activation. A. The crystal structure of inactive rhodopsin (Protein Data Bank id 1GZM). B. The crystal structure of constitutively active Glu3.28Gln mutant of rhodopsin with all-trans retinal unbound from Lys7.43 but still present in the binding site (Protein Data Bank id 2X72). Both structures include water molecules (shown as red spheres) which participate in hydrogen bond network. In the inactive rhodopsin there is a hydrophobic area consisted of five residues located in helices TM2, TM3 and TM6 (in yellow) which form a hydrophobic barrier (area in green) separating residues in CwxP (cyan) and nPxy (blue) motifs from those of (d/e)Ry motif (orange). In the activated rhodopsin a rotation of TM6 disrupts the water mediated link between TM6 and TM7 and reorganizes the hydrogen bond network. Two tyrosine residues Tyr5.58 and Tyr7.53 reposition and fill the uncovered gap between TM3 and TM6 to extend hydrogen bond network toward (d/e)Ry motif and a fragment of G protein GαCT (pink). Figure is based on [354]. The residues are numbered according to the Ballesteros-Weinstein numbering scheme [265].

## 10.7 Modeling of ligand binding to G protein coupled receptors: cannabinoid CB<sub>1</sub>, CB<sub>2</sub> and adrenergic $\beta$ 2AR

Cannabinoid and adrenergic receptors belong to the class A (similar to rhodopsin) GPCRs. Docking of agonists and antagonists to CB<sub>1</sub> and CB<sub>2</sub> cannabinoid receptors revealed the importance of a centrally located rotamer toggle switch, and its possible role in the mechanism of agonist/antagonist recognition. The switch is composed of two residues, F3.36 and W6.48, located on opposite transmembrane helices TM3 and TM6 in the central part of the membranous domain of cannabinoid receptors. The CB<sub>1</sub> and CB<sub>2</sub> receptor models were constructed based on the adenosine A<sub>2A</sub> receptor template. The two best scored conformations of each receptor were used for the docking procedure. In all poses (ligand-receptor conformations) characterized by the lowest ligand-receptor intermolecular energy and free energy of binding the ligand type matched the state of the rotamer toggle switch: antagonists maintained an inactive state of the switch, whereas agonists changed it. In case of agonists of  $\beta$ 2AR, the (R,R) and (S,S) stereoisomers of fenoterol, the molecular dynamics simulations provided evidence of different binding modes while preserving the same average position of ligands in the binding site. The (S,S) isomer was much more labile in the binding site and only one stable hydrogen bond was created. Such dynamical binding modes may also be valid for ligands of cannabinoid receptors because of the hydrophobic nature of their ligand-receptor interactions. However, only very long molecular dynamics simulations could verify the validity of such binding modes and how they affect the process of activation [294].



*Fig 10.5 The rotamer toggle switch in cannabinoid receptors is comprised of two residues, F3.36 and W6.48, which are located on transmembrane helices TM3 and TM6. Docking of agonists and antagonists to CB<sub>1</sub> and CB<sub>2</sub> cannabinoid receptors revealed the importance of this centrally located switch and its possible participation in the mechanism of agonist/antagonist sensing. The best scored poses (ligand-receptor conformations) were obtained for the ligands matching the switch state: antagonists maintained the state of the rotamer toggle switch, whereas agonists changed it.*

### 10.7.1 Introduction

The recent crystal structures of class A (similar to rhodopsin) G protein coupled receptors (GPCRs), namely  $\beta_1$ - and  $\beta_2$ -adrenergic receptors ( $\beta_1$ AR [298] and  $\beta_2$ AR [266,399,400]) and adenosine receptor ( $A_{2A}$ R [401]) showed nearly identical structures of transmembrane domain but also differences in states of molecular switches as compared to rhodopsin [297] which was the first crystallized GPCR. Based on experimental data it was proposed that agonist binding and the receptor activation occur through a series of conformational intermediates. Transition between these intermediate states involves the disruption, creation or reorganization of intramolecular interactions that stabilize the basal states of a receptor. Such changes are made by the action of molecular switches (also called microswitches). The major switches proposed so far for different GPCRs, reflecting shared activation mechanisms, include the “rotamer toggle switch” involving the CWxPx(F/H) sequence on TM6 [373], the switch based on the NPxxY(x)<sub>(5,6)</sub>F sequence linking TM7 and H8 [356], and the “ionic lock” linking transmembrane helices TM3 and TM6 and employing the (E/D)RY motif on TM3. There are also switches not connected to any particular sequence motifs like the “3-7 lock” involving the interaction connecting TM3 and TM7 and present only in selected receptor types: this switch involves Schiff base-counterion interaction in rhodopsin [71,402] and it was proposed to operate in opioid receptors [349–351].

In the rhodopsin structure which is completely inactive in the basal state i.e. when retinal is in 11-*cis* conformation, all switches are assumed to be in their *off* positions. However, recent crystal structures of GPCRs, assumed to be in *inactive* state because they were complexes with antagonists and inverse agonists, showed remarkable similarities of states of switches that are different from those of inactive rhodopsin: ionic lock is in open state (broken connection between helices TM3 and TM6), and a connection between TM7 and H8 is broken because of change of rotamer of Y7.53 (part of NPxxY motif on TM7). On the other side, the “rotamer toggle switch” involving W6.48 residue remains in its *off* state (identical as in inactive rhodopsin structure) in all receptor crystal structures even in Meta-II rhodopsin structure assumed to be in *activated* form [369,403]. Therefore, based on existing crystal structures, different states of switches may not describe *on* and *off* positions but different states adopted during activation processes because these receptors may be partially activated in their basal state. In our earlier papers we investigated early activation steps occurring simultaneously to ligand binding for opioid receptors MOR ( $\mu$ ), DOR ( $\delta$ ) and KOR ( $\kappa$ ). The first switch that was broken by agonist binding was “3-7 lock”, a hydrogen bond D3.32-Y7.43 linking transmembrane helices TM3 and TM7. It was the first activation event observed. We also detected action of the second switch: rotamer toggle switch involving simultaneous change of side chain conformations of W6.48 and adjacent residues, therefore called the extended toggle switch. In case

of opioid receptors the other residue in this extended switch was H6.52. This residue also participated in agonist-antagonist sensor based on propensity for creating a hydrogen bond with Y3.33 for antagonists and H6.52 for agonists. All studied ligands, being analogs of morphine – with common tyramine structural scaffold, created a salt bridge with D3.32 with their protonated nitrogen atom of tyramine group. This sensor was studied for MOR [349] and later for DOR and KOR [350]. The proposed mechanism of its action was later confirmed via molecular dynamics simulations of tightly related agonist-antagonist pair of KOR ligands: 5'- and 6'-GNTI [351].

Ligands of opioid receptor are similar to ligands of  $\beta_1$ - and  $\beta_2$ -adrenergic receptors because they interact in the binding site of the receptor in protonated form. This is not the case for ligands of cannabinoid receptors because of their hydrophobic properties. Following our earlier research on opioid receptors where we proposed agonist-antagonist sensor we investigated early activation states in  $CB_1$  and  $CB_2$  receptors focusing on the centrally located rotamer toggle switch involving residues W6.48 and F3.36 located on two different helices. Here we show our latest results on cannabinoid receptors  $CB_1$  and  $CB_2$  regarding binding of ligands and possible activation steps simultaneous to ligand binding. Ligands of cannabinoid receptors are mostly hydrophobic what reflects the presence of large number of hydrophobic residues in the binding sites of their receptors. Therefore, the microswitches may be modified but they are still operational according to the structural mimicry rules. A mechanism for differentiation between agonists and antagonists is not clear for ligands of cannabinoid receptors because their structures are mostly highly flexible so finding the agonist-antagonist sensor is more difficult using simulation techniques, therefore we decided to start with simple docking method including flexible amino acid residues. This method was used to check the possibility whether a W6.48/F3.36 rotamer toggle switch may participate in mechanism of agonist/antagonist sensor.

We also show our results regarding interactions between  $\beta_2$ -adrenergic receptor ( $\beta_2AR$ ) model and two enantiomers of fenoterol: (*R,R*)-fenoterol and (*S,S*)-fenoterol investigated by molecular dynamics (MD) simulations. Fenoterol is a selective  $\beta_2AR$  agonist and exists in four stereoisomers which significantly differ in  $\beta_2AR$  efficacies and selectivities. Racemic mixture of (*R,R*)- and (*S,S*)-isomers, *rac*-fenoterol is clinically used for the treatment of asthma. Radioligand binding studies evidence that fenoterol stereochemistry greatly influences the binding affinity to the  $\beta_2AR$  with relative order (*R,R*)>(*R,S*)>(*S,R*)>(*S,S*) and a similar trend was found in functional assays [404]. Moreover, fenoterol stereochemistry also affects the mode of coupling of activated  $\beta_2AR$  to G proteins studied in experiments with addition of pertussis toxin, a selective blocker of  $G_i$  mediated signaling. In this experiments the toxin has no effect on the activity of (*R,R*)-fenoterol what indicates that this stereoisomer upon binding to the receptor activates it to couple exclusively  $G_s$  protein [405].

When other stereoisomers of fenoterol were tested in such experiments, the addition of pertussis toxin significantly reduced a functional effect what implies that binding of non (*R,R*)- stereoisomers activates  $\beta_2$ AR to the form which couple both  $G_i$  and  $G_s$  protein [405].

### 10.7.2 Results

To investigate influence of ligand binding on amino acid residues in the binding site of  $CB_1$  and  $CB_2$  receptors to see a potential action of the rotamer toggle switch we used two best scored by Modeller conformations of cannabinoid receptors. These two conformations of the same receptor were similar to each other in transmembrane domain but quite different in extracellular loops area especially for  $CB_2$  receptor (Fig. 10.7). Calculated root mean square distance (RMSD) between the two conformations of  $CB_1$  model was 0.258 nm and of  $CB_2$  0.265 nm (counting all heavy atoms in the receptor structure excluding N- and C-termini). The residues forming the rotamer toggle switch were located in the same positions whereas most of other residues from the binding site took different conformations. For each pose of ligand-receptor pair the free energy of binding and the ligand-receptor intermolecular energy were calculated. Intermolecular energy ( $E_{int}$ ) is defined as a sum of four components: energies of van der Waals interactions, electrostatic interactions, hydrogen bonds and desolvation of a ligand. An error of  $E_{int}$  estimation is about 2.5 kcal/mol.  $\Delta G$  is defined as a sum of  $E_{int}$  and the entropy term which is a change of entropy of a ligand and of a set of flexible amino acid side chains of a protein. The results obtained for two different conformations of each receptor are presented in Table 1. Usually  $\Delta G$  and  $E_{int}$  had similar values with the exception of anandamide, which flexibility increased  $\Delta G$  values. Nevertheless, the highly negative value of  $E_{int}$  in case of AEA compared to other ligands preserved a negative value of  $\Delta G$ . Such exceptionally low value of intermolecular energy of anandamide inside a binding pocket of cannabinoid receptors is also associated with its flexibility leading to the highest match with the amino acid from the binding site, especially hydrophobic ones. Regardless of which conformation of the receptor was used the poses with lowest  $E_{int}$  and  $\Delta G$  were obtained for the ligands matching rotamer switch state (involving residues W6.48 and F3.36): antagonists maintained state of the rotamer toggle switch, whereas agonists changed it (values of  $E_{int}$  and  $\Delta G$  in bold in Table 1). They were selected as having the lowest energy from two receptor conformations and from two states of the switch. The deviations from the pure *gauche+* and *trans* rotamers of W6.48 and F3.36 for chi1 angle were high, usually about  $\pm 30^\circ$  and sometimes even higher up to  $\pm 60^\circ$  which is half the way between pure *gauche+* and *trans* states.



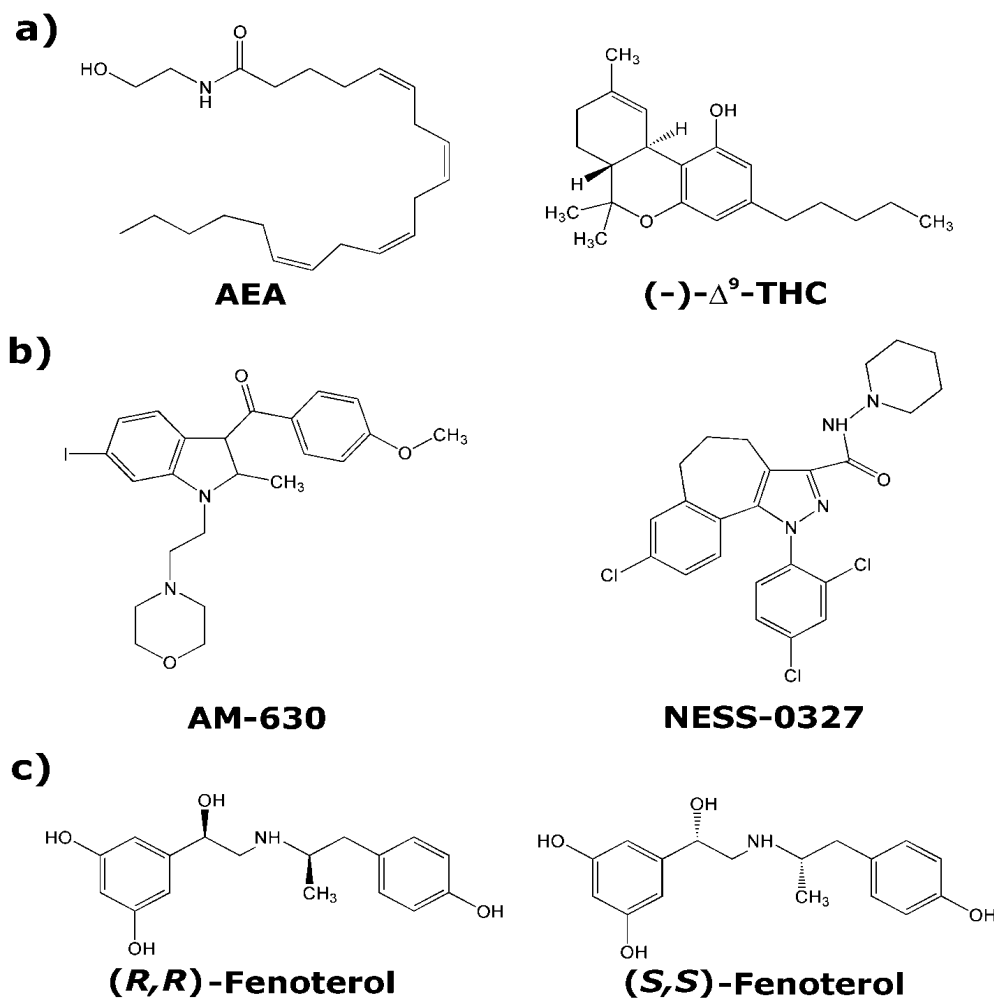


Fig 10.6 Agonists **(a)** anandamide (AEA) and (-)- $\Delta^9$ -THC (both unselective for  $CB_1$  and  $CB_2$ ), and antagonists **(b)** AM630 ( $CB_1$  selective) and NESS-0327 ( $CB_2$  selective) used in the study. **(c)** (R,R)- and (S,S)-fenoterol - agonists of  $\beta_2AR$ .

For the best pose (characterized with the lowest energies  $E_{int}$  and  $\Delta G$ ) of the selective antagonist NESS-0327 in complex with  $CB_1$  receptor (Fig.10.6a) we observed the  $\pi$ - $\pi$  aromatic interactions of the ligand with F268 residue. The piperidine ring of the ligand was located close to helices TM2 and TM7 while other rings in the area among helices TMs:3-4-5-6. For the selective antagonist of  $CB_2$ , AM-630, we analyzed both the protonated and unprotonated form and got similar results of docking but slightly different energies of binding (Table ). The protonated form tended to bind to conformation one of  $CB_2$ , while the unprotonated form to conformation two (although the energies for both conformations were very similar). Interestingly, the charged ligand was classified with higher energies than the ligand in unprotonated form for both receptor conformations. The ligand was bound to the receptor with the morpholine ring located close to the rotamer toggle switch area (Fig. 10.6b). The cation- $\pi$  interaction was possible after a slight rotation of the phenyl ring (change of the  $\chi_2$  angle by  $25^\circ$  with no change of the  $\chi_1$  angle). An oxygen atom from the morpholine ring

formed a hydrogen bond with S5.42<sup>(193)</sup>, rather than S4.53<sup>(161)</sup> and S4.57<sup>(165)</sup> residues which are more distant to the ligand and covered partly by helix TM3. These serine residues in TM4 were important for binding of antagonist SR-144528 but not CP-55940 or WIN-55212-2 as pointed out by mutagenesis studies described in [406]. WIN-55212-2 contains a morpholine ring similarly to AM-630 and is also CB<sub>2</sub> selective. Interestingly, the serine residues, S4.53 and S4.57, are present only in CB<sub>2</sub> receptor so the other residue should be responsible for selectivity. According to mutagenesis studies done in Reggio group [407] such a residue is F5.46<sup>(197)</sup> and in fact in our model this residue is close to morpholine ring of AM-630. The anisole ring was located close to helices TM6 and TM7, while the iodobenzene ring close to helices TM3, TM4 and TM5. Among many analyzed poses of the antagonist AM-630 at the binding site of the CB<sub>2</sub> receptor one conformation appeared noteworthy due to a salt bridge formed by a protonated nitrogen atom of the morpholine ring of the ligand and the carboxylic group of E5.37 (not shown). Nevertheless, that pose was energetically less favorable (as scored by docking procedure). It means that appropriate hydrophobic interactions were a predominant feature of the best poses of this antagonist.

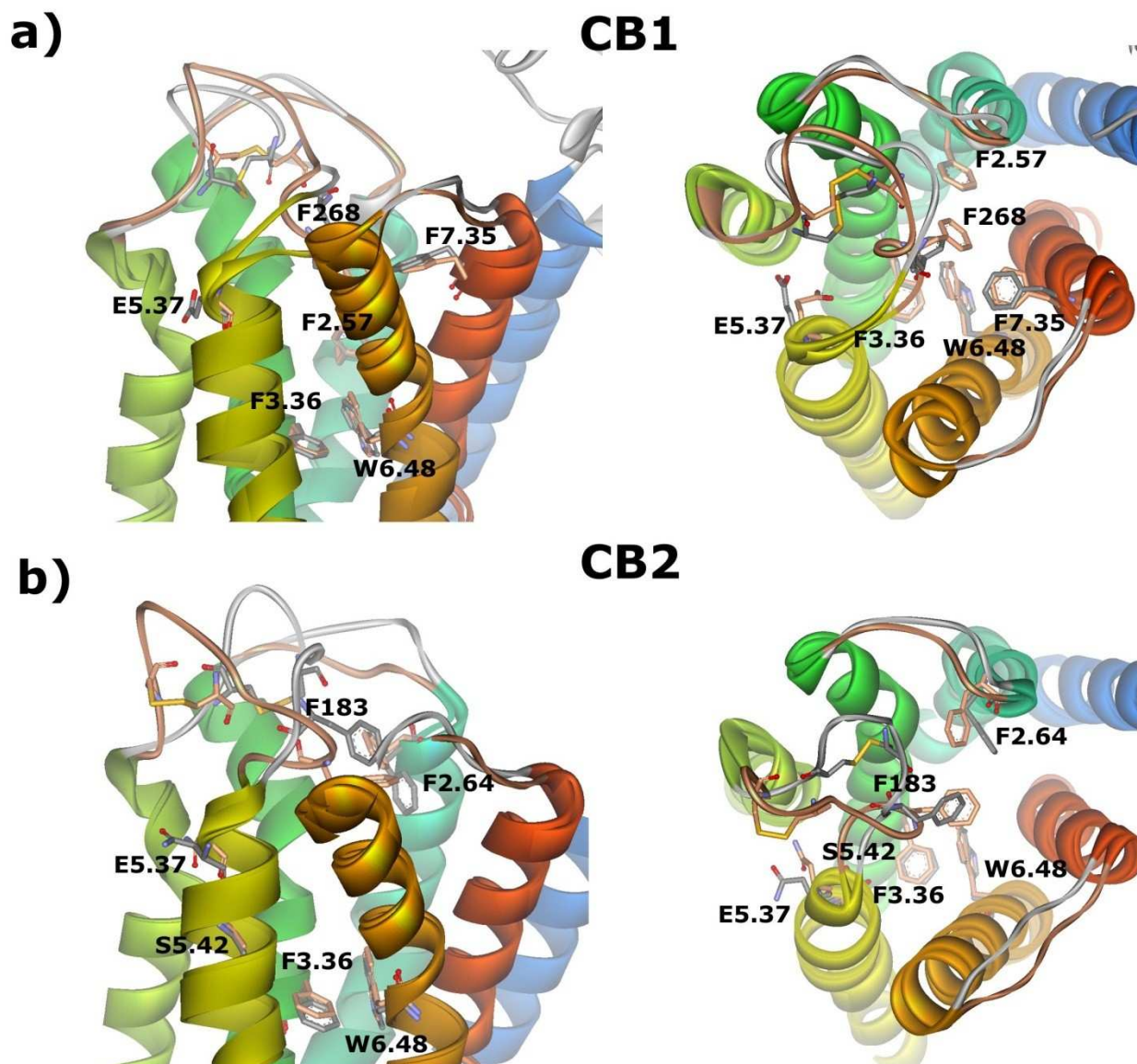


Fig 10.7 Two views of the binding site, parallel to the membrane and from the extracellular side, of CB<sub>1</sub> (a) and CB<sub>2</sub> (b) receptors of two the best scored conformations for each receptor superimposed. All presented models are based on the A<sub>2A</sub>R template. Residues forming the rotamer toggle switch and some characteristic residues of the binding site are shown.

Binding of nonselective agonists (-)- $\Delta^9$ -THC and anandamide (AEA) changed the rotamers of the W6.48 and F3.36 residues in the rotamer toggle switch in both receptors (Fig.10.9). In the CB<sub>1</sub> receptor an alkyl tail of AEA interacted with F268 (EC2 loop) through  $\sigma$ - $\pi$  interactions. The polar part of AEA was located between helices TMs:3-4-5, near the residues T3.33 and E5.37. The same location of the polar part of AEA was observed at the binding site of the CB<sub>2</sub> receptor, although the polar head of AEA was translated more toward TM4, possibly because TM4 in CB<sub>2</sub> contains more polar residues (two serine residues S5.53 and S5.57 instead of alanine residues in CB<sub>1</sub>). Additionally, a nitrogen atom in the polar head of AEA formed a hydrogen bond with S5.42. The alkyl tail of the ligand was elongated and its end was located in the hydrophobic pocket of CB<sub>2</sub> between TM2 and

TM3 so it could interact with phenylalanine residues F2.61, F2.64, F3.25 and the hydrophobic part of K3.28 (charged part of this lysine residue projected outward of the receptor). The middle part of the tail of AEA interacted with F183 from the EC2 loop, similarly as in the AEA-CB<sub>1</sub> complex. Such interactions between phenylalanine in EC2 and CB<sub>1</sub> agonists was confirmed by mutagenesis in [259]. As regarding the other agonist, (-)- $\Delta^9$ -THC, its hydroxyl group formed a hydrogen bond with E5.37 in the THC-CB<sub>1</sub> complex, while its alkyl tail was located in the hydrophobic area between helices TMs:3-6-7 close to the rotamer toggle switch. In the THC-CB<sub>2</sub> complex the similar hydrogen bond was created between the hydroxyl group of a ligand and the carboxylic group of E5.37. The alkyl part of THC was located between helices TMs:3-5-6 of the receptor and also close to the switch, while the opposite part of the ligand interacted with F2.64 ( $\sigma$ - $\pi$  interactions).

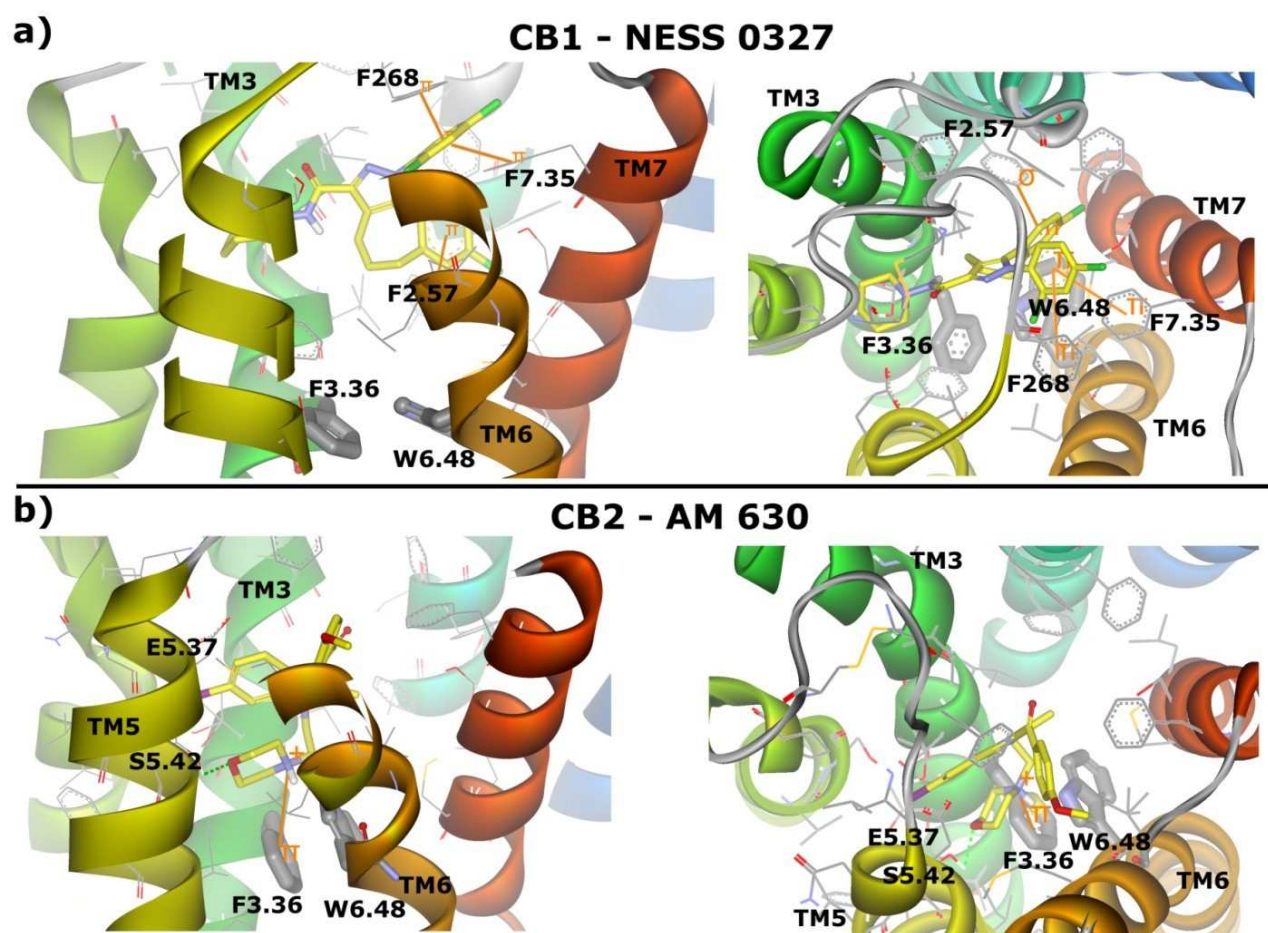


Fig 10.8 Selective antagonists NESS-0327 and AM630 bound to the CB<sub>1</sub> (a) and CB<sub>2</sub> (b) respectively, in complexes characterized by the lowest energies. Interactions involving  $\pi$  orbitals are shown as orange solid lines and hydrogen bonds as green dashed lines. Residues important for the binding of the ligand are labeled.

For analyzed  $\beta_2$ AR complexes with fenoterol the molecular dynamics simulations revealed different binding modes for (RR)- and (SS)-fenoterol isomers. The starting positions of two ligands were very

similar. Two hydroxyl groups of ligand's 1,3-benzenediol moiety were located in the vicinity of S5.42 and S5.46 residues from TM5, whereas ligand's protonated nitrogen atom was at close distance to residues D3.32 from TM3 and N7.39 from TM7. During MD simulation of (*R,R*)-fenoterol the ligand's hydroxyl group located at first stereogenic center formed hydrogen bonds with side chains of D3.32 and N7.39 residues. The position of D3.32 side chain was also stabilized by interaction with Y7.43 residue from TM7 helix. The hydrogen bond between hydroxyl group of ligand's phenolic moiety and D192 residue located in EC2 loop was also well preserved during MD simulation. The two hydroxyl groups of ligand's 1,3-benzenediol moiety were oriented toward S5.42 and S5.46 residues from TM5, but the distance was too large to form a direct hydrogen bond. In addition, the N6.55 residue, which is believed to play a crucial role in agonist binding and receptor activation [250], did not form any stable interaction with ligand during MD simulation. In contrast to (*R,R*) enantiomer, the (*S,S*)-fenoterol isomer did not maintain a stable position in the receptor binding cavity. Only one well preserved contact during each of 11 MD simulations was formed between protonated nitrogen atom of (*S,S*)-fenoterol and a carboxyl group of D3.32 residue. Interaction between N7.39 residue and the ligand was temporarily formed but frequently broken. The average distance between the hydroxyl group of the ligand phenolic moiety and the carboxyl group of D192 (EC2 loop) residue was too large ( $> 0.45$  nm) to establish a stable hydrogen bond. The movement of the ligand in the binding cavity allowed two hydroxyl groups of the molecule 1,3-benzenediol moiety to move closer toward S5.42 and S5.46 residues from TM5, but no stable connections were made. Representative conformations of both (*R,R*)- and (*S,S*)-fenoterol isomers were extracted from 22 MD simulations based on the lowest RMSD values when compared to average position of the ligand observed during 5 ns MD trajectory. Superimposition of all extracted structures for each complex is shown in Fig 10.10.

**Table 1** The free energy of binding and intermolecular energy of the receptor-ligand complexes. The numbers in bold indicate the lowest energies for particular complexes and hence the most probable conformations of their structures

Complexes	Rotamer toggle switch state is matching the ligand type <sup>a</sup>				Opposite state of the switch to the ligand type			
	Conformation 1 <sup>b</sup>		Conformation 2 <sup>c</sup>		Conformation 1		Conformation 2	
	$\Delta G$ <sup>d</sup>	$E_{int}$ <sup>e</sup>	$\Delta G$	$E_{int}$	$\Delta G$	$E_{int}$	$\Delta G$	$E_{int}$
<b>Antagonists</b>								
NESS-0327 – CB <sub>1</sub>	<b>-6.60</b>	<b>-7.50</b>	-3.93	-4.82	-2.16	-3.06	-4.28	-5.17
AM630(+) – CB <sub>2</sub>	<b>-2.62</b>	<b>-4.41</b>	-1.91	-3.70	0.38	-1.41	3.73	1.94
AM630 – CB <sub>2</sub>	-3.32	-5.11	<b>-3.37</b>	<b>-5.16</b>	-1.01	-2.80	-2.11	-3.90
<b>Agonists</b>								
AEA – CB <sub>1</sub>	-0.01	-5.07	<b>-2.40</b>	<b>-7.47</b>	-0.34	-5.42	0.76	-4.32
THC – CB <sub>1</sub>	-3.73	-5.23	<b>-4.54</b>	<b>-6.03</b>	-3.06	-4.55	-4.20	-5.69
AEA – CB <sub>2</sub>	-3.90	-8.97	<b>-5.21</b>	<b>-10.28</b>	-1.71	-6.78	-3.33	-8.40
THC – CB <sub>2</sub>	<b>-4.32</b>	<b>-5.81</b>	-3.82	-5.32	-1.70	-3.19	-2.27	-3.76

<sup>a</sup> Matching means *right-right* state of the switch for binding of antagonists and *left-left* state for binding of agonists (Fig. 6)

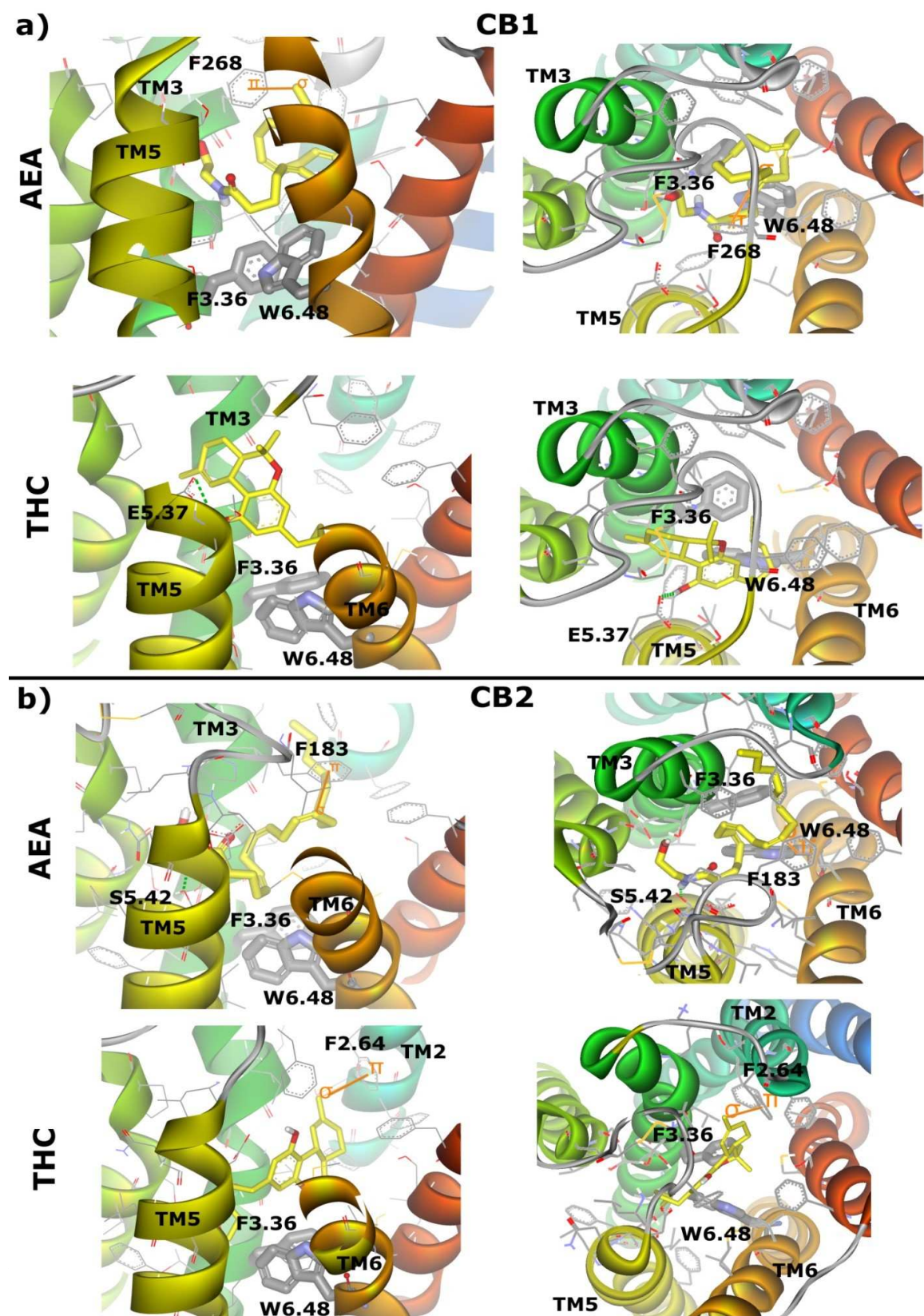
<sup>b</sup> The first of the two best receptor conformations generated by Modeller from the A<sub>2A</sub>R template

<sup>c</sup> The second of the two best receptor conformations generated by Modeller from the A<sub>2A</sub>R template

<sup>d</sup> Free energy of binding (calculated by Autodock) [kcal mol<sup>-1</sup>]

<sup>e</sup> Intermolecular energy of the protein-ligand complex (calculated by Autodock) [kcal mol<sup>-1</sup>] AM630(+) – protonated form of AM630

Fig 10.9 Nonselective agonists (-)- $\Delta^9$ -THC and AEA bound to the CB<sub>1</sub> (a) and CB<sub>2</sub> (b) in complexes characterized by the lowest energies. Interactions involving  $\pi$  orbitals are shown as orange solid lines and hydrogen bonds as green dashed lines. Residues important for the binding of the ligand are labeled.



changing its state simultaneously to the ligand binding. The state of the switch was able to change, even without a change of the backbone of the receptor, indicating ability to change in early activation stages of the receptor. Following the structural mimicry principle in class A of GPCRs one can identify this switch as a pair of residues W6.48/F3.36 in cannabinoid receptors CB<sub>1</sub> and CB<sub>2</sub>. These two residues being in contact with each other are located on two opposite helices TM6 and TM3. McAllister *et al.* [364] based on biochemical experiments highlighted the importance of this switch as a constraint for the CB<sub>1</sub>-inactive state that may need to break during activation. Their modeling studies using inactive and fully activated CB<sub>1</sub> receptor models indicated that the W6.48/F3.36 contact can exist in the inactive state of CB<sub>1</sub> and be broken in the activated state via a  $\chi_1$  rotamer switch: W6.48(*gauche*+ $\rightarrow$ *trans*) and F3.36(*trans* $\rightarrow$ *gauche*+). So it may represent a "toggle switch" for activation of CB<sub>1</sub>. Our docking studies confirm their findings, however, because our modeling did not include movement of backbone of helices one can regard such state as early activation state of the receptor. In such a state there is still a partial interaction between W6.48 and F3.36 residues (Fig 10.11) although they are located in new places ready to start rearrangement of adjacent residues through modification of the hydrogen bond network and imposing forces on adjacent helices for their later displacement.

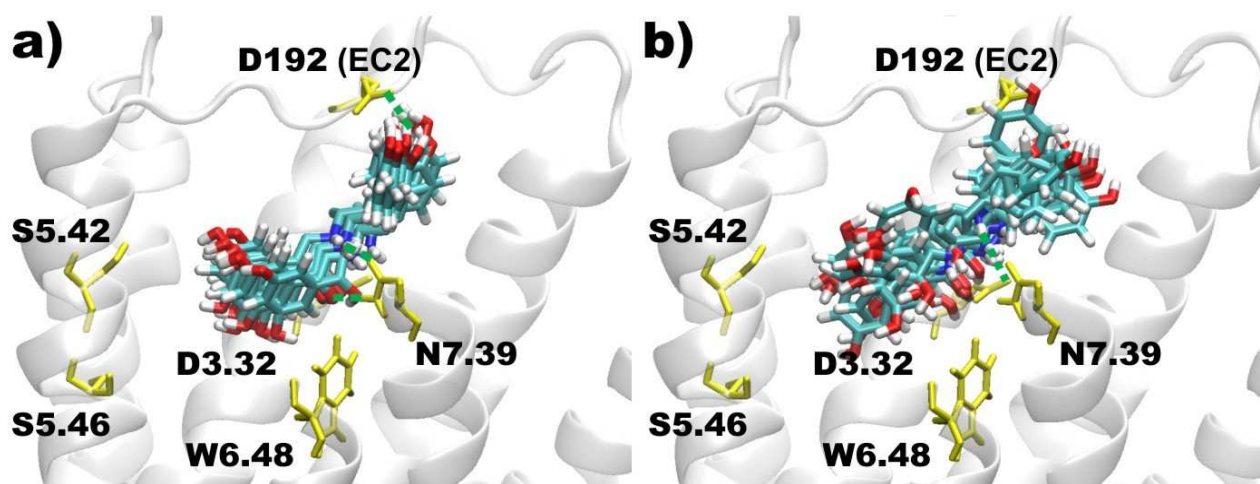


Fig 10.10 Superimposition of representative ligand conformations extracted from 22 molecular dynamics simulations of fenoterol- $\beta_2$ AR complex; 11 conformations per each isomer of fenoterol: (*R,R*)-fenoterol (a), and (*S,S*)-fenoterol (b). Green dashed lines indicate hydrogen bonds.

The presence of pure  $\chi_1$  *gauche*+ and  $\chi_1$  *trans* rotamers of W6.48 and F3.36 was rare in the obtained poses especially for agonists of cannabinoid receptors: frequently the  $\chi_1$  angle deviated by about  $\pm 30^\circ$  and sometimes (for the lowest energy pose of AEA-CB<sub>1</sub> complex) even by  $\pm 60^\circ$  which is in between *gauche*+ and *trans* states. This is possibly because of crowding in the binding site and the mismatch (which is a driving force for receptor activation) between the inactive structure of the

receptor and docked agonists, as we analyzed the initial stages of activation by binding of all ligands to the receptor in the inactive state. Because of such discrepancies in the  $\chi_1$  angle of W6.48 and F3.36 residues, we decided to introduce another notation of the state of the switch indicating the mutual positions of the two residues in relation to each other. In this way the so called inactive state (present in the template and maintained by antagonists) is right-right (R-R), whereas the state preferred by agonists is left-left (L-L). The definition is based on the position taken by these two residues with regard to the line connecting their  $C_\beta$  carbon atoms viewed from the extracellular side (it is further explained and visualized in Fig.10.11). The  $\chi_2$  angle of W6.48 and F3.36 residues was also able to change and helped to define the state of the switch: if a change of the  $\chi_1$  angle was not big enough to differentiate between gauche+ and trans rotamers, then the location of the ring (via  $\chi_2$  angle - especially for W6.48) indicated the specific R-R or L-L state. Therefore, by incorporating the  $\chi_1$  and  $\chi_2$  angles, the proposed naming rule properly differentiated inactive and activated states of the rotamer toggle switch in the early stages of receptor activation. The conformations of the receptor for the best scored poses of the ligand-receptor complex were different but the switch was always in the L-L state for agonists while it was in the R-R state in the case of antagonists (the values in bold in Table 1). The poses with an opposite state of the switch (also obtained during docking procedure) were always classified with higher free energies and intermolecular energies. The docking procedure also provided right-left (R-L) and left-right (L-R) states of the switch among the docking results but all of them were of higher energy and are not shown.



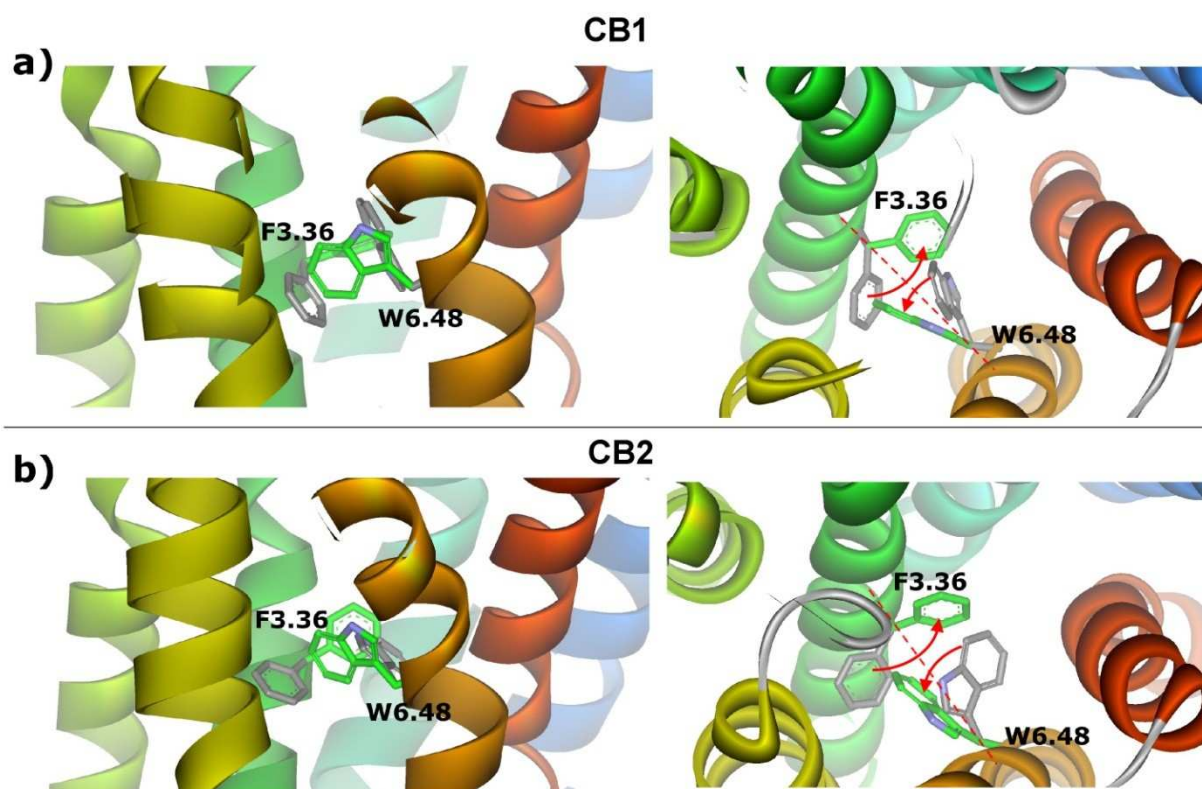


Fig 10.11 The location of W6.48 and F3.36 residues (forming the rotamer toggle switch) in molecular models of CB<sub>1</sub> (a) and CB<sub>2</sub> (b) receptors viewing from the membrane (left panels) and from extracellular side (right panels). Residues in grey indicate so called inactive state (from complexes with antagonists), residues in green indicate activated state (from the complexes with agonists). Red arrows show the change of positions of W6.48 and F3.36 residues during the activation process. Based on location of these two residues to the line connecting their C<sub>β</sub> carbon atoms, viewing from extracellular side, one can identify the states of the switch as right-right (R-R) state (both side chains on their right positions) or left-left (L-L) state (both side chains on their left positions).

The observations of the possible ligand poses in the cavity of cannabinoid receptors led us to the conclusion, that the most important for the ligand binding are the van der Waals interactions while hydrogen bonds and even ionic interactions are less important. The highly scored poses (i.e. with the lowest energies) were often lacking strong hydrogen bonds (defined as donor-acceptor distance less than 0.25 nm), although some weak (donor-acceptor distance less than 0.45 nm [408]) could still be formed. The common interactions between the ligands and cannabinoid receptors involved  $\pi$  orbitals in phenyl rings, which are present in large amounts in the CB<sub>1</sub> and CB<sub>2</sub> receptors' binding sites, especially in the cavity between helices TM2, TM7 and EC2 loop (Fig. 3 and Fig. 4). For the most flexible ligand, AEA, about  $5 \cdot 10^6$  poses generated for each ligand for each receptor conformation by docking procedure ( $2.5 \cdot 10^5$  energy evaluations \* 20 runs) was still too small number to properly sample conformational space of AEA even in confined space of the receptor binding site. However, obtaining the *activated* state of the switch for the lowest energy poses of this agonist indicates that

such poses are not rare and maybe they are dominant even for inactive receptor structure that was used for docking.

In the computational studies Gonzalez *et al.* [409] performed docking on homology modeled CB<sub>1</sub> receptor. They docked WIN55,212-2 (agonist), SR-141716a (inverse agonist) and CP-55940 (non-classical agonist) ligands to the conformation of CB<sub>1</sub> in basal state and to the model of activated receptor obtained by translation and rotation of TM6. Docking results were similar to obtained earlier by McAllister *et al.* [410] and confirmed preferential binding of ligands using two aromatic microdomains of CB<sub>1</sub> confined by helices TMs:3-5-6-7 and helices TMs:1-2-3-7. Docking revealed that all three ligands preferred the same binding site (TMs: 3-5-6-7), and only in the model of activated receptor (R\*), one ligand, the non-classical agonist CP-55940, was docked to the second binding site. Another computational study on CB<sub>1</sub> and CB<sub>2</sub> receptors modeled on β<sub>2</sub>AR and rhodopsin templates [411] revealed binding modes of AMG3 ligand. Both templates proved to be useful and provided similar docking results. Aromatic residues were also an important part of the binding site. Shim *et al.* [412] conducted 105 ns molecular-dynamics simulations of empty CB<sub>1</sub> receptor embedded in POPC bilayer for obtaining the structural and functional features of CB<sub>1</sub> over time. The helical bundle maintained a structure very similar to the x-ray structures of GPCRs. It was also revealed that the CB<sub>1</sub> receptor is stabilized by the formation of extensive, water-mediated hydrogen bond networks and aromatic stacking interactions within the helical core region. It is likely that these interactions, involving specific functional motifs, are the molecular constraints imposed on the inactive state of the CB<sub>1</sub> receptor. It was also suggested that disruption of these specific interactions is necessary to release the molecular constraints to achieve a conformational change of the receptor suitable for G-protein activation.

Based on accumulated biophysical and biochemical knowledge in the recent review of Shim [412] the initial stages of CB<sub>1</sub> receptor activation as well as a usefulness of functional residues of CB<sub>1</sub> receptor for drug discovery are discussed. The role of loops between transmembrane helices in ligand binding and differentiation for CB<sub>1</sub> was tested by Ahn *et al.* [259] based on alanine scanning mutagenesis of CB<sub>1</sub> EC2 loop. Their findings are consistent with a dual role for EC2 in stabilizing receptor assembly and in ligand binding. Selectivity issues i.e. how to develop compounds with high affinity for the CB<sub>2</sub> receptor and little affinity for the CB<sub>1</sub> receptor, are discussed in [413] for classical cannabinoids and cannabimimetic indoles. Interestingly, two opposite approaches, receptor mutations and molecular modeling, have been employed to obtain binding data. The selectivity of cannabinoid receptors was also investigated using molecular modeling and automated docking procedures [414]. An analysis of the interaction of WIN55212-2 with both receptors showed that CB<sub>2</sub>/CB<sub>1</sub> selectivity is mainly determined by the interaction with the nonconserved residues S3.31

and F5.46 in CB<sub>2</sub> receptor. Importance of these residues was suggested by site-directed mutagenesis experiments.

Another type of selectivity involving coupling of  $\beta_2$ AR to G<sub>s</sub> and G<sub>i</sub> proteins was recently reported for fenoterol stereoisomers [405]. Based on conducted MD simulations of  $\beta_2$ AR-fenoterol complexes we observed different binding modes for (R,R)- and (S,S)-fenoterol isomers which are in agreement with recent results from docking experiments [414]. (R,R)-fenoterol adopted stable conformation inside the binding cavity and created a network of hydrogen bonds involving D3.32 and N7.39 residues, which was well preserved during simulation. In contrast, (S,S)-isomer showed much higher conformational flexibility. Interaction between ligand's protonated amine group and D3.32 residue was the only one well preserved during MD simulation. This findings can be important in explaining the differences of the  $\beta_2$ AR active states leading to G<sub>s</sub> or G<sub>s</sub>/G<sub>i</sub> selective coupling. W6.48 did not change its state during any of the 5 ns simulations possibly because of short length of them. However, such length was enough to show critical differences in ligand binding modes of tightly related compounds.

A number of conformational switches in  $\beta_2$ AR has been reported so far. Two the best characterized are the R3.50-E6.30 ionic lock and W6.48 rotamer toggle switch. It was demonstrated that norepinephrine and dopamine break the ionic lock and engage the rotamer toggle switch whereas salbutamol, a noncatechol partial agonist, only breaks the ionic lock while the weak agonist catechol only engages the rotamer toggle switch [391,415,416]. The activation mechanism is linked to the disruption of the network of interactions in the ionic lock. According to Romo *et al.* [391] the ionic lock can exist in three states: closed (or locked), semi-open with a bridging water molecule, and open. Recently Bokoch *et al.* [367] suggested that the extracellular domains of  $\beta_2$ AR also take part in the activation process. NMR spectroscopy and X-ray studies showed the functional role of the extracellular surface in ligand-specific conformational changes around a salt bridge linking D192 (EC2 loop) and K7.32 of  $\beta_2$ AR. This connection is formed in the unliganded  $\beta_2$ AR as well as in the complex with an inverse agonist – carazolol [298,339] but certain agonists (including formoterol, a molecule structurally related to fenoterol) were reported to induce its breaking. Our results show that (RR)-fenoterol isomer forms well preserved interaction between the hydroxyl group of the ligand's phenolic moiety and the carboxyl group of D192 (EC2 loop) residue during MD simulations in a similar manner as suggested by Bokoch *et al.* [367]. Such interaction is not observed in simulations on  $\beta_2$ AR - (S,S)-fenoterol complex.

Presence of the ionic lock switch is not questionable in nearly all GPCRs, however its influence for basal receptor activity is not obvious. Using D6.30N mutant of CB<sub>1</sub> and CB<sub>2</sub> receptors Nebane *et al.*

[417] found that D6.30 is essential for full activation of both cannabinoid receptors. Both CB<sub>1</sub> and CB<sub>2</sub> D6.30N mutants demonstrated a level of constitutive activity not greater than that of their wild type counterparts, indicating that either D6.30 does not participate in a salt bridge with R3.50 (DRY motif), or the salt bridge is not critical for keeping cannabinoid receptors in the inactive conformation. The same conclusion was drawn from 2 ns simulations of empty  $\beta_2$ AR conducted to resolve the question about the open ionic lock in the crystal structure of this receptor with inverse agonist bound [388]. It turned out that ionic lock in empty receptor may be open or not indicating two distinct *inactive* conformations of  $\beta_2$ AR in the basal state.

### 10.7.4 Conclusions

The flexible docking procedure for cannabinoid receptor CB<sub>1</sub> and CB<sub>2</sub> and molecular dynamics simulations for  $\beta_2$ -adrenergic receptor revealed important differences in binding modes and provided additional insights into the activation processes of GPCRs. New structures of these receptors awaited due to progress in stabilization, overexpression and crystallization techniques may help explain to what extent the structural mimicry is applied to GPCRs during activation processes. The longer molecular dynamics simulations even beyond the microsecond and reaching millisecond timescale will provide additional information about dynamics of ligand-receptor complexes and possibly also the course of the action of particular micro-switches. Taken into account the hypothesis of an ensemble of receptor conformations in the basal state one can also consider an ensemble of possible conformations of flexible ligands inside the binding site of the receptor, like anandamide in cannabinoid receptors and (S,S)-fenoterol in  $\beta_2$ AR, leading to dynamical binding modes in case of some ligand-receptor pairs that also could be investigated using simulations techniques.

## 10.8 Water mediated activation mechanism of Formyl peptide receptor 1 (FPR1)

### 10.8.1 Introduction

Human N-formyl peptide receptors (FPRs) are G protein-coupled receptors (GPCRs) involved in many physiological processes, including host defense against bacterial infection and resolving inflammation [418–424]. The three human FPRs (FPR1, FPR2 and FPR3) share significant sequence homology and perform their action via coupling to G<sub>i</sub> protein. Activation of FPRs induces a variety of responses, which are dependent on the agonist, cell type, receptor subtype, and also species involved. FPRs are expressed mainly by phagocytic leukocytes. Together, these receptors bind a large number of structurally diverse groups of agonistic ligands, including N-formyl and nonformyl peptides of different composition, that chemoattract and activate phagocytes. For example, N-formyl-Met-Leu-Phe (fMLF), an FPR1 agonist, activates human phagocyte inflammatory responses, such as intracellular calcium mobilization, production of cytokines, generation of reactive oxygen species, and chemotaxis [425]. This ligand can efficiently activate the major bactericidal neutrophil functions and it was one of the first characterized bacterial chemotactic peptides [426]. Whereas fMLF is by far the most frequently used chemotactic peptide in studies of neutrophil functions, atomistic descriptions for fMLF-FPR1 binding mode are still scarce mainly because of the absence of a crystal structure of this receptor. Elucidating the binding modes may contribute to designing novel and more efficient non-peptide FPR1 drug candidates. Molecular modeling of FPR1, on the other hand, can provide an efficient way to reveal details of ligand binding and activation of the receptor. However, recent modelings of FPRs were confined only to bovine rhodopsin [297,427] as a template.

Recently, Fujita *et al.* [428] investigated binding of calpain inhibitors as well as short peptides including fMLF to FPR1 and FPR2 receptors. Their findings suggest that potent calpain inhibitors could stimulate phagocyte functions via activation of FPR1, FPR2 and/or other G-protein coupled receptors depending on the inhibitors used. Using molecular docking they obtained different binding modes of fMLF in the above receptors and compared qualitatively the estimated energies of ligand binding to experimental data. They also provided a list of residues in a vicinity of a ligand but they did not show ligand-receptor interactions in the binding site. In another paper, Khlebnikov *et al.* [429] investigated binding of a set of benzimidazole derivatives as well as other agonists of FPR1 including fMLF. After the docking the 2 ns molecular dynamics (MD) simulations confined to the binding site were conducted. The rest of the FPR1 structure was kept rigid. In the best scored pose of fMLF-FPR1 the C-terminus of the ligand interacted with R205<sup>5.42</sup> while the formylated N-terminus interacted with main chains of residues L198<sup>5.35</sup>-V200<sup>5.37</sup> which could suggest that this part of the helix was unfolded. In another report Movitz *et al.* [430] identified the shortest sequence of the FPR1

ligand annexin A1 [431] which was still able to activate FPR1 and they also investigated the binding modes of this tetrapeptide. The Gln<sup>9</sup>-Phe<sup>12</sup> (Ac-QAWF) peptide was the shortest peptide of annexin A1 possessing the capacity both to trigger a neutrophil NADPH oxidase response and to inhibit the activity induced by other FPR agonists. Two alternative binding modes of Ac-QAWF were found having the same position of N-terminus of this peptide close to residues D106<sup>3,33</sup>, R201<sup>5,38</sup> and R205<sup>5,42</sup>. However, in both configurations there was no interaction with R86<sup>2,65</sup> predicted to be a part of the binding site for fMLF by the mutagenesis experiments [432]. In all the above studies the rhodopsin structure was taken as a template and no molecular dynamics simulations of the receptor in the membrane were performed to investigate an influence of the ligand on the receptor structure.

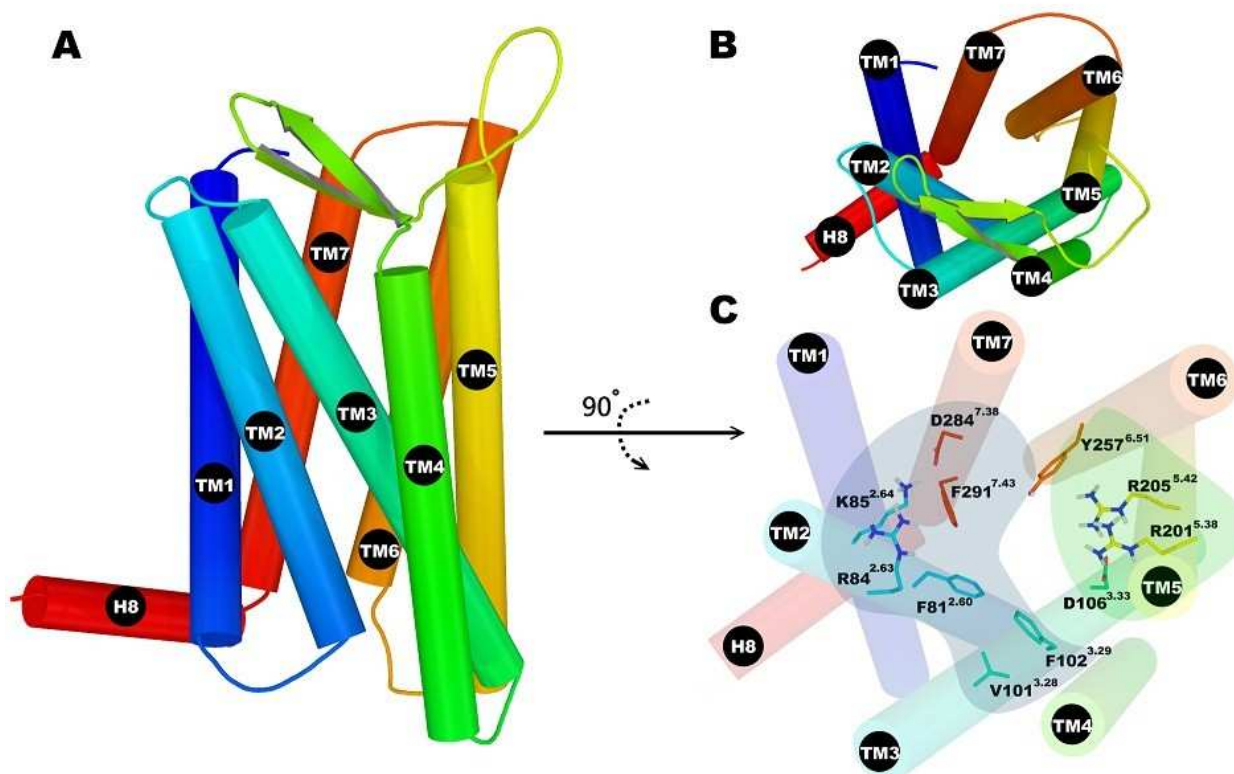
To locate specific ligand-receptor interactions based on a more appropriate template than rhodopsin we generated the homology models of FPR1 using the crystal structure of the chemokine receptor CXCR4 [316], which shares over 30% sequence identity with FPR1 and is located in the same  $\gamma$  branch of phylogenetic tree of GPCRs (rhodopsin is located in  $\alpha$  branch). Docking and model refinement procedures were pursued afterward. Finally, 40 ns full-atom MD simulations were conducted for the Apo form as well as for complexes of fMLF (agonist) and tBocMLF (antagonist) with FPR1 in the membrane. Based on locations of the N- and C-termini of the ligand the FPR1 extracellular pocket can be divided into two zones, namely, the anchor and activation regions. The formylated M1 residue of fMLF bound to the activation region led to a series of conformational changes of conserved residues. Internal water molecules participating in extended hydrogen bond networks were found to play a crucial role in transmitting the agonist-receptor interactions. A mechanism of initial steps of the activation concurrent with ligand binding is proposed.

## 10.8.2 Results

### FPR1 structure and the binding pocket

Currently, in the  $\gamma$  branch of the most populated family A of GPCRs there are five receptors whose structure has been determined, the chemokine receptor CXCR4 [316] and opioid receptors:  $\mu$ OR [433],  $\delta$ OR [434],  $\kappa$ OR [435] and the nociceptin FQ receptor [436]. For the homology modeling of FPR1 we used the one most similar in sequence and the closest in the phylogenetic tree, the chemokine receptor. The model obtained for the FPR1 structure consists of a seven transmembrane (TM) helix bundle (TM1 to TM7), a cytosol helix H8 and a  $\beta$ -hairpin loop between TM4 and TM5 (Fig 10.12A). Although the structure of CXCR4 does not contain helix H8 it exists in all crystal structures of opioid receptors which suggests that H8 is unfolded in the crystal of CXCR4 because of crystal packing. The model of FPR1 was relaxed in a POPE membrane using detailed relaxation procedure in Desmond program and subjected to ligand docking. The fMLF binding site of modeled

FPR1 is quasi symmetrical (Figure 10.12C). At both ends of the binding site there are positively charged residues: R84<sup>2.63</sup> and K85<sup>2.64</sup> located in TM2 as well as R201<sup>5.38</sup> and R205<sup>5.42</sup> on helix TM5. They are complemented by negatively charged residues: D284<sup>7.38</sup> in TM7 interacting with K85<sup>2.64</sup> and, at the other end, D106<sup>3.33</sup> in TM3 interacting with R201<sup>5.38</sup>. However, D106<sup>3.33</sup> is located much deeper in the receptor structure than D284<sup>7.38</sup> and is tightly interacting with R201<sup>5.38</sup>. Between both areas there are hydrophobic residues separating these charged areas and also interacting with the ligand. They can also be divided into two zones: residues F81<sup>2.60</sup>, V101<sup>3.28</sup> and F102<sup>3.29</sup> on helices TM2 and TM3 are located on one side of the ligand whereas Y257<sup>6.51</sup> and F291<sup>7.43</sup> on helices TM6 and TM7 on the other side. All abovementioned residues are located in the vicinity of 4 Å from the ligand. Because of such distribution of residues the entrance to the binding site is nearly uniformly positively charged so that the negatively charged ligands will be selectively attracted and in case of both agonist fMLF and antagonist tBocMLF (Fig 10.13) they would enter the binding site most preferably with the negatively charged C-terminus. The residue D106<sup>3.33</sup> is buried under R201<sup>5.42</sup> and is not visible in the figure. The red spot of negative potential in the center of the receptor comes from residue N108<sup>3.35</sup>. To facilitate comparison of our structure to other GPCRs the Ballesteros-Weinstein numbering scheme [265] was used (numbers in superscript) apart from the sequence numbers of FPR1 residues.

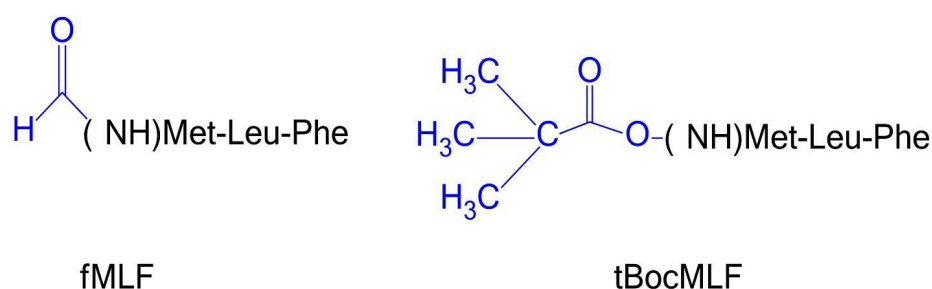


**Fig 10.12 The structure of homology model of FPR1 and its binding pocket.** (A) overall view of FPR1 model; (B) alternative view of FPR1 model from extracellular side; (C) important residues in binding site of FPR1. The whole pocket was visually divided into two zones: the anchor region (on left - in blue) and the activation region (on right - in green).

### Interactions of ligands with binding site of FPR1

The three best scored binding configurations of the 2000 conformations for agonist and antagonist, respectively, were characterized by the C-terminus of both ligands bound to the charged area at TM2 (the anchor region) whereas the N-terminus of the peptide was bound to the second charged area at TM5 (the activation region). Employing docking and the Glide docking score we found that the agonist fMLF was bound more strongly to the binding site of FPR1 (-7.8 kcal/mol) than the antagonist tBocMLF (-7.2 kcal/mol). The same hydrophobic residues of both ligands can suggest similar preferential binding modes. Next we conducted equilibration calculations of both complexes in a model of POPE membrane. After equilibration the C-terminal residue F3 of the agonist was engaged in a stable hydrogen bond network (Fig 10.14A) formed by the side chains of R84<sup>2.63</sup>, K85<sup>2.64</sup> and D284<sup>7.38</sup> while a water molecule mediated the hydrogen bonds between fMLF carbonyl group and D284<sup>7.38</sup>. The hydrophobic side chain of residue F3 was surrounded by four hydrophobic residues, namely F81<sup>2.60</sup>, V101<sup>3.28</sup>, F102<sup>3.29</sup> and F291<sup>7.43</sup>. Similarly to the agonist in the C-terminal region, the antagonist tBocMLF also formed the hydrogen bonds directly with R84<sup>2.63</sup> and K85<sup>2.64</sup> (Figure 5A), while the hydrophobic side chain of F3 was also stabilized by residues F81<sup>2.60</sup>, V101<sup>3.28</sup>, F102<sup>3.29</sup> and F291<sup>7.43</sup>. Differently from the agonist a hydrogen bond of tBocMLF with D284<sup>7.38</sup> was not created or even bridged by a water molecule but instead by a NH group of the peptide bond in residue F3 which formed a hydrogen bond with D284<sup>7.38</sup> directly.

At the other end of the two ligands the N-terminal formyl group of the agonist (Figure 10.14B) was involved in a complex water-mediated hydrogen bond network including residues R205<sup>5.42</sup> and D106<sup>3.33</sup> while the carbonyl group of the peptide bond in residue M1 formed a hydrogen bond with Y257<sup>6.51</sup>. Similarly to the agonist no direct interactions with charged residues of the receptor were found in the N-terminus of tBocMLF and only a water mediated hydrogen bond network was located between the carbonyl group of tBoc and both arginine residues R201<sup>5.38</sup> and R205<sup>5.42</sup>. Moreover, there was also a direct hydrogen bond between Y257<sup>6.51</sup> and the main chain of the antagonist.



**Fig 10.13 The chemical formulas of fMLF (agonist) and tBocMLF (antagonist).** Both ligands share most of the structure so only differences in N-termini are shown in detail and colored in blue.

To investigate the changes in FPR1 structure that can be induced concurrently with agonist binding

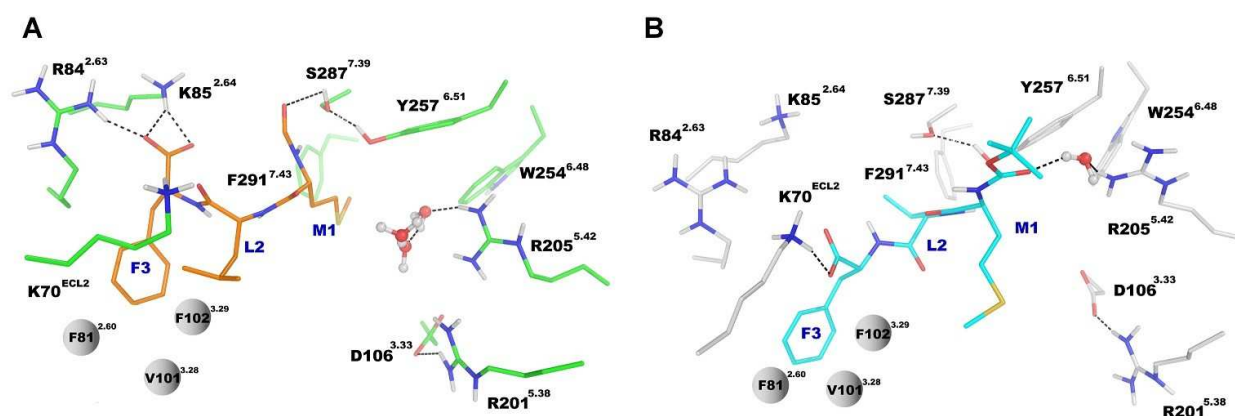


we performed 40 ns MD simulations starting from systems equilibrated in a model membrane. The simulations were conducted for FPR1 in its Apo form, as well as for complexes with agonist and antagonist. The root mean squares deviation (RMSD) plots of protein backbone show small rearrangements (0.7 Å) compared to the starting structures so the investigated structures were stable as early as at 5 ns of MD simulation indicating that equilibration procedure was sufficient to stabilize the receptor. The binding pocket remained similar to the starting conformations in all conducted simulations suggesting that only local rearrangements took place at least at this time-scale.

During the simulations both agonist and antagonist changed their positions, however, the agonist stayed bound to the anchor region for the whole simulation while the antagonist unbound and finally its charged C-terminus interacted directly with K170 from a long EC2 loop (between TM4 and TM5). Additionally the benzene ring of F3 of antagonist formed  $\pi$ - $\pi$  stacked interactions with W91 of the EC1 loop. In case of the agonist the side chain of F3 was stably located between F81<sup>2,60</sup>, W91<sup>EC1</sup> and F102<sup>3,29</sup> (Fig 10.14A). At the N-terminus of the antagonist there was a large movement of residue M1 from an interior position toward EC2 and especially residue F178. The tBoc group did not change much its position but a hydrogen bond to Y257<sup>6,51</sup> was lost (Fig 10.14B). In case of the agonist there was also a change of the M1 side chain but here towards the interior of FPR1 close to position previously occupied by formyl group of this ligand i.e. close to residues R205<sup>5,42</sup>, Y257<sup>6,51</sup> and W254<sup>6,48</sup>. M1 also displaced one water molecule and stayed close also to F291<sup>7,43</sup>. The formyl group interacted with S287<sup>7,39</sup> and indirectly with Y257<sup>6,51</sup>. The electrostatic interactions between D106<sup>3,33</sup> and both arginine residues, R201<sup>3,38</sup> and R205<sup>5,42</sup>, were stable in the Apo form of FPR1. However, for antagonist and agonist the interaction D106<sup>3,33</sup>-R201<sup>3,38</sup> was broken and restored many times. For both ligands the residue R205<sup>5,42</sup> moved away from D106<sup>3,33</sup> but in case of agonist it was separated by one water molecule only.

Similarly to other structures of GPCRs in partly activated state a hydrogen bond network has been found throughout the whole transmembrane region of FPR1 (Fig 10.15). This network started from W254<sup>6,48</sup> and consisted of residues N108<sup>3,35</sup> (the residue also present in CXCR4 and opioid receptors but not in muscarinic receptors), D71<sup>2,50</sup> (the most conserved residue in TM2), N297<sup>7,49</sup> and Y301<sup>7,53</sup> (from the NPxxY motif). The above residues were connected directly by hydrogen bonds. Y301<sup>7,53</sup> formed  $\pi$ - $\pi$  stacking interaction with Y64<sup>2,43</sup> but also participated in water-mediated hydrogen bond networks involving also residues at the cytoplasmic part of the receptor: Y64<sup>2,43</sup>, D122<sup>3,49</sup> and R123<sup>3,50</sup> (from the DRC motif – corresponding to DRY in other GPCRs) and also R137<sup>4,37</sup> interacting directly with D122<sup>3,49</sup> (Fig 10.15B). During MD simulation of the fMLF-FPR1 complex a water molecule initially located between R205<sup>5,42</sup> and W254<sup>6,48</sup> diffused toward the center of FPR1 and transiently (4-18 ns in MD simulation) bridged the hydrogen bond between W254<sup>6,48</sup> and N108<sup>3,35</sup>. A similar scenario was observed at the other side of FPR1 when water molecule from the cytoplasmic

side diffused into the receptor and bridged a hydrogen bond between N298<sup>7.49</sup> and Y301<sup>7.53</sup> in NPxxY motif. This bridging was stable till the end of simulation. The network of interactions between residues of the agonist-receptor complex as well as a movement of bridging water molecules is depicted schematically in Fig 10.16.



**Fig 10.14 The ligand-receptor interactions after 40 ns MD simulation.** View from extracellular side. (A) The agonist fMLF (in orange). (B) The antagonist tBocMLF (in cyan). The M1 residue of agonist went down toward W254<sup>6.48</sup> while that of antagonist went up toward EC2 loop.

### Models of FPR2 and FPR3 receptors

To obtain models of the related receptors FPR2 and FPR3 we performed homology modeling based on the equilibrated structure of FPR1. The FPR2, which shares 69% sequence identity with FPR1, is a low affinity receptor for fMLF with a  $K_d$  of 430 nM [437–439]. The obtained model of FPR2 showed many differences compared to FPR1 including residues in the binding site: (FPR1 to FPR2) F81L<sup>2.60</sup>, R84S<sup>2.63</sup>, K85M<sup>2.64</sup>, F102H<sup>3.29</sup>, Y257F<sup>6.51</sup> and D284N<sup>7.38</sup> (Figure 10.16A). Since K85<sup>2.64</sup> and R84<sup>2.63</sup> has been experimentally proven to be crucial for fMLF binding [432], the mutations at these two positions in FPR2 might be responsible for the low binding affinity of fMLF. We also performed docking of this agonist and the obtained scores had indicated that fMLF binding in FPR1 is more favorable than in FPR2 with scores 7.8 kcal/mol and -6.1 kcal/mol, respectively. The FPR3, which shares 56% sequence identity with FPR1, binds the agonist fMLF with undetectable affinity [432]. The obtained homology model of FPR3 also exhibited many differences including residues in the binding site: (FPR1 to FPR3) F81R<sup>2.60</sup>, R84S<sup>2.63</sup>, K85V<sup>2.64</sup>, F102H<sup>3.29</sup> and D284N<sup>7.38</sup> (Fig 10.17B). The loss of fMLF binding affinity can be attributed to the mutations K85V<sup>2.64</sup> and R84S<sup>2.63</sup> both of which had been shown to be important for binding. Moreover, F81R<sup>2.60</sup> could also contribute to the lack of fMLF binding since hydrophobic properties were lost at the position where the F3 residue of the ligand is located. Furthermore, residues in the activation zone at positions 201 and 205, namely R205H<sup>5.42</sup> and R201F<sup>5.38</sup>, were also found to have properties different from FPR1 indicating that the activation must be performed in another way than for FPR1 and FPR2.

### 10.8.3 Discussion

#### The choice of CXCR4 as template structure

Our study is the first attempt, to our knowledge, to show changes in the molecular structure of FPR1 that occur upon agonist binding. The structure was constructed based on a novel template, the chemokine receptor CXCR4, belonging to the same  $\gamma$  branch of the phylogenetic tree of GPCRs as the formyl receptors. Molecular dynamics simulations were conducted including an all-atom model of the membrane. Because there are two structures of CXCR4 complexed with different antagonists we chose the one in which helices are not distorted by the presence of detergent used for crystallization. The structure with small agonist (PDB id 3ODU) contains two detergent molecules between helices TM5 and TM6. They do not affect the binding of antagonist IT1t of CXCR4 which is bound mainly to helices TM2, TM3 and TM7. In case of FPR1 the experimental evidence is that TM5 participates extensively in binding of agonists and antagonists. Therefore, we decided to use the structure of CXCR4 complexed with cyclic peptide CVX15 being also an antagonist of this receptor (PDB id 3OE0) in spite of its lower resolution 3.2 Å compared to 2.5 Å of the structure with IT1t.

#### Comparison of structures based on rhodopsin and CXCR4 templates

Earlier modeling attempts of FPR1 [428–430] were all based on the rhodopsin template. There are several important differences between the rhodopsin and CXCR4 structures which can affect homology modeling and binding of ligands. First, the EC2 loop is outside the binding site of CXCR4 so there is much more space for binding of ligand, and second, there is a bulge at extracellular part of TM2 of rhodopsin (located at positions S76<sup>2.55</sup> and T77<sup>2.56</sup> of FPR1) which is not present in the CXCR4 structure. Using the CXCR4 template this part of TM2 is rotated about 100° compared to the rhodopsin template so another part of TM2 is facing the binding site (Fig 10.17). Especially residue R84<sup>2.63</sup> which was predicted, based on rhodopsin structure, to be outside the binding site can now interact with ligand jointly with K85<sup>2.64</sup>. Interestingly, these both residues were predicted by Mills *et al.* [432] to strongly interact with ligands of FPR1. Additional confirmation of the obtained structure is a presence on a salt bridge between K85<sup>2.64</sup> and D284<sup>7.38</sup> which was proposed by Mills based on site-specific fluorescent photoaffinity labeling and mass spectrometry [432]. The mutual location of helices other than TM2 is also different in both templates so the binding site is dissimilar enough to prefer other ligand binding modes. The presence of the bulge in TM2 in rhodopsin could severely influence the structure and interactions in the binding site of homology models and it was one of major reasons for very poor docking results during modeling of complex of CXCR4 structure during GPCR Dock 2010 assessment [440].

According to the location of residues in contact with docked ligands, the binding pocket of FPR1 can

be visually divided into two zones: the activation zone where the modified N-terminus of fMLF and tBocMLF is bound, and the binding zone where the C-terminus of ligand is bound. The nearly symmetrical binding site of FPR1 enables reversed binding configurations of agonist and antagonist and one cannot exclude that initially the ligands can bind in both ways with N-terminus docked either to TM5 or to TM2. However, only one way can be appropriate for activation of the receptor. The receptor binding site is more spacious close to TM5 because this helix is located farther from the EC2 loop and this could be the reason for preferential docking of tBoc moiety to this area as well as further unbinding of the C-terminus of the antagonist from TM2 during MD simulation. fMLF was stably bound to TM2 during the entire MD simulation. Such findings are also supported by experimental data because fMLF shows higher binding affinity than tBocMLF in the case of native FPR1.

### **Binding of tripeptide and tetrapeptide ligands**

Movitz *et al.* [430] identified a tetrapeptide of the ligand annexin A1, Gln<sup>9</sup>-Phe<sup>12</sup> (Ac-QAWF), as the shortest sequence of annexin A1 which is still able to activate FPR1. Although the modeling that was also performed in this paper was based on the rhodopsin template the authors proposed a possible binding mode of this peptide. It directly binds to both TM2 and TM5 and spans across the entire binding site. Tripeptides like fMLF are shorter than the Ac-QAWF so the binding mode must be different. We found that the water molecules can bridge interactions between N-terminus of agonist and charged residues D106<sup>3.33</sup>, R201<sup>5.38</sup> and R205<sup>5.42</sup> and therefore they can participate in the activation process. Based on experimental results [441–444], four hydrophobic residues, namely F81<sup>2.60</sup>, V101<sup>3.28</sup>, F102<sup>3.29</sup> and F291<sup>7.43</sup>, had been shown to be important for fMLF binding. Moreover, the residues Y257<sup>6.51</sup>, K85<sup>2.64</sup> and R84<sup>2.63</sup> [432,444] were also identified by mutagenesis to have a significant effect on FPR1 binding affinity, while D106<sup>3.33</sup> [432,443], R201<sup>5.38</sup> and R205<sup>5.42</sup> [444] were confirmed to be crucial for FPR1 activation. All these residues were found in close vicinity of the docked, optimized and simulated agonist fMLF.

During MD simulation the side chain of the M1 residue of the agonist went down toward the center of the receptor close to residues regarded to be crucial for activation of most of GPCRs including W254<sup>6.48</sup> from CWxP motif in helix TM6. This residue participates in the so called transmission switch, the action of which leads to rearrangements of residues in the central part of GPCRs and is a prerequisite for outward movement of cytoplasmic part of helix TM6 (a recent review on the action of molecular switches in GPCRs can be found in [295]). Contrary to the agonist the side chain of residue M1 in tBocMLF was displaced toward the EC2 loop. There is a similarity of the location of the side chain of the first amino acid of fMLF and of the tetrapeptide Ac-QAWF. In both cases this side chain is located in close vicinity of W254<sup>6.48</sup>. The tetrapeptide was manually docked in [430] to preserve interactions with residues in TM3 and TM5 known to participate in activation. In our

simulation the formyl part of the M1 residue was initially bound to TM5 via bridging water molecules while the side chain of M1 was outside of this region (Figure 4B and 7A) but after MD simulation the side chain of M1 was localized much deeper than the formyl group (Figure 6 and 8) similarly to the binding mode of Ac-QAWF proposed by Movitz *et al.* [430].

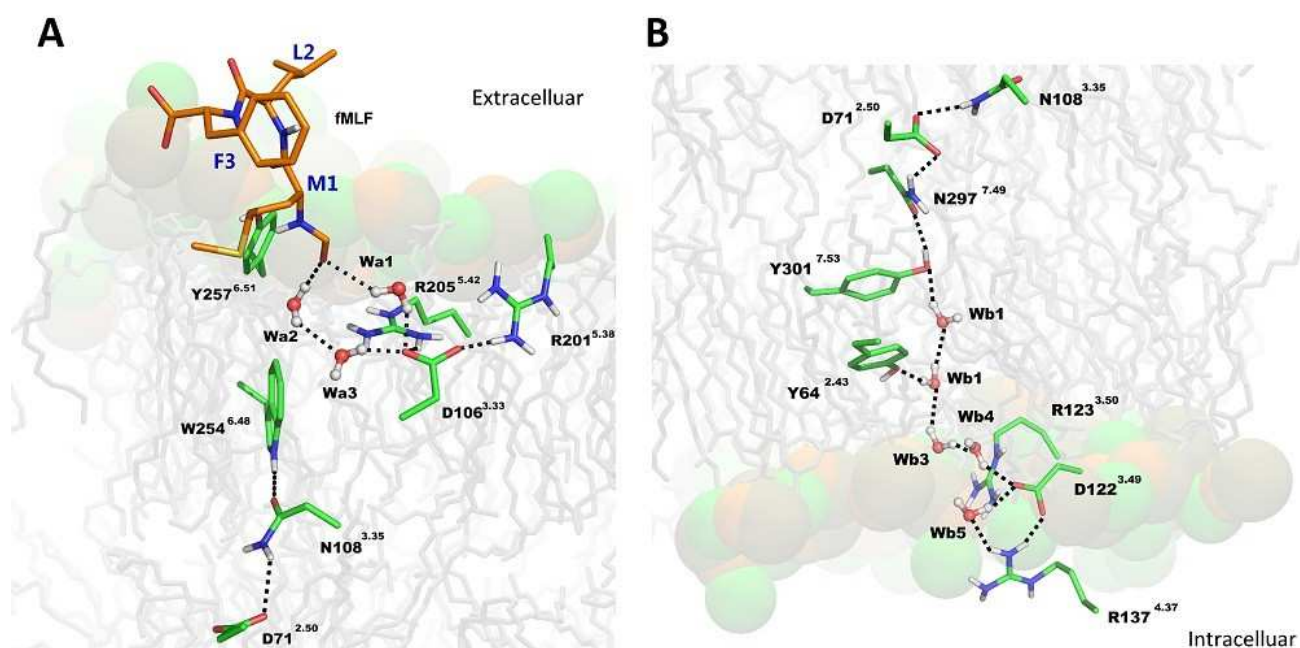
### Role of water molecules in ligand binding

Water molecules were found important also in a recent paper of Vanni *et al.* [445] in 800 ns MD simulation of  $\beta_2$ -adrenergic receptor. They bridged interactions between agonists and serines located in TM5 while the ligands were closely bound to D113<sup>3.32</sup> in TM3 with their protonated amine group. Displacement of these water molecules may be a step toward an activation of the receptor because it was found that the binding site of  $\beta_2$ -AR is shrinking during activation [367]. Two water molecules were also found to bridge interaction between phenolic hydroxyl groups of antagonists and the side chain of H(6.52) in three crystal structures of opioid receptors  $\mu$ OR,  $\delta$ OR and  $\kappa$ OR. Identical arrangements of these water molecules in three different receptors suggest that their presence is crucial to stabilize the antagonist and possibly they participate in receptor activation when an agonist is bound. In our earlier papers on activation of opioid receptors [349–351] we postulated, based on MD simulations, that antagonists can bind to residues in TM3, namely D(3.32) and Y(3.33), but agonists can swap from Y(3.33) to H(6.52) in helix TM6 and such change of location is probably one of the first activation steps. Since no structures of opioid receptors with agonists are available, this hypothesis still needs to be validated. Possibly, during the activation these water molecules are displaced and the agonist can bind directly to H(6.52). This can shrink the binding site and facilitate rearrangement of residues of the central part of the receptor being a part of the transmission switch. This switch was previously called the rotamer toggle switch and was linked only to residue W(6.48), however, the suggested action of this switch was not confirmed by later crystal structures of GPCRs with agonists.

During MD simulation of fMLF-FPR1 we observed a movement of water molecules toward the receptor center from the extracellular as well as from the cytoplasmic side (Fig 10.15). Because of the small simulation time one cannot postulate these movements as being part of activation process but rather as thermal movements of water tending to populate the receptor interior because the initial structure of the receptor was free of individual water molecules separated from bulk water. In a recent structure of the muscarinic receptor M2 [446] there is an aqueous channel extending from the extracellular surface into the transmembrane core with well-ordered water molecules. This channel is interrupted by a layer of hydrophobic residues located in helices TM2, TM3 and TM6 close to residue Y(7.53) in NPxxY motif. Although the Tyr toggle switch is in an active state (i.e. the side chain of Y(7.53) is directed toward the receptor center contrary to that residue in the rhodopsin structure which is directed toward a cytoplasmic helix H8) [295] there is no a hydrogen bond

network linking Y(7.53) with N(7.49). Possibly, after the action of the transmission switch in the muscarinic M2 receptor the channel will be rearranged and an extended hydrogen bond network will connect both sides of the receptor to enable final stages of receptor activation. Such an extended network of hydrogen bonds involving water molecules crossing the hydrophobic barrier was found recently in the structure of constitutively active rhodopsin [354]. In the model of FPR1 we also found an extended network of hydrogen bonds (Fig 10.15). Such network was broken at residue Y301<sup>7.53</sup> since it created a  $\pi$ - $\pi$  stack interaction with Y64<sup>2.43</sup>. However, because the switch involving Y301<sup>7.53</sup> is already in its active state and Y301<sup>7.53</sup> forms a hydrogen bond with N297<sup>7.49</sup> (from the same NPxxY motif) the water molecules can easily pass the hydrophobic barrier even in a relatively short 40 ns MD simulation.

To resolve unanswered questions of activation details and ligand docking as well as ligand selectivity the MD simulations in a microsecond time scale have to be conducted, preferably based on the solved crystal structures of FPRs. Knowledge of these structures and the activation processes initiated by binding of the diverse ligands will lead to better understanding of mechanisms of action of these highly elusive receptors and also to a design of safer and more efficient drugs.



**Fig 10.15 A hydrogen bond network in the structure of the agonist-FPR1 complex.** A side view of initial equilibrated structure. (A) The binding site showing the hydrogen bond network involving water molecules. (B) A continuation of the hydrogen bond network of the same complex at intracellular side.

Fig 10.16 A scheme of interactions in the final structure of the agonist-FPR1 complex after MD simulation. A movement of two water molecules during MD simulation is shown. These molecules can bridge the hydrogen bonds between some residues. A water molecule transiently bridge a hydrogen bond between W254<sup>6,48</sup> and N108<sup>3,35</sup> while another water molecule from cytoplasmic side enters the receptor cavity and can bridge the hydrogen bond between residues N297<sup>7,49</sup> and Y301<sup>7,53</sup> from NPxxY motif.

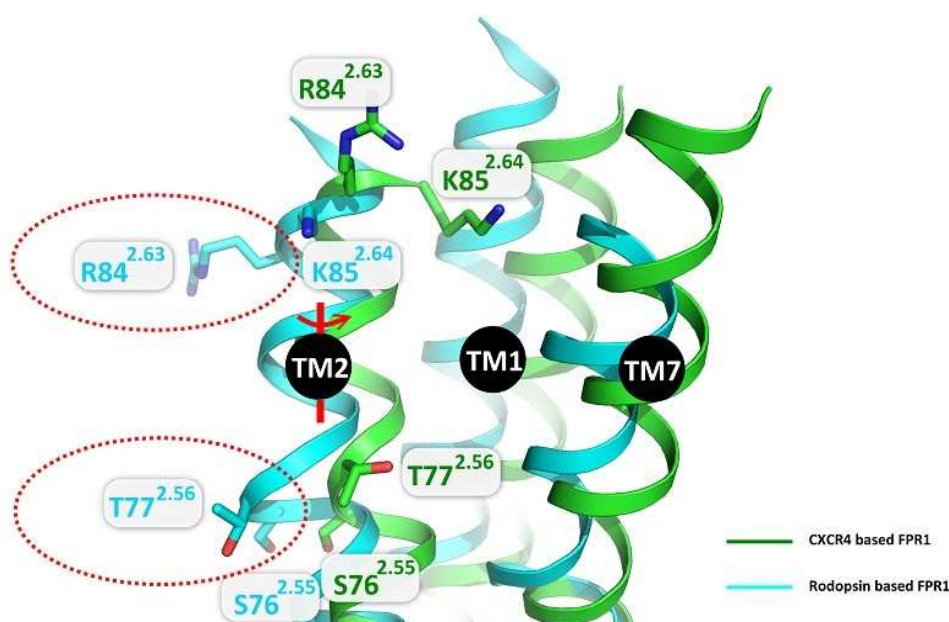
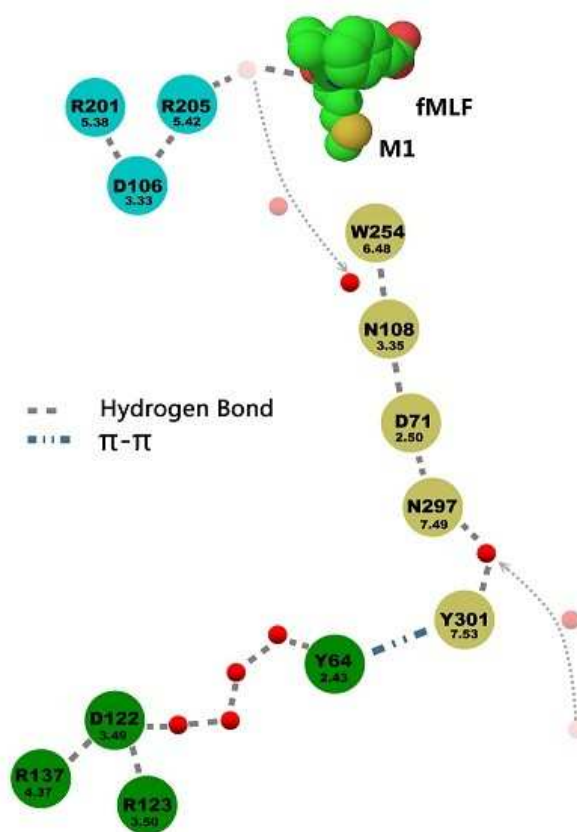


Fig 10.17 Comparison of FPR1 models constructed on different templates. A model based on rhodopsin is colored in cyan while that based on CXCR4 in green. Some residues in TM2 are shown in red dashed ellipses to exemplify differences between both models. A change of a template from rhodopsin to CXCR4 leads to the rotation about 100° of extracellular part of TM2 starting from S76<sup>2,55</sup> and removal of a bulge at T77<sup>2,56</sup>.

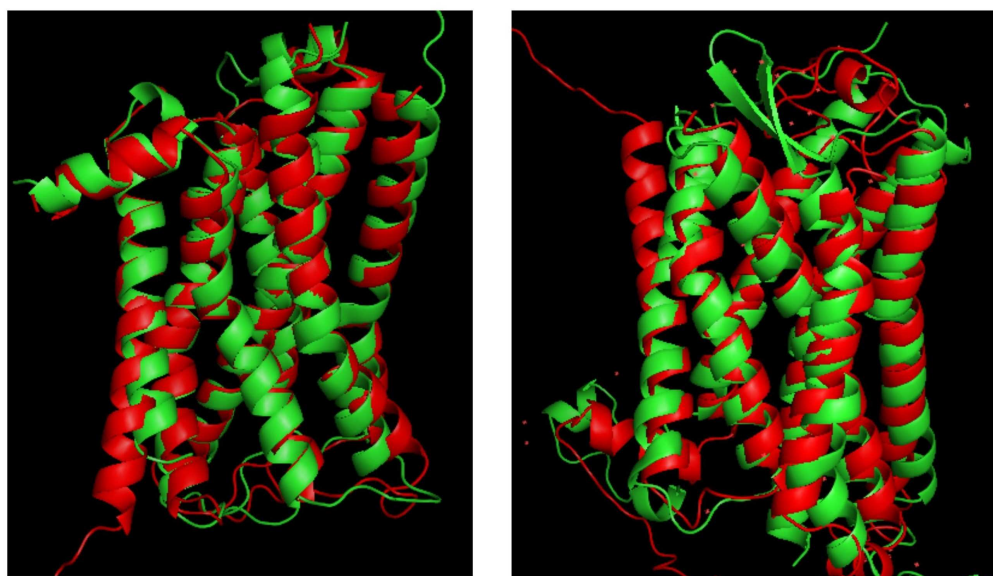
## 10.9 GPCR Dock 2010 Results

Results from the competition are given in the following table and figure showing the RMSD differences of the models and crystal structures.

**Table 2. GPCR Dock 2010 Models that Score One or More SD above the Average**

Group	Model	TM Rmsd, Å (# residues)	Fraction TM Superimposed, Å (rmsd)	ECL2 Rmsd (Å)	Pocket Rmsd, Å (# residues)	TM Pocket Rmsd, Å (# residues)	W94, D97 Rotation (°)	Fraction of Pocket Predicted	Ligand Rmsd (Å)	Atomic Contacts (# residues)/ Reference	Correct Contact Strength (% reference)	Critical Contacts	Z-Score
D3/Eticlopride													
PompeuFabra	3	1.38 (205)	63% (0.52)	2.87	1.50 (15)	1.16 (14)	n/a	63%	0.96	36 (9)/65 (15)	39.95 (58%)	D110	2.39
CDD-CMBI	1	1.95 (205)	66% (0.66)	4.38	2.25 (15)	1.67 (14)	n/a	50%	2.13	36 (12)/65 (15)	38.75 (57%)	D110	2.05
COH-Vaidehi	1	1.58 (205)	69% (0.79)	4.11	2.04 (15)	1.36 (14)	n/a	59%	1.22	31 (10)/65 (15)	34.78 (51%)	D110	2.00
UCSF-Shoichet-1	4	1.45 (205)	65% (0.60)	3.53	1.53 (15)	1.41 (14)	n/a	74%	1.23	29 (10)/65 (15)	32.46 (47%)	D110	1.85
Schrödinger	5	1.41 (205)	60% (0.53)	12.57	1.57 (15)	1.48 (14)	n/a	38%	1.77	23 (11)/65 (15)	30.94 (45%)	D110 <sup>a</sup>	1.63
CDD-CMBI	2	1.95 (205)	67% (0.66)	4.38	2.24 (15)	1.67 (14)	n/a	46%	2.27	26 (11)/65 (15)	30.5 (44%)	D110 <sup>a</sup>	1.49
CDD-CMBI	3	1.95 (205)	66% (0.65)	4.35	2.17 (15)	1.59 (14)	n/a	43%	2.38	30 (11)/65 (15)	30.1 (44%)	D110	1.44
Warsaw	4	2.13 (205)	62% (0.64)	4.71	1.68 (15)	1.64 (14)	n/a	52%	2.34	27 (9)/65 (15)	26.47 (39%)		1.22
CDD-CMBI	4	1.94 (205)	67% (0.66)	4.17	1.99 (15)	1.62 (14)	n/a	35%	2.13	19 (8)/65 (15)	24.84 (36%)		1.16
CDD-CMBI	5	1.94 (205)	67% (0.67)	4.00	1.87 (15)	1.67 (14)	n/a	41%	2.44	21 (8)/65 (15)	25.57 (37%)	D110 <sup>a</sup>	1.14
Helsinki-Xhaard	1	1.55 (205)	67% (0.67)	3.74	1.43 (15)	1.29 (14)	n/a	47%	3.42	21 (7)/65 (15)	28.95 (42%)	D110 <sup>a</sup>	1.13
COH-Vaidehi	2	1.57 (205)	70% (0.80)	4.01	2.16 (15)	1.50 (14)	n/a	57%	2.96	25 (7)/65 (15)	26.75 (39%)	D110	1.10
Monash-Sexton-2	3	1.61 (205)	63% (0.62)	4.81	2.47 (15)	2.30 (14)	n/a	52%	1.94	17 (6)/65 (15)	22.51 (33%)		1.05
Schrödinger	1	1.43 (205)	63% (0.55)	n/a <sup>b</sup>	1.50 (15)	1.42 (14)	n/a	61%	2.24	19 (10)/65 (15)	22.81 (33%)	D110	1.01

Fig 10.18: Performance of modeling groups in GPCR Dock 2010



D3 Receptor, RMSD: **2.13 Å**

CXCR4 Receptor, RMSD: **3.21 Å**

Fig 10.19: Superimposition of D3 and CXCR4 models with the crystal structure



## 11 References

- 1 Alzheimer A (1907) A New Disease of the Cortex, (Ger). *Allg. Z Psychiatr* **64**, 146–148.
- 2 2012 Alzheimer's disease facts and figures (2012) *Alzheimer's & Dementia* **8**, 131–168.
- 3 Sugimoto H (2010) *YAKUGAKU ZASSHI* **130**, 521–526.
- 4 Katzman R (ed.) (1983) *Biological Aspects of Alzheimer's Disease* Cold Spring Harbor Laboratory Pr.
- 5 Lichtenthaler SF, Haass C & Steiner H (2011) Regulated intramembrane proteolysis – lessons from amyloid precursor protein processing. *Journal of Neurochemistry* **117**, 779–796.
- 6 Erez E, Fass D & Bibi E (2009) How intramembrane proteases bury hydrolytic reactions in the membrane. *Nature* **459**, 371–378.
- 7 Lal M & Caplan M (2011) Regulated intramembrane proteolysis: signaling pathways and biological functions. *Physiology (Bethesda)* **26**, 34–44.
- 8 Brown MS, Ye J, Rawson RB & Goldstein JL (2000) Regulated intramembrane proteolysis: a control mechanism conserved from bacteria to humans. *Cell* **100**, 391–398.
- 9 Fluhrer R, Steiner H & Haass C (2009) Intramembrane proteolysis by signal peptide peptidases: a comparative discussion of GXGD-type aspartyl proteases. *J. Biol. Chem.* **284**, 13975–13979.
- 10 Freeman M (2009) Rhomboids: 7 years of a new protease family. *Semin. Cell Dev. Biol.* **20**, 231–239.
- 11 Friedmann E, Hauben E, Maylandt K, Schleegeer S, Vreugde S, Lichtenthaler SF, Kuhn P-H, Stauffer D, Rovelli G & Martoglio B (2006) SPPL2a and SPPL2b promote intramembrane proteolysis of TNFalpha in activated dendritic cells to trigger IL-12 production. *Nat. Cell Biol.* **8**, 843–848.
- 12 Faça VM, Ventura AP, Fitzgibbon MP, Pereira-Faça SR, Pitteri SJ, Green AE, Ireton RC, Zhang Q, Wang H, O'Brian KC, Drescher CW, Schummer M, McIntosh MW, Knudsen BS & Hanash SM (2008) Proteomic analysis of ovarian cancer cells reveals dynamic processes of protein secretion and shedding of extra-cellular domains. *PLoS ONE* **3**, e2425.
- 13 Wang Y, Zhang Y & Ha Y (2006) Crystal structure of a rhomboid family intramembrane protease. *Nature* **444**, 179–180.
- 14 Baker RP, Young K, Feng L, Shi Y & Urban S (2007) Enzymatic analysis of a rhomboid intramembrane protease implicates transmembrane helix 5 as the lateral substrate gate. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 8257–8262.
- 15 Wu Z, Yan N, Feng L, Oberstein A, Yan H, Baker RP, Gu L, Jeffrey PD, Urban S & Shi Y (2006) Structural analysis of a rhomboid family intramembrane protease reveals a gating mechanism for substrate entry. *Nat. Struct. Mol. Biol.* **13**, 1084–1091.
- 16 Barrett PJ, Song Y, Van Horn WD, Hustedt EJ, Schafer JM, Hadziselimovic A, Beel AJ & Sanders CR (2012) The amyloid precursor protein has a flexible transmembrane domain and binds cholesterol. *Science* **336**, 1168–1171.
- 17 Miyashita N, Straub JE & Thirumalai D (2009) Structures of beta-amyloid peptide 1-40, 1-42, and 1-55-the 672-726 fragment of APP-in a membrane environment with implications for interactions with gamma-secretase. *J. Am. Chem. Soc.* **131**, 17843–17852.
- 18 Miyashita N, Straub JE, Thirumalai D & Sugita Y (2009) Transmembrane structures of amyloid precursor protein dimer predicted by replica-exchange molecular dynamics simulations. *J. Am. Chem. Soc.* **131**, 3438–3439.
- 19 McCarthy JV, Twomey C & Wujek P (2009) Presenilin-dependent regulated intramembrane proteolysis and gamma-secretase activity. *Cell. Mol. Life Sci.* **66**, 1534–1555.
- 20 Kimberly WT, LaVoie MJ, Ostaszewski BL, Ye W, Wolfe MS & Selkoe DJ (2003) Gamma-secretase is a membrane protein complex comprised of presenilin, nicastrin, Aph-1, and Pen-2. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 6382–6387.
- 21 Wolfe MS, Xia W, Ostaszewski BL, Diehl TS, Kimberly WT & Selkoe DJ (1999) Two transmembrane aspartates in presenilin-1 required for presenilin endoproteolysis and gamma-secretase activity. *Nature* **398**, 513–517.
- 22 Ahn K, Shelton CC, Tian Y, Zhang X, Gilchrist ML, Sisodia SS & Li Y-M (2010) Activation and intrinsic gamma-secretase activity of presenilin 1. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 21435–21440.
- 23 Shih I-M & Wang T-L (2007) Notch signaling, gamma-secretase inhibitors, and cancer therapy. *Cancer Res.* **67**, 1879–1882.
- 24 Lee S-F, Shah S, Li H, Yu C, Han W & Yu G (2002) Mammalian APH-1 interacts with presenilin and nicastrin and is required for intramembrane proteolysis of amyloid-beta precursor protein and Notch. *J. Biol. Chem.* **277**, 45013–45019.
- 25 Yu G, Chen F, Levesque G, Nishimura M, Zhang DM, Levesque L, Rogueva E, Xu D, Liang Y, Duthie M, St George-Hyslop PH & Fraser PE (1998) The presenilin 1 protein is a component of a high molecular weight intracellular complex that contains beta-catenin. *J. Biol. Chem.* **273**, 16470–16475.
- 26 Serneels L, Van Biervliet J, Craessaerts K, Dejaegere T, Horrè K, Van Houtvin T, Esselmann H, Paul S, Schäfer MK, Berezovska O, Hyman BT, Sprangers B, Sciot R, Moons L, Jucker M, Yang Z, May PC, Karran E, Wiltfang J, D'Hooge R & De Strooper B (2009) gamma-Secretase heterogeneity in the Aph1 subunit: relevance for Alzheimer's disease. *Science* **324**, 639–642.

- 27 Shirotani K, Tomioka M, Kremmer E, Haass C & Steiner H (2007) Pathological activity of familial Alzheimer's disease-associated mutant presenilin can be executed by six different gamma-secretase complexes. *Neurobiol. Dis.* **27**, 102–107.
- 28 Lazarov VK, Fraering PC, Ye W, Wolfe MS, Selkoe DJ & Li H (2006) Electron microscopic structure of purified, active gamma-secretase reveals an aqueous intramembrane chamber and two pores. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 6889–6894.
- 29 Osenkowski P, Li H, Ye W, Li D, Aeschbach L, Fraering PC, Wolfe MS, Selkoe DJ & Li H (2009) Cryoelectron microscopy structure of purified gamma-secretase at 12 Å resolution. *J. Mol. Biol.* **385**, 642–652.
- 30 Renzi F, Zhang X, Rice WJ, Torres-Arancivia C, Gomez-Llorente Y, Diaz R, Ahn K, Yu C, Li Y-M, Sisodia SS & Ubarretxena-Belandia I (2011) Structure of gamma-secretase and its trimeric pre-activation intermediate by single-particle electron microscopy. *J. Biol. Chem.* **286**, 21440–21449.
- 31 Kim S-H & Sisodia SS (2005) Evidence that the “NF” motif in transmembrane domain 4 of presenilin 1 is critical for binding with PEN-2. *J. Biol. Chem.* **280**, 41953–41966.
- 32 Watanabe N, Tomita T, Sato C, Kitamura T, Morohashi Y & Iwatsubo T (2005) Pen-2 is incorporated into the gamma-secretase complex through binding to transmembrane domain 4 of presenilin 1. *J. Biol. Chem.* **280**, 41967–41975.
- 33 Fassler M, Li X & Kaether C (2011) Polar transmembrane-based amino acids in presenilin 1 are involved in endoplasmic reticulum localization, Pen2 protein binding, and  $\gamma$ -secretase complex stabilization. *J. Biol. Chem.* **286**, 38390–38396.
- 34 Kornilova AY, Kim J, Laudon H & Wolfe MS (2006) Deducing the transmembrane domain organization of presenilin-1 in gamma-secretase by cysteine disulfide cross-linking. *Biochemistry* **45**, 7598–7604.
- 35 Kornilova AY, Bihel F, Das C & Wolfe MS (2005) The initial substrate-binding site of gamma-secretase is located on presenilin near the active site. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 3230–3235.
- 36 Sato C, Takagi S, Tomita T & Iwatsubo T (2008) The C-terminal PAL motif and transmembrane domain 9 of presenilin 1 are involved in the formation of the catalytic pore of the gamma-secretase. *J. Neurosci.* **28**, 6264–6271.
- 37 Chiang P-M, Fortna RR, Price DL, Li T & Wong PC (2012) Specific domains in anterior pharynx-defective 1 determine its intramembrane interactions with nicastrin and presenilin. *Neurobiol. Aging* **33**, 277–285.
- 38 Spasic D & Annaert W (2008) Building gamma-secretase: the bits and pieces. *J. Cell. Sci.* **121**, 413–420.
- 39 Takasugi N, Tomita T, Hayashi I, Tsuruoka M, Niimura M, Takahashi Y, Thinakaran G & Iwatsubo T (2003) The role of presenilin cofactors in the gamma-secretase complex. *Nature* **422**, 438–441.
- 40 Kaether C, Scheuermann J, Fassler M, Zilow S, Shirotani K, Valkova C, Novak B, Kacmar S, Steiner H & Haass C (2007) Endoplasmic reticulum retention of the gamma-secretase complex component Pen2 by Rer1. *EMBO Rep.* **8**, 743–748.
- 41 Fassler M, Zocher M, Klare S, de la Fuente AG, Scheuermann J, Capell A, Haass C, Valkova C, Veerappan A, Schneider D & Kaether C (2010) Masking of transmembrane-based retention signals controls ER export of gamma-secretase. *Traffic* **11**, 250–258.
- 42 Cheng H, Vetrivel KS, Drisdell RC, Meckler X, Gong P, Leem JY, Li T, Carter M, Chen Y, Nguyen P, Iwatsubo T, Tomita T, Wong PC, Green WN, Kounnas MZ & Thinakaran G (2009) S-palmitoylation of gamma-secretase subunits nicastrin and APH-1. *J. Biol. Chem.* **284**, 1373–1384.
- 43 Podlisny MB, Citron M, Amarante P, Sherrington R, Xia W, Zhang J, Diehl T, Levesque G, Fraser P, Haass C, Koo EH, Seubert P, St George-Hyslop P, Teplow DB & Selkoe DJ (1997) Presenilin proteins undergo heterogeneous endoproteolysis between Thr291 and Ala299 and occur as stable N- and C-terminal fragments in normal and Alzheimer brain tissue. *Neurobiol. Dis.* **3**, 325–337.
- 44 Fukumori A, Fluhrer R, Steiner H & Haass C (2010) Three-amino acid spacing of presenilin endoproteolysis suggests a general stepwise cleavage of gamma-secretase-mediated intramembrane proteolysis. *J. Neurosci.* **30**, 7853–7862.
- 45 Knappenberger KS, Tian G, Ye X, Sobotka-Briner C, Ghanekar SV, Greenberg BD & Scott CW (2004) Mechanism of gamma-secretase cleavage activation: is gamma-secretase regulated through autoinhibition involving the presenilin-1 exon 9 loop? *Biochemistry* **43**, 6208–6218.
- 46 Kaether C, Capell A, Edbauer D, Winkler E, Novak B, Steiner H & Haass C (2004) The presenilin C-terminus is required for ER-retention, nicastrin-binding and gamma-secretase activity. *EMBO J.* **23**, 4738–4748.
- 47 Spasic D, Raemaekers T, Dillen K, Declerck I, Baert V, Serneels L, Füllekrug J & Annaert W (2007) Rer1p competes with APH-1 for binding to nicastrin and regulates gamma-secretase complex assembly in the early secretory pathway. *J. Cell Biol.* **176**, 629–640.
- 48 Schwappach B (2008) An overview of trafficking and assembly of neurotransmitter receptors and ion channels (Review). *Mol. Membr. Biol.* **25**, 270–278.
- 49 Area-Gomez E, Del Carmen Lara Castillo M, Tambini MD, Guardia-Laguarta C, de Groof AJC, Madra M, Ikenouchi J, Umeda M, Bird TD, Sturley SL & Schon EA (2012) Upregulated function of mitochondria-associated ER membranes in Alzheimer disease. *EMBO J.* **31**, 4106–4123.
- 50 Sato T, Diehl TS, Narayanan S, Funamoto S, Ihara Y, De Strooper B, Steiner H, Haass C & Wolfe MS (2007) Active

- gamma-secretase complexes contain only one of each component. *J. Biol. Chem.* **282**, 33985–33993.
- 51 Schroeter EH, Ilagan MXG, Brunkan AL, Hecimovic S, Li Y, Xu M, Lewis HD, Saxena MT, De Strooper B, Coonrod A, Tomita T, Iwatsubo T, Moore CL, Goate A, Wolfe MS, Shearman M & Kopan R (2003) A presenilin dimer at the core of the gamma-secretase enzyme: insights from parallel analysis of Notch 1 and APP proteolysis. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 13075–13080.
- 52 Laudon H, Hansson EM, Melén K, Bergman A, Farmery MR, Winblad B, Lendahl U, von Heijne G & Näslund J (2005) A nine-transmembrane domain topology for presenilin 1. *J. Biol. Chem.* **280**, 35352–35360.
- 53 Sobhanifar S, Schneider B, Löhr F, Gottstein D, Ikeya T, Mlynarczyk K, Pulawski W, Ghoshdastider U, Kolinski M, Filipek S, Güntert P, Bernhard F & Dötsch V (2010) Structural investigation of the C-terminal catalytic fragment of presenilin 1. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 9644–9649.
- 54 Crystal AS, Morais VA, Pierson TC, Pijak DS, Carlin D, Lee VM-Y & Doms RW (2003) Membrane topology of gamma-secretase component PEN-2. *J. Biol. Chem.* **278**, 20117–20123.
- 55 Francis R, McGrath G, Zhang J, Ruddy DA, Sym M, Apfeld J, Nicoll M, Maxwell M, Hai B, Ellis MC, Parks AL, Xu W, Li J, Gurney M, Myers RL, Himes CS, Hiesch R, Ruble C, Nye JS & Curtis D (2002) *aph-1* and *pen-2* are required for Notch pathway signaling, gamma-secretase cleavage of betaAPP, and presenilin protein accumulation. *Dev. Cell* **3**, 85–97.
- 56 Hasegawa H, Sanjo N, Chen F, Gu Y-J, Shier C, Petit A, Kawarai T, Katayama T, Schmidt SD, Mathews PM, Schmitt-Ulms G, Fraser PE & St George-Hyslop P (2004) Both the sequence and length of the C terminus of PEN-2 are critical for intermolecular interactions and function of presenilin complexes. *J. Biol. Chem.* **279**, 46455–46463.
- 57 Isoo N, Sato C, Miyashita H, Shinohara M, Takasugi N, Morohashi Y, Tsuji S, Tomita T & Iwatsubo T (2007) Aβ42 Overproduction Associated with Structural Changes in the Catalytic Pore of γ-Secretase COMMON EFFECTS OF PEN-2 N-TERMINAL ELONGATION AND FENOFIBRATE. *J. Biol. Chem.* **282**, 12388–12396.
- 58 Bammens L, Chávez-Gutiérrez L, Tolia A, Zwijsen A & De Strooper B (2011) Functional and topological analysis of Pen-2, the fourth subunit of the gamma-secretase complex. *J. Biol. Chem.* **286**, 12271–12282.
- 59 Yu G, Nishimura M, Arawaka S, Levitan D, Zhang L, Tandon A, Song YQ, Rogaeva E, Chen F, Kawarai T, Supala A, Levesque L, Yu H, Yang DS, Holmes E, Milman P, Liang Y, Zhang DM, Xu DH, Sato C, Rogaeve E, Smith M, Janus C, Zhang Y, Aebersold R, Farrer LS, Sorbi S, Bruni A, Fraser P & St George-Hyslop P (2000) Nicastrin modulates presenilin-mediated notch/glp-1 signal transduction and betaAPP processing. *Nature* **407**, 48–54.
- 60 Pardossi-Piquard R, Yang S-P, Kanemoto S, Gu Y, Chen F, Böhm C, Sevalle J, Li T, Wong PC, Checler F, Schmitt-Ulms G, St George-Hyslop P & Fraser PE (2009) APH1 polar transmembrane residues regulate the assembly and activity of presenilin complexes. *J. Biol. Chem.* **284**, 16298–16307.
- 61 Murphy MP, Das P, Nyborg AC, Rochette MJ, Dodson MW, Loosbrock NM, Souder TM, McLendon C, Merit SL, Piper SC, Jansen KR & Golde TE (2003) Overexpression of nicastrin increases Abeta production. *FASEB J.* **17**, 1138–1140.
- 62 Zhang X, Hoey RJ, Lin G, Koide A, Leung B, Ahn K, Dolios G, Paduch M, Ikeuchi T, Wang R, Li Y-M, Koide S & Sisodia SS (2012) Identification of a tetratricopeptide repeat-like domain in the nicastrin subunit of γ-secretase using synthetic antibodies. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 8534–8539.
- 63 Chen F, Hasegawa H, Schmitt-Ulms G, Kawarai T, Bohm C, Katayama T, Gu Y, Sanjo N, Glista M, Rogaeva E, Wakutani Y, Pardossi-Piquard R, Ruan X, Tandon A, Checler F, Marambaud P, Hansen K, Westaway D, St George-Hyslop P & Fraser P (2006) TMP21 is a presenilin complex component that modulates gamma-secretase but not epsilon-secretase activity. *Nature* **440**, 1208–1212.
- 64 Tolia A, Chávez-Gutiérrez L & De Strooper B (2006) Contribution of presenilin transmembrane domains 6 and 7 to a water-containing cavity in the gamma-secretase complex. *J. Biol. Chem.* **281**, 27633–27642.
- 65 Wang J, Brunkan AL, Hecimovic S, Walker E & Goate A (2004) Conserved “PAL” sequence in presenilins is essential for gamma-secretase activity, but not required for formation or stabilization of gamma-secretase complexes. *Neurobiol. Dis.* **15**, 654–666.
- 66 Wang J, Behr D, Nyborg AC, Shearman MS, Golde TE & Goate A (2006) C-terminal PAL motif of presenilin and presenilin homologues required for normal active site conformation. *J. Neurochem.* **96**, 218–227.
- 67 Takagi S, Tominaga A, Sato C, Tomita T & Iwatsubo T (2010) Participation of transmembrane domain 1 of presenilin 1 in the catalytic pore structure of the γ-secretase. *J. Neurosci.* **30**, 15943–15950.
- 68 Steiner H, Kostka M, Romig H, Basset G, Pesold B, Hardy J, Capell A, Meyn L, Grim ML, Baumeister R, Fichteler K & Haass C (2000) Glycine 384 is required for presenilin-1 function and is conserved in bacterial polytopic aspartyl proteases. *Nat. Cell Biol.* **2**, 848–851.
- 69 Watanabe N, Image Image II, Takagi S, Image Image II, Tominaga A, Image Image I, Tomita T, Image Image II, Iwatsubo T & Image Image I (2010) Functional analysis of the transmembrane domains of presenilin 1: participation of transmembrane domains 2 and 6 in the formation of initial substrate-binding site of gamma-secretase. *J. Biol. Chem.* **285**, 19738–19746.
- 70 Weidemann A, Eggert S, Reinhard FBM, Vogel M, Paliga K, Baier G, Masters CL, Beyreuther K & Evin G (2002) A novel epsilon-cleavage within the transmembrane domain of the Alzheimer amyloid precursor protein demonstrates homology with Notch processing. *Biochemistry* **41**, 2825–2835.

- 71 Zhao G, Mao G, Tan J, Dong Y, Cui M-Z, Kim S-H & Xu X (2004) Identification of a new presenilin-dependent zeta-cleavage site within the transmembrane domain of amyloid precursor protein. *J. Biol. Chem.* **279**, 50647–50650.
- 72 Zhao G, Cui M-Z, Mao G, Dong Y, Tan J, Sun L & Xu X (2005) gamma-Cleavage is dependent on zeta-cleavage during the proteolytic processing of amyloid precursor protein within its transmembrane domain. *J. Biol. Chem.* **280**, 37689–37697.
- 73 Qi-Takahara Y, Morishima-Kawashima M, Tanimura Y, Dolios G, Hirotsu N, Horikoshi Y, Kametani F, Maeda M, Saido TC, Wang R & Ihara Y (2005) Longer forms of amyloid beta protein: implications for the mechanism of intramembrane cleavage by gamma-secretase. *J. Neurosci.* **25**, 436–445.
- 74 Takami M, Nagashima Y, Sano Y, Ishihara S, Morishima-Kawashima M, Funamoto S & Ihara Y (2009) gamma-Secretase: successive tripeptide and tetrapeptide release from the transmembrane domain of beta-carboxyl terminal fragment. *J. Neurosci.* **29**, 13042–13052.
- 75 Munter L-M, Voigt P, Harmeier A, Kaden D, Gottschalk KE, Weise C, Pipkorn R, Schaefer M, Langosch D & Multhaup G (2007) GxxxG motifs within the amyloid precursor protein transmembrane sequence are critical for the etiology of Abeta42. *EMBO J.* **26**, 1702–1712.
- 76 Okochi M, Steiner H, Fukumori A, Tani H, Tomita T, Tanaka T, Iwatsubo T, Kudo T, Takeda M & Haass C (2002) Presenilins mediate a dual intramembraneous gamma-secretase cleavage of Notch-1. *EMBO J.* **21**, 5408–5416.
- 77 Jonsson T, Atwal JK, Steinberg S, Snaedal J, Jonsson PV, Bjornsson S, Stefansson H, Sulem P, Gudbjartsson D, Maloney J, Hoyte K, Gustafson A, Liu Y, Lu Y, Bhargava T, Graham RR, Huttenlocher J, Bjornsdottir G, Andreassen OA, Jönsson EG, Palotie A, Behrens TW, Magnusson OT, Kong A, Thorsteinsdottir U, Watts RJ & Stefansson K (2012) A mutation in APP protects against Alzheimer's disease and age-related cognitive decline. *Nature* **488**, 96–99.
- 78 Chávez-Gutiérrez L, Bammens L, Benilova I, Vandersteen A, Benurwar M, Borgers M, Lismont S, Zhou L, Van Cleynebreugel S, Esselmann H, Wiltfang J, Serneels L, Karran E, Gijzen H, Schymkowitz J, Rousseau F, Broersen K & De Strooper B (2012) The mechanism of  $\gamma$ -Secretase dysfunction in familial Alzheimer disease. *EMBO J.* **31**, 2261–2274.
- 79 Esler WP, Kimberly WT, Ostaszewski BL, Ye W, Diehl TS, Selkoe DJ & Wolfe MS (2002) Activity-dependent isolation of the presenilin- gamma -secretase complex reveals nicastrin and a gamma substrate. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 2720–2725.
- 80 Dries DR, Shah S, Han Y-H, Yu C, Yu S, Shearman MS & Yu G (2009) Glu-333 of nicastrin directly participates in gamma-secretase activity. *J. Biol. Chem.* **284**, 29714–29724.
- 81 Zhao G, Liu Z, Ilagan MXG & Kopan R (2010) Gamma-secretase composed of PS1/Pen2/Aph1a can cleave notch and amyloid precursor protein in the absence of nicastrin. *J. Neurosci.* **30**, 1648–1656.
- 82 Chen AC, Guo LY, Ostaszewski BL, Selkoe DJ & LaVoie MJ (2010) Aph-1 associates directly with full-length and C-terminal fragments of gamma-secretase substrates. *J. Biol. Chem.* **285**, 11378–11391.
- 83 Hemming ML, Elias JE, Gygi SP & Selkoe DJ (2008) Proteomic profiling of gamma-secretase substrates and mapping of substrate requirements. *PLoS Biol.* **6**, e257.
- 84 Roher AE, Lowenson JD, Clarke S, Woods AS, Cotter RJ, Gowing E & Ball MJ (1993) beta-Amyloid-(1-42) is a major component of cerebrovascular amyloid deposits: implications for the pathology of Alzheimer disease. *Proc. Natl. Acad. Sci. U.S.A.* **90**, 10836–10840.
- 85 Scheuner D, Eckman C, Jensen M, Song X, Citron M, Suzuki N, Bird TD, Hardy J, Hutton M, Kukull W, Larson E, Levy-Lahad E, Viitanen M, Peskind E, Poorkaj P, Schellenberg G, Tanzi R, Wasco W, Lannfelt L, Selkoe D & Younkin S (1996) Secreted amyloid beta-protein similar to that in the senile plaques of Alzheimer's disease is increased in vivo by the presenilin 1 and 2 and APP mutations linked to familial Alzheimer's disease. *Nat. Med.* **2**, 864–870.
- 86 Heilig EA, Xia W, Shen J & Kelleher RJ 3rd (2010) A presenilin-1 mutation identified in familial Alzheimer disease with cotton wool plaques causes a nearly complete loss of gamma-secretase activity. *J. Biol. Chem.* **285**, 22350–22359.
- 87 Thathiah A & De Strooper B (2011) The role of G protein-coupled receptors in the pathology of Alzheimer's disease. *Nat. Rev. Neurosci.* **12**, 73–87.
- 88 Hu J, Xue Y, Lee S & Ha Y (2011) The crystal structure of GXGD membrane protease FlaK. *Nature* **475**, 528–531.
- 89 Feng L, Yan H, Wu Z, Yan N, Wang Z, Jeffrey PD & Shi Y (2007) Structure of a site-2 protease family intramembrane metalloprotease. *Science* **318**, 1608–1612.
- 90 Coles M, Bicknell W, Watson AA, Fairlie DP & Craik DJ (1998) Solution structure of amyloid beta-peptide(1-40) in a water-micelle environment. Is the membrane-spanning domain where we think it is? *Biochemistry* **37**, 11064–11077.
- 91 Soto C, Brañes MC, Alvarez J & Inestrosa NC (1994) Structural determinants of the Alzheimer's amyloid beta-peptide. *J. Neurochem.* **63**, 1191–1198.
- 92 Sievers SA, Karanicolas J, Chang HW, Zhao A, Jiang L, Zirafi O, Stevens JT, Münch J, Baker D & Eisenberg D (2011) Structure-based design of non-natural amino-acid inhibitors of amyloid fibril formation. *Nature* **475**, 96–100.

- 93 Selkoe DJ (2001) Alzheimer's disease: genes, proteins, and therapy. *Physiol. Rev.* **81**, 741–766.
- 94 von Bergen M, Friedhoff P, Biernat J, Heberle J, Mandelkow EM & Mandelkow E (2000) Assembly of tau protein into Alzheimer paired helical filaments depends on a local sequence motif ((306)VQIVYK(311)) forming beta structure. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 5129–5134.
- 95 Goux WJ, Kopplin L, Nguyen AD, Leak K, Rutkofsky M, Shanmuganandam VD, Sharma D, Inouye H & Kirschner DA (2004) The formation of straight and twisted filaments from short tau peptides. *J. Biol. Chem.* **279**, 26868–26875.
- 96 Münch J, Rucker E, Ständker L, Adermann K, Goffinet C, Schindler M, Wildum S, Chinnadurai R, Rajan D, Specht A, Giménez-Gallego G, Sánchez PC, Fowler DM, Koulov A, Kelly JW, Mothes W, Grivel J-C, Margolis L, Keppler OT, Forssmann W-G & Kirchhoff F (2007) Semen-derived amyloid fibrils drastically enhance HIV infection. *Cell* **131**, 1059–1071.
- 97 Goldschmidt L, Teng PK, Riek R & Eisenberg D (2010) Identifying the amyloids, proteins capable of forming amyloid-like fibrils. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 3487–3492.
- 98 Aguzzi A & O'Connor T (2010) Protein aggregation diseases: pathogenicity and therapeutic perspectives. *Nat Rev Drug Discov* **9**, 237–248.
- 99 Schnabel J (2010) Protein folding: The dark side of proteins. *Nature* **464**, 828–829.
- 100 Thompson MJ, Sievers SA, Karanicolas J, Ivanova MI, Baker D & Eisenberg D (2006) The 3D profile method for identifying fibril-forming segments of proteins. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 4074–4078.
- 101 Sunde M & Blake CC (1998) From the globular to the fibrous state: protein structure and structural conversion in amyloid formation. *Q. Rev. Biophys.* **31**, 1–39.
- 102 Chiti F & Dobson CM (2009) Amyloid formation by globular proteins under native conditions. *Nat. Chem. Biol.* **5**, 15–22.
- 103 Khoshnoodi J, Cartiailler J-P, Alvares K, Veis A & Hudson BG (2006) Molecular recognition in the assembly of collagens: terminal noncollagenous domains are key recognition modules in the formation of triple helical protomers. *J. Biol. Chem.* **281**, 38117–38121.
- 104 Cejas MA, Kinney WA, Chen C, Vinter JG, Almond HR Jr, Balss KM, Maryanoff CA, Schmidt U, Breslav M, Mahan A, Lacy E & Maryanoff BE (2008) Thrombogenic collagen-mimetic peptides: Self-assembly of triple helix-based fibrils driven by hydrophobic interactions. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 8513–8518.
- 105 Carlier M-F & Pantaloni D (2007) Control of actin assembly dynamics in cell motility. *J. Biol. Chem.* **282**, 23005–23009.
- 106 Binder WH & Smrzka OW (2006) Self-assembly of fibers and fibrils. *Angew. Chem. Int. Ed. Engl.* **45**, 7324–7328.
- 107 Luheshi LM & Dobson CM (2009) Bridging the gap: from protein misfolding to protein misfolding diseases. *FEBS Lett.* **583**, 2581–2586.
- 108 Caughey B & Lansbury PT (2003) Protofibrils, pores, fibrils, and neurodegeneration: separating the responsible protein aggregates from the innocent bystanders. *Annu. Rev. Neurosci.* **26**, 267–298.
- 109 Conway KA, Lee SJ, Rochet JC, Ding TT, Williamson RE & Lansbury PT Jr (2000) Acceleration of oligomerization, not fibrillization, is a shared property of both alpha-synuclein mutations linked to early-onset Parkinson's disease: implications for pathogenesis and therapy. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 571–576.
- 110 Apetri MM, Maiti NC, Zagorski MG, Carey PR & Anderson VE (2006) Secondary structure of alpha-synuclein oligomers: characterization by raman and atomic force microscopy. *J. Mol. Biol.* **355**, 63–71.
- 111 Bartolini M & Andrisano V (2010) Strategies for the inhibition of protein aggregation in human diseases. *Chembiochem* **11**, 1018–1035.
- 112 Petkova AT, Ishii Y, Balbach JJ, Antzutkin ON, Leapman RD, Delaglio F & Tycko R (2002) A structural model for Alzheimer's beta -amyloid fibrils based on experimental constraints from solid state NMR. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 16742–16747.
- 113 Smith JF, Knowles TPJ, Dobson CM, Macphee CE & Welland ME (2006) Characterization of the nanoscale properties of individual amyloid fibrils. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 15806–15811.
- 114 Lührs T, Ritter C, Adrian M, Riek-Loher D, Bohrmann B, Döbeli H, Schubert D & Riek R (2005) 3D structure of Alzheimer's amyloid-beta(1-42) fibrils. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 17342–17347.
- 115 Sawaya MR, Sambashivan S, Nelson R, Ivanova MI, Sievers SA, Apostol MI, Thompson MJ, Balbirnie M, Wiltzius JJW, McFarlane HT, Madsen AØ, Riek C & Eisenberg D (2007) Atomic structures of amyloid cross-beta spines reveal varied steric zippers. *Nature* **447**, 453–457.
- 116 Wiltzius JJW, Landau M, Nelson R, Sawaya MR, Apostol MI, Goldschmidt L, Soriaga AB, Cascio D, Rajashankar K & Eisenberg D (2009) Molecular mechanisms for protein-encoded inheritance. *Nat. Struct. Mol. Biol.* **16**, 973–978.
- 117 Bhak G, Choe Y-J & Paik SR (2009) Mechanism of amyloidogenesis: nucleation-dependent fibrillation versus double-concerted fibrillation. *BMB Rep* **42**, 541–551.
- 118 Žerovnik E, Stoka V, Mirtič A, Gunčar G, Grdadolnik J, Staniforth RA, Turk D & Turk V (2011) Mechanisms of amyloid fibril formation—focus on domain-swapping. *FEBS J.* **278**, 2263–2282.
- 119 Goedert M (2001) Alpha-synuclein and neurodegenerative diseases. *Nat. Rev. Neurosci.* **2**, 492–501.
- 120 Luk KC, Song C, O'Brien P, Stieber A, Branch JR, Brunden KR, Trojanowski JQ & Lee VM-Y (2009) Exogenous

- alpha-synuclein fibrils seed the formation of Lewy body-like intracellular inclusions in cultured cells. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 20051–20056.
- 121 Jaikaran ET & Clark A (2001) Islet amyloid and type 2 diabetes: from molecular misfolding to islet pathophysiology. *Biochim. Biophys. Acta* **1537**, 179–203.
- 122 Walsh DM, Hartley DM, Kusumoto Y, Fezoui Y, Condron MM, Lomakin A, Benedek GB, Selkoe DJ & Teplow DB (1999) Amyloid beta-protein fibrillogenesis. Structure and biological activity of protofibrillar intermediates. *J. Biol. Chem.* **274**, 25945–25952.
- 123 Koch KM (1992) Dialysis-related amyloidosis. *Kidney Int.* **41**, 1416–1429.
- 124 Bates G (2003) Huntingtin aggregation and toxicity in Huntington's disease. *Lancet* **361**, 1642–1644.
- 125 Bellotti V, Mangione P & Merlini G (2000) Review: immunoglobulin light chain amyloidosis--the archetype of structural and pathogenic variability. *J. Struct. Biol.* **130**, 280–289.
- 126 Pepys MB, Hawkins PN, Booth DR, Vigushin DM, Tennent GA, Soutar AK, Totty N, Nguyen O, Blake CC & Terry CJ (1993) Human lysozyme gene mutations cause hereditary systemic amyloidosis. *Nature* **362**, 553–557.
- 127 Prusiner SB (1998) Prions. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 13363–13383.
- 128 Kelly JW, Colon W, Lai Z, Lashuel HA, McCulloch J, McCutchen SL, Miroy GJ & Peterson SA (1997) Transthyretin quaternary and tertiary structural changes facilitate misassembly into amyloid. *Adv. Protein Chem.* **50**, 161–181.
- 129 Meyer-Luehmann M, Coomaraswamy J, Bolmont T, Kaeser S, Schaefer C, Kilger E, Neuenschwander A, Abramowski D, Frey P, Jaton AL, Vigouret J-M, Paganetti P, Walsh DM, Mathews PM, Ghiso J, Staufenbiel M, Walker LC & Jucker M (2006) Exogenous induction of cerebral beta-amyloidogenesis is governed by agent and host. *Science* **313**, 1781–1784.
- 130 Ohhashi Y, Ito K, Toyama BH, Weissman JS & Tanaka M (2010) Differences in prion strain conformations result from non-native interactions in a nucleus. *Nat. Chem. Biol.* **6**, 225–230.
- 131 Buell AK, Jamie R, Blundell, Dobson CM, Welland ME, Terentjev EM & Knowles TPJ (2010) Frequency Factors in a Landscape Model of Filamentous Protein Aggregation. *Phys. Rev. Lett.* **104**, 228101.
- 132 White DA, Buell AK, Knowles TPJ, Welland ME & Dobson CM (2010) Protein aggregation in crowded environments. *J. Am. Chem. Soc.* **132**, 5170–5175.
- 133 Campioni S, Mannini B, Zampagni M, Pensalfini A, Parrini C, Evangelisti E, Relini A, Stefani M, Dobson CM, Cecchi C & Chiti F (2010) A causative link between the structure of aberrant protein oligomers and their toxicity. *Nat. Chem. Biol.* **6**, 140–147.
- 134 Morris AM, Watzky MA & Finke RG (2009) Protein aggregation kinetics, mechanism, and curve-fitting: a review of the literature. *Biochim. Biophys. Acta* **1794**, 375–397.
- 135 Wood SJ, Wypych J, Steavenson S, Louis JC, Citron M & Biere AL (1999) alpha-synuclein fibrillogenesis is nucleation-dependent. Implications for the pathogenesis of Parkinson's disease. *J. Biol. Chem.* **274**, 19509–19512.
- 136 Newcomer ME (2002) Protein folding and three-dimensional domain swapping: a strained relationship? *Curr. Opin. Struct. Biol.* **12**, 48–53.
- 137 Guo Z & Eisenberg D (2006) Runaway domain swapping in amyloid-like fibrils of T7 endonuclease I. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 8042–8047.
- 138 Wahlbom M, Wang X, Lindström V, Carlemalm E, Jaskolski M & Grubb A (2007) Fibrillogenic oligomers of human cystatin C are formed by propagated domain swapping. *J. Biol. Chem.* **282**, 18318–18326.
- 139 Pallitto MM & Murphy RM (2001) A mathematical model of the kinetics of beta-amyloid fibril growth from the denatured state. *Biophys. J.* **81**, 1805–1822.
- 140 Maji SK, Ogorzalek Loo RR, Inayathullah M, Spring SM, Vollers SS, Condron MM, Bitan G, Loo JA & Teplow DB (2009) Amino acid position-specific contributions to amyloid beta-protein oligomerization. *J. Biol. Chem.* **284**, 23580–23591.
- 141 Ono K, Condron MM & Teplow DB (2009) Structure-neurotoxicity relationships of amyloid beta-protein oligomers. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 14745–14750.
- 142 Murray MM, Bernstein SL, Nyugen V, Condron MM, Teplow DB & Bowers MT (2009) Amyloid beta protein: Abeta40 inhibits Abeta42 oligomerization. *J. Am. Chem. Soc.* **131**, 6316–6317.
- 143 Bernstein SL, Dupuis NF, Lazo ND, Wyttenbach T, Condron MM, Bitan G, Teplow DB, Shea J-E, Ruotolo BT, Robinson CV & Bowers MT (2009) Amyloid- $\beta$  protein oligomerization and the importance of tetramers and dodecamers in the aetiology of Alzheimer's disease. *Nat Chem* **1**, 326–331.
- 144 Calamai M, Chiti F & Dobson CM (2005) Amyloid fibril formation can proceed from different conformations of a partially unfolded protein. *Biophys. J.* **89**, 4201–4210.
- 145 Ostapchenko VG, Sawaya MR, Makarava N, Savtchenko R, Nilsson KPR, Eisenberg D & Baskakov IV (2010) Two amyloid States of the prion protein display significantly different folding patterns. *J. Mol. Biol.* **400**, 908–921.
- 146 Edgeworth JA, Gros N, Alden J, Joiner S, Wadsworth JDF, Linehan J, Brandner S, Jackson GS, Weissmann C & Collinge J (2010) Spontaneous generation of mammalian prions. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 14402–14406.
- 147 Urbanc B, Betnel M, Cruz L, Bitan G & Teplow DB (2010) Elucidation of amyloid beta-protein oligomerization

- mechanisms: discrete molecular dynamics study. *J. Am. Chem. Soc.* **132**, 4266–4280.
- 148 Lemkul JA & Bevan DR (2010) Destabilizing Alzheimer's A $\beta$ (42) protofibrils with morin: mechanistic insights from molecular dynamics simulations. *Biochemistry* **49**, 3935–3946.
- 149 Lemkul JA & Bevan DR (2009) Perturbation of membranes by the amyloid beta-peptide--a molecular dynamics study. *FEBS J.* **276**, 3060–3075.
- 150 Masman MF, Eisel ULM, Csizmadia IG, Penke B, Enriz RD, Marrink SJ & Luiten PGM (2009) In silico study of full-length amyloid beta 1-42 tri- and penta-oligomers in solution. *J Phys Chem B* **113**, 11710–11719.
- 151 Fowler DM, Koulov AV, Balch WE & Kelly JW (2007) Functional amyloid--from bacteria to humans. *Trends Biochem. Sci.* **32**, 217–224.
- 152 Ionomidou VA, Vriend G & Hamodrakas SJ (2000) Amyloids protect the silkworm oocyte and embryo. *FEBS Lett.* **479**, 141–145.
- 153 Ionomidou VA & Hamodrakas SJ (2008) Natural protective amyloids. *Curr. Protein Pept. Sci.* **9**, 291–309.
- 154 Fowler DM, Koulov AV, Alory-Jost C, Marks MS, Balch WE & Kelly JW (2006) Functional amyloid formation within mammalian tissue. *PLoS Biol.* **4**, e6.
- 155 Maji SK, Perrin MH, Sawaya MR, Jessberger S, Vadodaria K, Rissman RA, Singru PS, Nilsson KPR, Simon R, Schubert D, Eisenberg D, Rivier J, Sawchenko P, Vale W & Riek R (2009) Functional amyloids as natural storage of peptide hormones in pituitary secretory granules. *Science* **325**, 328–332.
- 156 Yonemoto IT, Kroon GJA, Dyson HJ, Balch WE & Kelly JW (2008) Amylin proprotein processing generates progressively more amyloidogenic peptides that initially sample the helical state. *Biochemistry* **47**, 9900–9910.
- 157 Cherny I & Gazit E (2008) Amyloids: not only pathological agents but also ordered nanomaterials. *Angew. Chem. Int. Ed. Engl.* **47**, 4062–4069.
- 158 Ajayaghosh A & Praveen VK (2007) Pi-organogels of self-assembled p-phenylenevinylenes: soft materials with distinct size, shape, and functions. *Acc. Chem. Res.* **40**, 644–656.
- 159 Prasanthkumar S, Saeki A, Seki S & Ajayaghosh A (2010) Solution phase epitaxial self-assembly and high charge-carrier mobility nanofibers of semiconducting molecular gelators. *J. Am. Chem. Soc.* **132**, 8866–8867.
- 160 Reches M & Gazit E (2003) Casting metal nanowires within discrete self-assembled peptide nanotubes. *Science* **300**, 625–627.
- 161 Reches M & Gazit E (2006) Designed aromatic homo-dipeptides: formation of ordered nanostructures and potential nanotechnological applications. *Phys Biol* **3**, S10–19.
- 162 Adler-Abramovich L & Gazit E (2008) Controlled patterning of peptide nanotubes and nanospheres using inkjet printing technology. *J. Pept. Sci.* **14**, 217–223.
- 163 Yemini M, Reches M, Gazit E & Rishpon J (2005) Peptide nanotube-modified electrodes for enzyme-biosensor applications. *Anal. Chem.* **77**, 5155–5159.
- 164 Yemini M, Reches M, Rishpon J & Gazit E (2005) Novel electrochemical biosensing platform using self-assembled peptide nanotubes. *Nano Lett.* **5**, 183–186.
- 165 Beker P, Koren I, Amdursky N, Gazit E & Rosenman G (2010) Bioinspired peptide nanotubes as supercapacitor electrodes. *Journal of Materials Science* **45**, 6374–6378.
- 166 Adler-Abramovich L, Reches M, Sedman VL, Allen S, Tendler SJB & Gazit E (2006) Thermal and chemical stability of diphenylalanine peptide nanotubes: implications for nanotechnological applications. *Langmuir* **22**, 1313–1320.
- 167 Knowles TP, Fitzpatrick AW, Meehan S, Mott HR, Vendruscolo M, Dobson CM & Welland ME (2007) Role of intermolecular forces in defining material properties of protein nanofibrils. *Science* **318**, 1900–1903.
- 168 Adler-Abramovich L, Aronov D, Beker P, Yevnin M, Stempler S, Buzhansky L, Rosenman G & Gazit E (2009) Self-assembled arrays of peptide nanotubes by vapour deposition. *Nat Nanotechnol* **4**, 849–854.
- 169 Gras SL, Tickler AK, Squires AM, Devlin GL, Horton MA, Dobson CM & MacPhee CE (2008) Functionalised amyloid fibrils for roles in cell adhesion. *Biomaterials* **29**, 1553–1562.
- 170 Choi YS, Kim J, Bhak G, Lee D & Paik SR (2011) Photoelectric protein nanofibrils of  $\alpha$ -synuclein with embedded iron and phthalocyanine tetrasulfonate. *Angew. Chem. Int. Ed. Engl.* **50**, 6070–6074.
- 171 Lee D, Choe Y-J, Choi YS, Bhak G, Lee J & Paik SR (2011) Photoconductivity of pea-pod-type chains of gold nanoparticles encapsulated within dielectric amyloid protein nanofibrils of  $\alpha$ -synuclein. *Angew. Chem. Int. Ed. Engl.* **50**, 1332–1337.
- 172 Schwarz D, Junge F, Durst F, Frölich N, Schneider B, Reckel S, Sobhanifar S, Dötsch V & Bernhard F (2007) Preparative scale expression of membrane proteins in Escherichia coli-based continuous exchange cell-free systems. *Nat Protoc* **2**, 2945–2957.
- 173 Junge F, Haberstock S, Roos C, Stefer S, Proverbio D, Dötsch V & Bernhard F (2011) Advances in cell-free protein synthesis for the functional and structural analysis of membrane proteins. *N Biotechnol* **28**, 262–271.
- 174 Roos C, Kai L, Proverbio D, Ghoshdastider U, Filipek S, Dötsch V & Bernhard F (2012) Co-translational association of cell-free expressed membrane proteins with supplied lipid bilayers. *Molecular membrane biology.*
- 175 Roos C, Kai L, Haberstock S, Proverbio D, Ghoshdastider U, Ma Y, Filipek S, Wang X, Dötsch V & Bernhard F (2012) High Level Cell-Free Production of Membrane Proteins into Nanodiscs. *Methods Mol. Biol.*
- 176 Katzen F, Peterson TC & Kudlicki W (2009) Membrane protein expression: no cells required. *Trends Biotechnol.* **27**,

- 455–460.
- 177 Stögbauer T, Windhager L, Zimmer R & Rädler JO (2012) Experiment and mathematical modeling of gene expression dynamics in a cell-free system. *Integr Biol (Camb)* **4**, 494–501.
- 178 Kuruma Y, Stano P, Ueda T & Luisi PL (2009) A synthetic biology approach to the construction of membrane proteins in semi-synthetic minimal cells. *Biochim. Biophys. Acta* **1788**, 567–574.
- 179 Bayburt TH & Sligar SG (2010) Membrane protein assembly into Nanodiscs. *FEBS Lett.* **584**, 1721–1727.
- 180 Kim T-W, Keum J-W, Oh I-S, Choi C-Y, Kim H-C & Kim D-M (2007) An economical and highly productive cell-free protein synthesis system utilizing fructose-1,6-bisphosphate as an energy source. *J. Biotechnol.* **130**, 389–393.
- 181 Zocher M, Roos C, Wegmann S, Bosshart PD, Dötsch V, Bernhard F & Müller DJ (2012) Single-molecule force spectroscopy from nanodiscs: an assay to quantify folding, stability, and interactions of native membrane proteins. *ACS Nano* **6**, 961–971.
- 182 Reckel S, Gottstein D, Stehle J, Löhr F, Verhoeven M-K, Takeda M, Silvers R, Kainosho M, Glaubitz C, Wachtveitl J, Bernhard F, Schwalbe H, Güntert P & Dötsch V (2011) Solution NMR structure of proteorhodopsin. *Angew. Chem. Int. Ed. Engl.* **50**, 11942–11946.
- 183 von Buelow J, Mueller-Lucks A, Kai L, Bernhard F & Beitz E (2012) Functional characterization of a novel aquaporin from *Dictyostelium discoideum* amoebae implies a unique gating mechanism. *The Journal of Biological Chemistry*.
- 184 Kai L, Kaldenhoff R, Lian J, Zhu X, Dötsch V, Bernhard F, Cen P & Xu Z (2010) Preparative scale production of functional mouse aquaporin 4 using different cell-free expression modes. *PLoS ONE* **5**, e12972.
- 185 Junge F, Luh LM, Proverbio D, Schäfer B, Abele R, Beyermann M, Dötsch V & Bernhard F (2010) Modulation of G-protein coupled receptor sample quality by modified cell-free expression protocols: a case study of the human endothelin A receptor. *J. Struct. Biol.* **172**, 94–106.
- 186 Rath P, Demange P, Saurel O, Tropis M, Daffé M, Dötsch V, Ghazi A, Bernhard F & Milon A (2011) Functional expression of the PorAH channel from *Corynebacterium glutamicum* in cell-free expression systems: implications for the role of the naturally occurring mycolic acid modification. *J. Biol. Chem.* **286**, 32525–32532.
- 187 Keller T, Schwarz D, Bernhard F, Dötsch V, Hunte C, Gorboulev V & Koepsell H (2008) Cell free expression and functional reconstitution of eukaryotic drug transporters. *Biochemistry* **47**, 4552–4564.
- 188 Schwarz D, Daley D, Beckhaus T, Dötsch V & Bernhard F (2010) Cell-free expression profiling of *E. coli* inner membrane proteins. *Proteomics* **10**, 1762–1779.
- 189 Ma Y, Münch D, Schneider T, Sahl H-G, Bouhss A, Ghoshdastider U, Wang J, Dötsch V, Wang X & Bernhard F (2011) Preparative scale cell-free production and quality optimization of MraY homologues in different expression modes. *J. Biol. Chem.* **286**, 38844–38853.
- 190 Ma Y, Ghoshdastider U, Wang J, Ye W, Dötsch V, Filipek S, Bernhard F & Wang X (2012) Cell-free expression of human glucosamine 6-phosphate N-acetyltransferase (HsGNA1) for inhibitor screening. *Protein Expr. Purif.* **86**, 120–126.
- 191 Stefer S, Reitz S, Wang F, Wild K, Pang Y-Y, Schwarz D, Bomke J, Hein C, Löhr F, Bernhard F, Denic V, Dötsch V & Sinning I (2011) Structural basis for tail-anchored membrane protein biogenesis by the Get3-receptor complex. *Science* **333**, 758–762.
- 192 Matthies D, Haberstock S, Joos F, Dötsch V, Vonck J, Bernhard F & Meier T (2011) Cell-free expression and assembly of ATP synthase. *J. Mol. Biol.* **413**, 593–603.
- 193 Haberstock S, Roos C, Hoevens Y, Dötsch V, Schnapp G, Pautsch A & Bernhard F (2012) A systematic approach to increase the efficiency of membrane protein production in cell-free expression systems. *Protein Expr. Purif.* **82**, 308–316.
- 194 Klammt C, Löhr F, Schäfer B, Haase W, Dötsch V, Rüterjans H, Glaubitz C & Bernhard F (2004) High level cell-free expression and specific labeling of integral membrane proteins. *Eur. J. Biochem* **271**, 568–580.
- 195 Reckel S, Sobhanifar S, Schneider B, Junge F, Schwarz D, Durst F, Löhr F, Güntert P, Bernhard F & Dötsch V (2008) Transmembrane segment enhanced labeling as a tool for the backbone assignment of alpha-helical membrane proteins. *Proc. Natl. Acad. Sci. U.S.A* **105**, 8262–8267.
- 196 Klammt C, Schwarz D, Fendler K, Haase W, Dötsch V & Bernhard F (2005) Evaluation of detergents for the soluble expression of alpha-helical and beta-barrel-type integral membrane proteins by a preparative scale individual cell-free expression system. *FEBS J* **272**, 6024–6038.
- 197 Kai L, Roos C, Haberstock S, Proverbio D, Ma Y, Junge F, Karbyshev M, Dötsch V & Bernhard F (2012) Systems for the cell-free synthesis of proteins. *Methods Mol. Biol.* **800**, 201–225.
- 198 Denisov IG, Grinkova YV, Lazarides AA & Sligar SG (2004) Directed self-assembly of monodisperse phospholipid bilayer Nanodiscs with controlled size. *J. Am. Chem. Soc.* **126**, 3477–3487.
- 199 Grinkova YV, Denisov IG & Sligar SG (2010) Engineering extended membrane scaffold proteins for self-assembly of soluble nanoscale lipid bilayers. *Protein Eng. Des. Sel.* **23**, 843–848.
- 200 Nesterenko MV, Tilley M & Upton SJ (1994) A simple modification of Blum's silver stain method allows for 30 minute detection of proteins in polyacrylamide gels. *J. Biochem. Biophys. Methods* **28**, 239–242.
- 201 Ryckaert JP, Ciccotti G & Berendsen HJC (1977) Numerical integration of the cartesian equations of motion of a



- system with constraints: molecular dynamics of *n*-alkanes. *Journal of Computational Physics* **23**, 327–341.
- 202 Hess B, Bekker H, Berendsen HJC & Fraaije JGEM (1997) LINCS: A linear constraint solver for molecular simulations. *Journal of Computational Chemistry* **18**, 1463–1472.
- 203 Hess B (2008) P-LINCS: A Parallel Linear Constraint Solver for Molecular Simulation. *J. Chem. Theory Comput.* **4**, 116–122.
- 204 Hockney R, Goel S & Eastwood J (1974) Quiet high-resolution computer models of a plasma. *Journal of Computational Physics* **14**, 148–158.
- 205 Marrink SJ, Risselada HJ, Yefimov S, Tieleman DP & de Vries AH (2007) The MARTINI force field: coarse grained model for biomolecular simulations. *J Phys Chem B* **111**, 7812–7824.
- 206 Monticelli L, Kandasamy SK, Periole X, Larson RG, Tieleman DP & Marrink S-J (2008) The MARTINI Coarse-Grained Force Field: Extension to Proteins. *J. Chem. Theory Comput.* **4**, 819–834.
- 207 Periole X & Marrink S-J (2013) The martini coarse-grained force field. *Methods Mol. Biol.* **924**, 533–565.
- 208 Marrink SJ & Mark AE (2003) The mechanism of vesicle fusion as revealed by molecular dynamics simulations. *J. Am. Chem. Soc.* **125**, 11144–11145.
- 209 Kasson PM, Kelley NW, Singhal N, Vrljic M, Brunger AT & Pande VS (2006) Ensemble molecular dynamics yields submillisecond kinetics and intermediates of membrane fusion. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 11916–11921.
- 210 Risselada HJ & Marrink SJ (2008) The molecular face of lipid rafts in model membranes. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 17367–17372.
- 211 Bond PJ & Sansom MSP (2006) Insertion and assembly of membrane proteins via simulation. *J. Am. Chem. Soc.* **128**, 2697–2704.
- 212 Bogaart G van den, Meyenberg K, Risselada HJ, Amin H, Willig KI, Hubrich BE, Dier M, Hell SW, Grubmüller H, Diederichsen U & Jahn R (2011) Membrane protein sequestering by ionic protein-lipid interactions. *Nature* **479**, 552–555.
- 213 Louhivuori M, Risselada HJ, van der Giessen E & Marrink SJ (2010) Release of content through mechano-sensitive gates in pressurized liposomes. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 19856–19860.
- 214 Lazaridis T & Karplus M (1999) Effective energy function for proteins in solution. *Proteins: Structure, Function, and Bioinformatics* **35**, 133–152.
- 215 Lazaridis T (2003) Effective energy function for proteins in lipid membranes. *Proteins* **52**, 176–192.
- 216 Kabsch W & Sander C (1983) Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**, 2577–2637.
- 217 Hess B (2000) Similarities between principal components of protein dynamics and random diffusion. *Physical Review E* **62**, 8438.
- 218 Lodish H, Berk A, Kaiser CA, Krieger M, Bretscher A, Ploegh H, Amon A & Scott MP (2012) *Molecular Cell Biology*, Seventh ed. W. H. Freeman.
- 219 Lomize AL, Pogozheva ID & Mosberg HI (2011) Anisotropic solvent model of the lipid bilayer. 2. Energetics of insertion of small molecules, peptides, and proteins in membranes. *J Chem Inf Model* **51**, 930–946.
- 220 Yarov-Yarovoy V, Schonbrun J & Baker D (2006) Multipass membrane protein structure prediction using Rosetta. *Proteins* **62**, 1010–1025.
- 221 Barth P, Schonbrun J & Baker D (2007) Toward high-resolution prediction and design of transmembrane helical protein structures. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 15682–15687.
- 222 Barth P, Wallner B & Baker D (2009) Prediction of membrane protein structures with complex topologies using limited constraints. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 1409–1414.
- 223 Viklund H & Elofsson A (2008) OCTOPUS: improving topology prediction by two-track ANN-based preference scores and an extended topological grammar. *Bioinformatics* **24**, 1662–1668.
- 224 Söding J (2005) Protein homology detection by HMM-HMM comparison. *Bioinformatics* **21**, 951–960.
- 225 Söding J, Biegert A & Lupas AN (2005) The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res.* **33**, W244–248.
- 226 Zhang Y (2008) I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics* **9**, 40.
- 227 Roy A, Kucukural A & Zhang Y (2010) I-TASSER: a unified platform for automated protein structure and function prediction. *Nat Protoc* **5**, 725–738.
- 228 Wallner B & Elofsson A (2003) Can correct protein models be identified? *Protein Sci.* **12**, 1073–1086.
- 229 Larsson P, Skwark MJ, Wallner B & Elofsson A (2009) Assessment of global and local model quality in CASP8 using Pcons and ProQ. *Proteins* **77 Suppl 9**, 167–172.
- 230 Rohl CA, Strauss CEM, Misura KMS & Baker D (2004) Protein structure prediction using Rosetta. *Meth. Enzymol.* **383**, 66–93.
- 231 Brooks BR, Brucoleri RE, Olafson BD, States DJ, Swaminathan S & Karplus M (1983) CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *Journal of Computational Chemistry* **4**, 187–217.
- 232 Oostenbrink C, Villa A, Mark AE & Van Gunsteren WF (2004) A biomolecular force field based on the free enthalpy of hydration and solvation: The GROMOS force-field parameter sets 53A5 and 53A6. *Journal of Computational Chemistry* **25**, 1656–1676.

- 233 Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE & Berendsen HJC (2005) GROMACS: fast, flexible, and free. *J Comput Chem* **26**, 1701–1718.
- 234 Hess B, Kutzner C, Van Der Spoel D & Lindahl E (2008) GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comp.* **4**, 435–447.
- 235 Eswar N, Eramian D, Webb B, Shen M-Y & Sali A (2008) Protein structure modeling with MODELLER. *Methods Mol. Biol.* **426**, 145–159.
- 236 Walters R & DeGrado W (2006) Helix-packing motifs in membrane proteins. *Proceedings of the National Academy of Sciences* **103**, 13658–13663.
- 237 Kandt C, Ash WL & Tieleman DP (2007) Setting up and running molecular dynamics simulations of membrane proteins. *Methods* **41**, 475–488.
- 238 Berendsen H, Postma J, van Gunsteren W & Hermans J (1981) Interaction models for water in relation to protein hydration. *Intermolecular Forces*, 331–342.
- 239 Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A & Haak JR (1984) Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics* **81**, 3684.
- 240 Nosé S (1984) A unified formulation of the constant temperature molecular dynamics methods. *The Journal of Chemical Physics* **81**, 511–519.
- 241 Hoover (1985) Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev., A* **31**, 1695–1697.
- 242 Parrinello M (1981) Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied Physics* **52**, 7182.
- 243 Nosé S & Klein ML (1983) Constant pressure molecular dynamics for molecular systems. *Molecular Physics* **50**, 1055–1076.
- 244 E N, S F & M K (1996) Simulation of Activation Free-Energies in Molecular-Systems. *J. Chem. Phys.* **105**, 1902–1921.
- 245 Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW & Klein ML (1983) Comparison of simple potential functions for simulating liquid water. *The Journal of chemical physics* **79**, 926.
- 246 Hornak V, Abel R, Okur A, Strockbine B, Roitberg A & Simmerling C (2006) Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins: Structure, Function, and Bioinformatics* **65**, 712–725.
- 247 van der Spoel D, Lindahl E, Hess B, van Buuren A, Apol E, Meulenhoff P, Tieleman D, Sijbers A, Feenstra K & van Drunen R (2010) Gromacs User Manual version 4.5. 4. .
- 248 Byrd RH, Lu P, Nocedal J & Zhu C (1995) A limited memory algorithm for bound constrained optimization. *SIAM Journal on Scientific Computing* **16**, 1190–1208.
- 249 Zhu C, Byrd RH, Lu P & Nocedal J (1997) Algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound-constrained optimization. *ACM Transactions on Mathematical Software (TOMS)* **23**, 550–560.
- 250 Darden T, York D & Pedersen L (1993) Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *The Journal of Chemical Physics* **98**, 10089–10092.
- 251 Essmann U, Perera L, Berkowitz M, Darden T, Lee H & Pedersen L (1995) A smooth particle mesh Ewald method. *The Journal of Chemical Physics* **103**, 8577–8593.
- 252 Miyamoto S & Kollman PA (1992) SETTLE: an analytical version of the SHAKE and RATTLE algorithm for rigid water models. *Journal of computational chemistry* **13**, 952–962.
- 253 Delano W (2002) The PyMOL Molecular Graphics System. *Palo Alto, CA: DeLano Scientific LLC.*
- 254 Humphrey W, Dalke A & Schulten K (1996) VMD: visual molecular dynamics. *J Mol Graph* **14**, 33–38, 27–28.
- 255 Mandell DJ, Coutsias EA & Kortemme T (2009) Sub-angstrom accuracy in protein loop reconstruction by robotics-inspired conformational sampling. *Nat. Methods* **6**, 551–552.
- 256 Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797.
- 257 Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ & Higgins DG (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947–2948.
- 258 Gouy M, Guindon S & Gascuel O (2010) SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol. Biol. Evol.* **27**, 221–224.
- 259 Ahn KH, Bertalovitz AC, Mierke DF & Kendall DA (2009) Dual role of the second extracellular loop of the cannabinoid receptor 1: ligand binding and receptor localization. *Mol. Pharmacol.* **76**, 833–842.
- 260 Shen M-Y & Sali A (2006) Statistical potential for assessment and prediction of protein structures. *Protein Sci.* **15**, 2507–2524.
- 261 Ligprep V 2.4,(2010). *Schrodinger, LLC, New York.*
- 262 Shelley JC, Cholleti A, Frye LL, Greenwood JR, Timlin MR & Uchimaya M (2007) Epik: a software program for pK a prediction and protonation state generation for drug-like molecules. *Journal of computer-aided molecular design* **21**, 681–691.
- 263 Reggio PH (2003) Pharmacophores for ligand recognition and activation/inactivation of the cannabinoid receptors. *Curr. Pharm. Des.* **9**, 1607–1633.

- 264 Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS & Olson AJ (2009) AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J Comput Chem* **30**, 2785–2791.
- 265 Ballesteros J & Weinstein H (1995) Integrated methods for modeling G-protein coupled receptors. *Methods Neurosci* **25**, 366–428.
- 266 Cherezov V, Rosenbaum DM, Hanson MA, Rasmussen SGF, Thian FS, Kobilka TS, Choi H-J, Kuhn P, Weis WI, Kobilka BK & Stevens RC (2007) High-resolution crystal structure of an engineered human beta2-adrenergic G protein-coupled receptor. *Science* **318**, 1258–1265.
- 267 Kolinski A (2004) Protein modeling and structure prediction with a reduced representation. *Acta Biochim. Pol.* **51**, 349–371.
- 268 Kukol A (2009) Lipid models for united-atom molecular dynamics simulations of proteins. *Journal of Chemical Theory and Computation* **5**, 615–626.
- 269 van der Spoel D, van Maaren PJ & Berendsen HJC (1998) A systematic study of water models for molecular simulation: Derivation of water models optimized for use with a reaction field. *The Journal of Chemical Physics* **108**, 10220.
- 270 Schüttelkopf AW & van Aalten DMF (2004) PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 1355–1363.
- 271 Altschul SF, Gish W, Miller W, Myers EW & Lipman DJ (1990) Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410.
- 272 Sokoloff P, Giros B, Martres MP, Bouthenet ML & Schwartz JC (1990) Molecular cloning and characterization of a novel dopamine receptor (D3) as a target for neuroleptics. *Nature* **347**, 146–151.
- 273 Chabot DJ, Zhang PF, Quinnan GV & Broder CC (1999) Mutagenesis of CXCR4 identifies important domains for human immunodeficiency virus type 1 X4 isolate envelope-mediated membrane fusion and virus entry and reveals cryptic coreceptor activity for R5 isolates. *J. Virol.* **73**, 6598–6609.
- 274 Martelle JL & Nader MA (2008) A review of the discovery, pharmacological characterization, and behavioral effects of the dopamine D2-like receptor antagonist eticlopride. *CNS Neurosci Ther* **14**, 248–262.
- 275 Thoma G, Streiff MB, Kovarik J, Glickman F, Wagner T, Beerli C & Zerwes H-G (2008) Orally bioavailable isothioureas block function of the chemokine receptor CXCR4 in vitro and in vivo. *J. Med. Chem.* **51**, 7915–7920.
- 276 Glide V 5.6,(2010). *Schrodinger, LLC, New York.*
- 277 Oh YS & Turner RJ (2005) Topology of the C-terminal fragment of human presenilin 1. *Biochemistry* **44**, 11821–11828.
- 278 Spasic D, Tolia A, Dillen K, Baert V, De Strooper B, Vrijens S & Annaert W (2006) Presenilin-1 maintains a nine-transmembrane topology throughout the secretory pathway. *J. Biol. Chem.* **281**, 26569–26577.
- 279 Hastings WK (1970) Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57**, 97–109.
- 280 Kozakov D, Hall DR, Beglov D, Brenke R, Comeau SR, Shen Y, Li K, Zheng J, Vakili P, Paschalidis IC & Vajda S (2010) Achieving reliability and high accuracy in automated protein docking: ClusPro, PIPER, SDU, and stability analysis in CAPRI rounds 13-19. *Proteins* **78**, 3124–3130.
- 281 Tovchigrechko A & Vakser IA (2006) GRAMM-X public web server for protein-protein docking. *Nucleic acids research* **34**, W310–W314.
- 282 Dominguez C, Boelens R & Bonvin A (2003) HADDOCK: a protein-protein docking approach based on biochemical or biophysical data. *NMR-based docking of protein-protein complexes* **125**, 51.
- 283 Singh R, Barman A & Prabhakar R (2009) Computational insights into aspartyl protease activity of presenilin 1 (PS1) generating Alzheimer amyloid beta-peptides (Abeta40 and Abeta42). *J Phys Chem B* **113**, 2990–2999.
- 284 Joachims T (1999) SvmLight: Support vector machine. *SVM-Light Support Vector Machine* <http://svmlight.joachims.org/>, University of Dortmund **19**, 4.
- 285 Cheng Z, Garvin D, Paguio A, Stecha P, Wood K & Fan F (2010) Luciferase Reporter Assay System for Deciphering GPCR Pathways. *Curr Chem Genomics* **4**, 84–91.
- 286 Carpino PA & Goodwin B (2010) Diabetes area participation analysis: a review of companies and targets described in the 2008 - 2010 patent literature. *Expert Opin Ther Pat* **20**, 1627–1651.
- 287 Parma J, Duprez L, Van Sande J, Cochaux P, Gervy C, Mockel J, Dumont J & Vassart G (1993) Somatic mutations in the thyrotropin receptor gene cause hyperfunctioning thyroid adenomas. *Nature* **365**, 649–651.
- 288 Dragun D, Philippe A, Catar R & Hegner B (2009) Autoimmune mediated G-protein receptor activation in cardiovascular and renal pathologies. *Thromb. Haemost.* **101**, 643–648.
- 289 Vischer HF, Vink C & Smit MJ (2006) A viral conspiracy: hijacking the chemokine system through virally encoded pirated chemokine receptors. *Curr. Top. Microbiol. Immunol.* **303**, 121–154.
- 290 Lombardi MS, Kavelaars A & Heijnen CJ (2002) Role and modulation of G protein-coupled receptor signaling in inflammatory processes. *Crit. Rev. Immunol.* **22**, 141–163.
- 291 Shichida Y & Imai H (1998) Visual pigment: G-protein-coupled receptor for light signals. *Cell. Mol. Life Sci.* **54**, 1299–1315.
- 292 Stone LS & Molliver DC (2009) In search of analgesia: emerging roles of GPCRs in pain. *Mol. Interv.* **9**, 234–251.

- 293 Lappano R & Maggiolini M (2011) G protein-coupled receptors: novel targets for drug discovery in cancer. *Nat Rev Drug Discov* **10**, 47–60.
- 294 Latek D, Kolinski M, Ghoshdastider U, Debinski A, Bombolewski R, Plazinska A, Jozwiak K & Filipek S (2011) Modeling of ligand binding to G protein coupled receptors: cannabinoid CB1, CB2 and adrenergic  $\beta$  2 AR. *J Mol Model* **17**, 2353–2366.
- 295 Trzaskowski B, Latek D, Yuan S, Ghoshdastider U, Debinski A & Filipek S (2012) Action of molecular switches in GPCRs--theoretical and experimental studies. *Curr. Med. Chem.* **19**, 1090–1109.
- 296 Yuan S, Ghoshdastider U, Trzaskowski B, Latek D, Debinski A, Pulawski W, Wu R, Gerke V & Filipek S (2012) The role of water in activation mechanism of human Nformyl Peptide Receptor 1 (FPR1) based on molecular dynamics simulations. *PloS One*.
- 297 Palczewski K, Kumasaka T, Hori T, Behnke CA, Motoshima H, Fox BA, Le Trong I, Teller DC, Okada T, Stenkamp RE, Yamamoto M & Miyano M (2000) Crystal structure of rhodopsin: A G protein-coupled receptor. *Science* **289**, 739–745.
- 298 Warne T, Serrano-Vega MJ, Baker JG, Moukhametzianov R, Edwards PC, Henderson R, Leslie AGW, Tate CG & Schertler GFX (2008) Structure of a beta1-adrenergic G-protein-coupled receptor. *Nature* **454**, 486–491.
- 299 Tanuj Sapra K, Park PS-H, Filipek S, Engel A, Müller DJ & Palczewski K (2006) Detecting molecular interactions that stabilize native bovine rhodopsin. *J. Mol. Biol* **358**, 255–269.
- 300 Chung KY, Rasmussen SGF, Liu T, Li S, DeVree BT, Chae PS, Calinski D, Kobilka BK, Woods VL Jr & Sunahara RK (2011) Conformational changes in the G protein Gs induced by the  $\beta$ 2 adrenergic receptor. *Nature* **477**, 611–615.
- 301 Rasmussen SGF, DeVree BT, Zou Y, Kruse AC, Chung KY, Kobilka TS, Thian FS, Chae PS, Pardon E, Calinski D, Mathiesen JM, Shah STA, Lyons JA, Caffrey M, Gellman SH, Steyaert J, Skiniotis G, Weis WI, Sunahara RK & Kobilka BK (2011) Crystal structure of the  $\beta$ 2 adrenergic receptor-Gs protein complex. *Nature* **477**, 549–555.
- 302 Unger VM, Hargrave PA, Baldwin JM & Schertler GF (1997) Arrangement of rhodopsin transmembrane alpha-helices. *Nature* **389**, 203–206.
- 303 Ahuja S & Smith SO (2009) Multiple switches in G protein-coupled receptor activation. *Trends Pharmacol. Sci.* **30**, 494–502.
- 304 Nygaard R, Frimurer TM, Holst B, Rosenkilde MM & Schwartz TW (2009) Ligand binding and micro-switches in 7TM receptor structures. *Trends Pharmacol. Sci.* **30**, 249–259.
- 305 Hofmann KP (2009) A G protein-coupled receptor at work: the rhodopsin model. *Trends Biochem. Sci.* **34**, 540–552.
- 306 Strange PG (2008) Signaling mechanisms of GPCR ligands. *Curr Opin Drug Discov Devel* **11**, 196–202.
- 307 Deupi X & Standfuss J Structural insights into agonist-induced activation of G-protein-coupled receptors. *Current Opinion in Structural Biology* **In Press, Corrected Proof**.
- 308 Unal H & Karnik SS (2012) Domain coupling in GPCRs: the engine for induced conformational changes. *Trends Pharmacol. Sci.* **33**, 79–88.
- 309 Schiöth HB & Fredriksson R (2005) The GRAFS classification system of G-protein coupled receptors in comparative perspective. *Gen. Comp. Endocrinol.* **142**, 94–101.
- 310 Attwood TK & Findlay JB (1994) Fingerprinting G-protein-coupled receptors. *Protein Eng.* **7**, 195–203.
- 311 Kolakowski LF Jr (1994) GCRDb: a G-protein-coupled receptor database. *Recept. Channels* **2**, 1–7.
- 312 Strotmann R, Schröck K, Bösel I, Stäubert C, Russ A & Schöneberg T (2011) Evolution of GPCR: change and continuity. *Mol. Cell. Endocrinol.* **331**, 170–178.
- 313 Surgand J-S, Rodrigo J, Kellenberger E & Rognan D (2006) A chemogenomic analysis of the transmembrane binding cavity of human G-protein-coupled receptors. *Proteins* **62**, 509–538.
- 314 Chabbert M, Castel H, Pele J, Deville J, Legendre R & Rodien P (2012) Evolution of class A G-protein-coupled receptors: implications for molecular modeling. *Curr. Med. Chem.* **19**, 1110–1118.
- 315 Pelé J, Abdi H, Moreau M, Thybert D & Chabbert M (2011) Multidimensional scaling reveals the main evolutionary pathways of class A G-protein-coupled receptors. *PLoS ONE* **6**, e19094.
- 316 Wu B, Chien EYT, Mol CD, Fenalti G, Liu W, Katritch V, Abagyan R, Brooun A, Wells P, Bi FC, Hamel DJ, Kuhn P, Handel TM, Cherezov V & Stevens RC (2010) Structures of the CXCR4 chemokine GPCR with small-molecule and cyclic peptide antagonists. *Science* **330**, 1066–1071.
- 317 Wicher D, Schäfer R, Bauernfeind R, Stensmyr MC, Heller R, Heinemann SH & Hansson BS (2008) Drosophila odorant receptors are both ligand-gated and cyclic-nucleotide-activated cation channels. *Nature* **452**, 1007–1011.
- 318 Luecke H, Richter HT & Lanyi JK (1998) Proton transfer pathways in bacteriorhodopsin at 2.3 angstrom resolution. *Science* **280**, 1934–1937.
- 319 Fuhrman JA, Schwabach MS & Stingl U (2008) Proteorhodopsins: an array of physiological roles? *Nat. Rev. Microbiol.* **6**, 488–494.
- 320 Essen L-O (2002) Halorhodopsin: light-driven ion pumping made simple? *Curr. Opin. Struct. Biol.* **12**, 516–522.
- 321 Levoye A, Dam J, Ayoub MA, Guillaume J-L, Couturier C, Delagrèze P & Jockers R (2006) The orphan GPR50 receptor specifically inhibits MT1 melatonin receptor function through heterodimerization. *EMBO J.* **25**, 3012–3023.
- 322 Warne T (2011) The structural basis for agonist and partial agonist action on a [bgr]1-adrenergic receptor. *Nature*

- 469, 241–244.
- 323 Deupi X & Kobilka BK (2010) Energy landscapes as a tool to integrate GPCR structure, dynamics, and function. *Physiology (Bethesda)* **25**, 293–303.
- 324 Smith NJ, Bennett KA & Milligan G (2011) When simple agonism is not enough: emerging modalities of GPCR ligands. *Mol. Cell. Endocrinol.* **331**, 241–247.
- 325 Milligan G (2009) G protein-coupled receptor hetero-dimerization: contribution to pharmacology and function. *Br. J. Pharmacol.* **158**, 5–14.
- 326 Whorton MR (2007) A monomeric G protein-coupled receptor isolated in a high-density lipoprotein particle efficiently activates its G protein. *Proc. Natl Acad. Sci. USA* **104**, 7682–7687.
- 327 Bayburt TH, Vishnivetskiy SA, McLean MA, Morizumi T, Huang C-C, Tesmer JJG, Ernst OP, Sligar SG & Gurevich VV (2011) Monomeric rhodopsin is sufficient for normal rhodopsin kinase (GRK1) phosphorylation and arrestin-1 binding. *J. Biol. Chem.* **286**, 1420–1428.
- 328 Pin J-P, Galvez T & Prézeau L (2003) Evolution, structure, and activation mechanism of family 3/C G-protein-coupled receptors. *Pharmacol. Ther.* **98**, 325–354.
- 329 Harikumar KG, Morfis MM, Sexton PM & Miller LJ (2008) Pattern of intra-family hetero-oligomerization involving the G-protein-coupled secretin receptor. *J. Mol. Neurosci.* **36**, 279–285.
- 330 Kuszak AJ, Pitchiaya S, Anand JP, Mosberg HI, Walter NG & Sunahara RK (2009) Purification and functional reconstitution of monomeric mu-opioid receptors: allosteric modulation of agonist binding by Gi2. *J. Biol. Chem.* **284**, 26732–26741.
- 331 Milligan G & Bouvier M (2005) Methods to monitor the quaternary structure of G protein-coupled receptors. *FEBS J.* **272**, 2914–2925.
- 332 Fung JJ, Deupi X, Pardo L, Yao XJ, Velez-Ruiz GA, Devree BT, Sunahara RK & Kobilka BK (2009) Ligand-regulated oligomerization of beta(2)-adrenoceptors in a model lipid bilayer. *EMBO J.* **28**, 3315–3328.
- 333 Han Y, Moreira IS, Urizar E, Weinstein H & Javitch JA (2009) Allosteric communication between protomers of dopamine class A GPCR dimers modulates activation. *Nat. Chem. Biol.* **5**, 688–695.
- 334 Rondard P, Huang S, Monnier C, Tu H, Blanchard B, Oueslati N, Malhaire F, Li Y, Trinquet E, Labesse G, Pin J-P & Liu J (2008) Functioning of the dimeric GABA(B) receptor extracellular domain revealed by glycan wedge scanning. *EMBO J.* **27**, 1321–1332.
- 335 Canals M, Sexton PM & Christopoulos A (2011) Allostery in GPCRs: “MWC” revisited. *Trends Biochem. Sci.* **36**, 663–672.
- 336 Galvez T, Duthey B, Kniazeff J, Blahos J, Rovelli G, Bettler B, Prézeau L & Pin JP (2001) Allosteric interactions between GB1 and GB2 subunits are required for optimal GABA(B) receptor function. *EMBO J.* **20**, 2152–2159.
- 337 Maurel D, Comps-Agrar L, Brock C, Rives M-L, Bourrier E, Ayoub MA, Bazin H, Tinel N, Durroux T, Prézeau L, Trinquet E & Pin J-P (2008) Cell-surface protein-protein interaction analysis with time-resolved FRET and snap-tag technologies: application to GPCR oligomerization. *Nat. Methods* **5**, 561–567.
- 338 Smith NJ & Milligan G (2010) Allostery at G protein-coupled receptor homo- and heteromers: uncharted pharmacological landscapes. *Pharmacol. Rev.* **62**, 701–725.
- 339 Hanson MA, Cherezov V, Griffith MT, Roth CB, Jaakola V-P, Chien EYT, Velasquez J, Kuhn P & Stevens RC (2008) A specific cholesterol binding site is established by the 2.8 Å structure of the human beta2-adrenergic receptor. *Structure* **16**, 897–905.
- 340 James JR, Oliveira MI, Carmo AM, Iaboni A & Davis SJ (2006) A rigorous experimental framework for detecting protein oligomerization using bioluminescence resonance energy transfer. *Nat. Methods* **3**, 1001–1006.
- 341 Katritch V, Cherezov V & Stevens RC (2012) Diversity and modularity of G protein-coupled receptor structures. *Trends Pharmacol. Sci.* **33**, 17–27.
- 342 Liang Y, Fotiadis D, Filipek S, Saperstein DA, Palczewski K & Engel A (2003) Organization of the G protein-coupled receptors rhodopsin and opsin in native membranes. *J. Biol. Chem.* **278**, 21655–21662.
- 343 Filipek S, Krzysko KA, Fotiadis D, Liang Y, Saperstein DA, Engel A & Palczewski K (2004) A concept for G protein activation by G protein-coupled receptor dimers: the transducin/rhodopsin interface. *Photochem. Photobiol. Sci.* **3**, 628–638.
- 344 Ballesteros JA, Jensen AD, Liapakis G, Rasmussen SG, Shi L, Gether U & Javitch JA (2001) Activation of the beta 2-adrenergic receptor involves disruption of an ionic lock between the cytoplasmic ends of transmembrane segments 3 and 6. *J. Biol. Chem.* **276**, 29171–29177.
- 345 Visiers I, Ballesteros JA & Weinstein H (2002) Three-dimensional representations of G protein-coupled receptor structures and mechanisms. *Meth. Enzymol.* **343**, 329–371.
- 346 Chien EY (2010) Structure of the human dopamine D3 receptor in complex with a D2/D3 selective antagonist. *Science* **330**, 1091–1095.
- 347 Doré AS, Robertson N, Errey JC, Ng I, Hollenstein K, Tehan B, Hurrell E, Bennett K, Congreve M, Magnani F, Tate CG, Weir M & Marshall FH (2011) Structure of the adenosine A(2A) receptor in complex with ZM241385 and the xanthines XAC and caffeine. *Structure* **19**, 1283–1293.
- 348 Shimamura T, Shiroishi M, Weyand S, Tsujimoto H, Winter G, Katritch V, Abagyan R, Cherezov V, Liu W, Han GW, Kobayashi T, Stevens RC & Iwata S (2011) Structure of the human histamine H1 receptor complex with

- doxepin. *Nature* **475**, 65–70.
- 349 Kolinski M & Filipek S (2008) Molecular dynamics of mu opioid receptor complexes with agonists and antagonists. *TOSBJ* **2**, 8–20.
- 350 Kolinski M & Filipek S (2009) Studies of the activation steps concurrent to ligand binding in DOR and KOR opioid receptors based on molecular dynamics simulations. *TOSBJ* **3**, 51–63.
- 351 Kolinski M & Filipek S (2010) Study of a structurally similar kappa opioid receptor agonist and antagonist pair by molecular dynamics simulations. *J Mol Model* **16**, 1567–1576.
- 352 Kim J-M, Altenbach C, Kono M, Oprian DD, Hubbell WL & Khorana HG (2004) Structural origins of constitutive activation in rhodopsin: Role of the K296/E113 salt bridge. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 12508–12513.
- 353 Choe H-W, Kim YJ, Park JH, Morizumi T, Pai EF, Krausz N, Hofmann KP, Scheerer P & Ernst OP (2011) Crystal structure of metarhodopsin II. *Nature* **471**, 651–655.
- 354 Standfuss J, Edwards PC, D'Antona A, Fransen M, Xie G, Oprian DD & Schertler GFX (2011) The structural basis of agonist-induced activation in constitutively active rhodopsin. *Nature* **471**, 656–660.
- 355 Li J, Edwards PC, Burghammer M, Villa C & Schertler GFX (2004) Structure of bovine rhodopsin in a trigonal crystal form. *J. Mol. Biol.* **343**, 1409–1438.
- 356 Fritze O, Filipek S, Kuksa V, Palczewski K, Hofmann KP & Ernst OP (2003) Role of the conserved NPxxY(x)5,6F motif in the rhodopsin ground state and during activation. *Proc. Natl. Acad. Sci. U.S.A* **100**, 2290–2295.
- 357 Rasmussen SG (2011) Structure of a nanobody-stabilized active state of the [bgr]2 adrenoceptor. *Nature* **469**, 175–180.
- 358 Schwartz TW, Frimurer TM, Holst B, Rosenkilde MM & Elling CE (2006) Molecular mechanism of 7TM receptor activation—a global toggle switch model. *Annu. Rev. Pharmacol. Toxicol.* **46**, 481–519.
- 359 Elling CE (2006) Metal ion site engineering indicates a global toggle switch model for seven-transmembrane receptor activation. *J. Biol. Chem.* **281**, 17337–17346.
- 360 Cohen GB, Oprian DD & Robinson PR (1992) Mechanism of activation and inactivation of opsin: role of Glu113 and Lys296. *Biochemistry* **31**, 12592–12601.
- 361 Nichols DE & Nichols CD (2008) Serotonin receptors. *Chem. Rev.* **108**, 1614–1641.
- 362 Kristiansen K (2004) Molecular mechanisms of ligand binding, signaling, and regulation within the superfamily of G-protein-coupled receptors: molecular modeling and mutagenesis approaches to receptor structure and function. *Pharmacol. Ther.* **103**, 21–80.
- 363 Chelikani P, Hornak V, Eilers M, Reeves PJ, Smith SO, RajBhandary UL & Khorana HG (2007) Role of group-conserved residues in the helical core of beta2-adrenergic receptor. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 7027–7032.
- 364 McAllister SD, Hurst DP, Barnett-Norris J, Lynch D, Reggio PH & Abood ME (2004) Structural mimicry in class A G protein-coupled receptor rotamer toggle switches: the importance of the F3.36(201)/W6.48(357) interaction in cannabinoid CB1 receptor activation. *J. Biol. Chem.* **279**, 48024–48037.
- 365 Moore S, Jaeschke H, Kleinau G, Neumann S, Costanzi S, Jiang J, Childress J, Raaka BM, Colson A, Paschke R, Krause G, Thomas CJ & Gershengorn MC (2006) Evaluation of small-molecule modulators of the luteinizing hormone/choriogonadotropin and thyroid stimulating hormone receptors: structure-activity relationships and selective binding patterns. *J. Med. Chem.* **49**, 3888–3896.
- 366 Salom D (2006) Crystal structure of a photoactivated deprotonated intermediate of rhodopsin. *Proc. Natl Acad. Sci. USA* **103**, 16123–16128.
- 367 Bokoch MP, Zou Y, Rasmussen SGF, Liu CW, Nygaard R, Rosenbaum DM, Fung JJ, Choi H-J, Thian FS, Kobilka TS, Puglisi JD, Weis WI, Pardo L, Prosser RS, Mueller L & Kobilka BK (2010) Ligand-specific regulation of the extracellular surface of a G-protein-coupled receptor. *Nature* **463**, 108–112.
- 368 Ahuja S, Hornak V, Yan ECY, Syrett N, Goncalves JA, Hirshfeld A, Ziliox M, Sakmar TP, Sheves M, Reeves PJ, Smith SO & Eilers M (2009) Helix movement is coupled to displacement of the second extracellular loop in rhodopsin activation. *Nat. Struct. Mol. Biol.* **16**, 168–175.
- 369 Park JH, Scheerer P, Hofmann KP, Choe HW & Ernst OP (2008) Crystal structure of the ligand-free G-protein-coupled receptor opsin. *Nature* **454**, 183–187.
- 370 Schadel SA, Heck M, Maretzki D, Filipek S, Teller DC, Palczewski K & Hofmann KP (2003) Ligand channeling within a G-protein-coupled receptor. The entry and exit of retinals in native opsin. *J. Biol. Chem* **278**, 24896–24903.
- 371 Prioleau C, Visiers I, Ebersole BJ, Weinstein H & Sealfon SC (2002) Conserved helix 7 tyrosine acts as a multistate conformational switch in the 5HT2C receptor. Identification of a novel “locked-on” phenotype and double revertant mutations. *J. Biol. Chem.* **277**, 36577–36584.
- 372 Nakamichi H & Okada T (2006) Local peptide movement in the photoreaction intermediate of rhodopsin. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 12729–12734.
- 373 Shi L, Liapakis G, Xu R, Guarneri F, Ballesteros JA & Javitch JA (2002) Beta2 adrenergic receptor activation. Modulation of the proline kink in transmembrane 6 by a rotamer toggle switch. *J. Biol. Chem.* **277**, 40989–40996.
- 374 Chabre M & Breton J (1979) Orientation of aromatic residues in rhodopsin. Rotation of one tryptophan upon the

- meta I to meta II transition after illumination. *Photochem. Photobiol.* **30**, 295–299.
- 375 Greasley PJ, Fanelli F, Rossier O, Abuin L & Cotecchia S (2002) Mutagenesis and modelling of the alpha(1b)-adrenergic receptor highlight the role of the helix 3/helix 6 interface in receptor activation. *Mol. Pharmacol.* **61**, 1025–1032.
- 376 Röhrig UF, Guidoni L & Rothlisberger U (2002) Early steps of the intramolecular signal transduction in rhodopsin explored by molecular dynamics simulations. *Biochemistry* **41**, 10799–10809.
- 377 Crozier PS, Stevens MJ, Forrest LR & Woolf TB (2003) Molecular dynamics simulation of dark-adapted rhodopsin in an explicit membrane bilayer: coupling between local retinal and larger scale conformational change. *J. Mol. Biol.* **333**, 493–514.
- 378 Pogozheva ID, Przydzial MJ & Mosberg HI (2005) Homology modeling of opioid receptor-ligand complexes using experimental constraints. *AAPS J* **7**, E434–448.
- 379 Huang P, Visiers I, Weinstein H & Liu-Chen L-Y (2002) The local environment at the cytoplasmic end of TM6 of the mu opioid receptor differs from those of rhodopsin and monoamine receptors: introduction of an ionic lock between the cytoplasmic ends of helices 3 and 6 by a L6.30(275)E mutation inactivates the mu opioid receptor and reduces the constitutive activity of its T6.34(279)K mutant. *Biochemistry* **41**, 11972–11980.
- 380 Provasi D, Bortolato A & Filizola M (2009) Exploring molecular mechanisms of ligand recognition by opioid receptors with metadynamics. *Biochemistry* **48**, 10020–10029.
- 381 Niv MY, Skrabanek L, Filizola M & Weinstein H (2006) Modeling activated states of GPCRs: the rhodopsin template. *J. Comput. Aided Mol. Des.* **20**, 437–448.
- 382 Lin SW & Sakmar TP (1996) Specific tryptophan UV-absorbance changes are probes of the transition of rhodopsin to its active state. *Biochemistry* **35**, 11149–11159.
- 383 Balaraman GS, Bhattacharya S & Vaidehi N (2010) Structural insights into conformational stability of wild-type and mutant beta1-adrenergic receptor. *Biophys. J.* **99**, 568–577.
- 384 Crocker E, Eilers M, Ahuja S, Hornak V, Hirshfeld A, Sheves M & Smith SO (2006) Location of Trp265 in metarhodopsin II: implications for the activation mechanism of the visual receptor rhodopsin. *J. Mol. Biol.* **357**, 163–172.
- 385 Bhattacharya S, Hall SE & Vaidehi N (2008) Agonist-induced conformational changes in bovine rhodopsin: insight into activation of G-protein-coupled receptors. *J. Mol. Biol.* **382**, 539–555.
- 386 Hornak V, Ahuja S, Eilers M, Goncalves JA, Sheves M, Reeves PJ & Smith SO (2010) Light activation of rhodopsin: insights from molecular dynamics simulations guided by solid-state NMR distance restraints. *J. Mol. Biol.* **396**, 510–527.
- 387 Holst B, Nygaard R, Valentin-Hansen L, Bach A, Engelstoft MS, Petersen PS, Frimurer TM & Schwartz TW (2010) A conserved aromatic lock for the tryptophan rotameric switch in TM-VI of seven-transmembrane receptors. *J. Biol. Chem.* **285**, 3973–3985.
- 388 Dror RO, Arlow DH, Borhani DW, Jensen MØ, Piana S & Shaw DE (2009) Identification of two distinct inactive conformations of the beta2-adrenergic receptor reconciles structural and biochemical observations. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 4689–4694.
- 389 Vanni S, Neri M, Tavernelli I & Rothlisberger U (2009) Observation of “ionic lock” formation in molecular dynamics simulations of wild-type beta 1 and beta 2 adrenergic receptors. *Biochemistry* **48**, 4789–4797.
- 390 Bhattacharya S, Hall SE, Li H & Vaidehi N (2008) Ligand-stabilized conformational states of human beta(2) adrenergic receptor: insight into G-protein-coupled receptor activation. *Biophys. J.* **94**, 2027–2042.
- 391 Romo TD, Grossfield A & Pitman MC (2010) Concerted interconversion between ionic lock substates of the beta(2) adrenergic receptor revealed by microsecond timescale molecular dynamics. *Biophys. J.* **98**, 76–84.
- 392 Moukhametzianov R, Warne T, Edwards PC, Serrano-Vega MJ, Leslie AGW, Tate CG & Schertler GFX (2011) Two distinct conformations of helix 6 observed in antagonist-bound structures of a beta1-adrenergic receptor. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 8228–8232.
- 393 Neri M, Vanni S, Tavernelli I & Rothlisberger U (2010) Role of aggregation in rhodopsin signal transduction. *Biochemistry* **49**, 4827–4832.
- 394 Sgourakis NG & Garcia AE (2010) The membrane complex between transducin and dark-state rhodopsin exhibits large-amplitude interface dynamics on the sub-microsecond timescale: insights from all-atom MD simulations. *J. Mol. Biol.* **398**, 161–173.
- 395 Elgeti M, Kazmin R, Heck M, Morizumi T, Ritter E, Scheerer P, Ernst OP, Siebert F, Hofmann KP & Bartl FJ (2011) Conserved Tyr223(5.58) plays different roles in the activation and G-protein interaction of rhodopsin. *J. Am. Chem. Soc.* **133**, 7159–7165.
- 396 Kobilka BK (2011) Structural insights into adrenergic receptor function and pharmacology. *Trends Pharmacol. Sci.* **32**, 213–218.
- 397 Wisler JW (2007) A unique mechanism of beta-blocker action: carvedilol stimulates beta-arrestin signaling. *Proc. Natl. Acad. Sci. USA* **104**, 16657–16662.
- 398 Kenakin T (2003) Ligand-selective receptor conformations revisited: the promise and the problem. *Trends Pharmacol. Sci.* **24**, 346–354.
- 399 Rasmussen SGF, Choi H-J, Rosenbaum DM, Kobilka TS, Thian FS, Edwards PC, Burghammer M, Ratnala VRP,

- Sanishvili R, Fischetti RF, Schertler GFX, Weis WI & Kobilka BK (2007) Crystal structure of the human beta2 adrenergic G-protein-coupled receptor. *Nature* **450**, 383–387.
- 400 Rosenbaum DM, Cherezov V, Hanson MA, Rasmussen SGF, Thian FS, Kobilka TS, Choi H-J, Yao X-J, Weis WI, Stevens RC & Kobilka BK (2007) GPCR engineering yields high-resolution structural insights into beta2-adrenergic receptor function. *Science* **318**, 1266–1273.
- 401 Jaakola V-P, Griffith MT, Hanson MA, Cherezov V, Chien EYT, Lane JR, Ijzerman AP & Stevens RC (2008) The 2.6 angstrom crystal structure of a human A2A adenosine receptor bound to an antagonist. *Science* **322**, 1211–1217.
- 402 Porter JE, Hwa J & Perez DM (1996) Activation of the alpha1b-adrenergic receptor is initiated by disruption of an interhelical salt bridge constraint. *J. Biol. Chem.* **271**, 28318–28323.
- 403 Scheerer P, Park JH, Hildebrand PW, Kim YJ, Krauss N, Choe H-W, Hofmann KP & Ernst OP (2008) Crystal structure of opsin in its G-protein-interacting conformation. *Nature* **455**, 497–502.
- 404 Jozwiak K, Khalid C, Tanga MJ, Berzetei-Gurske I, Jimenez L, Kozocas JA, Woo A, Zhu W, Xiao R-P, Abernethy DR & Wainer IW (2007) Comparative molecular field analysis of the binding of the stereoisomers of fenoterol and fenoterol derivatives to the beta2 adrenergic receptor. *J. Med. Chem.* **50**, 2903–2915.
- 405 Woo AY-H, Wang T-B, Zeng X, Zhu W, Abernethy DR, Wainer IW & Xiao R-P (2009) Stereochemistry of an agonist determines coupling preference of beta2-adrenoceptor to different G proteins in cardiomyocytes. *Mol. Pharmacol.* **75**, 158–165.
- 406 Gouldson PR, Higgs C, Smith RE, Dean MK, Gkoutos GV & Reynolds CA (2000) Dimerization and Domain Swapping in G-Protein-Coupled Receptors:: A Computational Study. *Neuropsychopharmacology* **23**, S60–S77.
- 407 Song ZH, Slowey CA, Hurst DP & Reggio PH (1999) The difference between the CB(1) and CB(2) cannabinoid receptors at position 5.46 is crucial for the selectivity of WIN55212-2 for CB(2). *Mol. Pharmacol.* **56**, 834–840.
- 408 Jeffrey GA (1997) *An Introduction to Hydrogen Bonding*, 1st ed. Oxford University Press, USA.
- 409 Gonzalez A, Duran LS, Araya-Secchi R, Garate JA, Pessoa-Mahana CD, Lagos CF & Perez-Acle T (2008) Computational modeling study of functional microdomains in cannabinoid receptor type 1. *Bioorg. Med. Chem.* **16**, 4378–4389.
- 410 McAllister SD, Rizvi G, Anavi-Goffer S, Hurst DP, Barnett-Norris J, Lynch DL, Reggio PH & Abood ME (2003) An aromatic microdomain at the cannabinoid CB(1) receptor constitutes an agonist/inverse agonist binding region. *J. Med. Chem.* **46**, 5139–5152.
- 411 Durdagi S, Papadopoulos MG, Zoumpoulakis PG, Koukoulitsa C & Mavromoustakos T (2010) A computational study on cannabinoid receptors and potent bioactive cannabinoid ligands: homology modeling, docking, de novo drug design and molecular dynamics analysis. *Mol. Divers.* **14**, 257–276.
- 412 Shim J-Y (2009) Transmembrane helical domain of the cannabinoid CB1 receptor. *Biophys. J.* **96**, 3251–3262.
- 413 Poso A & Huffman JW (2008) Targeting the cannabinoid CB2 receptor: modelling and structural determinants of CB2 selective ligands. *Br. J. Pharmacol.* **153**, 335–346.
- 414 Tuccinardi T, Ferrarini PL, Manera C, Ortore G, Saccomanni G & Martinelli A (2006) Cannabinoid CB2/CB1 selectivity. Receptor modeling and automated docking analysis. *J. Med. Chem.* **49**, 984–994.
- 415 Ghanouni P, Gryczynski Z, Steenhuis JJ, Lee TW, Farrens DL, Lakowicz JR & Kobilka BK (2001) Functionally Different Agonists Induce Distinct Conformations in the G Protein Coupling Domain of the  $\beta_2$  Adrenergic Receptor. *Journal of Biological Chemistry* **276**, 24433.
- 416 Swaminath G (2004) Sequential binding of agonists to the beta2 adrenoceptor. Kinetic evidence for intermediate conformational states. *J. Biol. Chem.* **279**, 686–691.
- 417 Nebane NM, Kellie B & Song Z-H (2006) The effects of charge-neutralizing mutation D6.30N on the functions of CB1 and CB2 cannabinoid receptors. *FEBS Lett.* **580**, 5392–5398.
- 418 Yousefi S, Cooper PR, Potter SL, Mueck B & Jarai G (2001) Cloning and expression analysis of a novel G-protein-coupled receptor selectively expressed on granulocytes. *J. Leukoc. Biol.* **69**, 1045–1052.
- 419 Zhang Y, Syed R, Uygur C, Pallos D, Gorry MC, Firatli E, Cortelli JR, VanDyke TE, Hart PS, Feingold E & Hart TC (2003) Evaluation of human leukocyte N-formylpeptide receptor (FPR1) SNPs in aggressive periodontitis patients. *Genes Immun.* **4**, 22–29.
- 420 Fu H, Karlsson J, Bylund J, Movitz C, Karlsson A & Dahlgren C (2006) Ligand recognition and activation of formyl peptide receptors in neutrophils. *J. Leukoc. Biol.* **79**, 247–256.
- 421 Rabiet M-J, Huet E & Boulay F (2007) The N-formyl peptide receptors and the anaphylatoxin C5a receptors: an overview. *Biochimie* **89**, 1089–1106.
- 422 Brandenburg L-O, Seyferth S, Wruck CJ, Koch T, Rosenstiel P, Lucius R & Pufe T (2009) Involvement of Phospholipase D 1 and 2 in the subcellular localization and activity of formyl-peptide-receptors in the human colonic cell line HT29. *Mol. Membr. Biol.* **26**, 371–383.
- 423 Filep JG & El Kebir D (2009) Neutrophil apoptosis: a target for enhancing the resolution of inflammation. *J. Cell. Biochem.* **108**, 1039–1046.
- 424 Jin Y, Arita M, Zhang Q, Saban DR, Chauhan SK, Chiang N, Serhan CN & Dana R (2009) Anti-angiogenesis effect of the novel anti-inflammatory and pro-resolving lipid mediators. *Invest. Ophthalmol. Vis. Sci.* **50**, 4743–4752.
- 425 Migeotte I, Communi D & Parmentier M (2006) Formyl peptide receptors: a promiscuous subfamily of G protein-coupled receptors controlling immune responses. *Cytokine Growth Factor Rev.* **17**, 501–519.



- 426 Ye RD, Boulay F, Wang JM, Dahlgren C, Gerard C, Parmentier M, Serhan CN & Murphy PM (2009) International Union of Basic and Clinical Pharmacology. LXXIII. Nomenclature for the formyl peptide receptor (FPR) family. *Pharmacol. Rev.* **61**, 119–161.
- 427 Okada T (2004) The retinal conformation and its environment in rhodopsin in light of a new 2.2 Å crystal structure. *J. Mol. Biol.* **342**, 571–583.
- 428 Fujita H, Kato T, Watanabe N, Takahashi T & Kitagawa S (2011) Stimulation of human formyl peptide receptors by calpain inhibitors: homology modeling of receptors and ligand docking simulation. *Arch. Biochem. Biophys.* **516**, 121–127.
- 429 Khlebnikov AI, Schepetkin IA, Kirpotina LN, Brive L, Dahlgren C, Jutila MA & Quinn MT (2012) Molecular docking of 2-(benzimidazol-2-ylthio)-N-phenylacetamide-derived small-molecule agonists of human formyl peptide receptor 1. *J. Mol. Model.* **18**, 2831–2843.
- 430 Movitz C, Brive L, Hellstrand K, Rabiet M-J & Dahlgren C (2010) The annexin I sequence gln(9)-ala(10)-trp(11)-phe(12) is a core structure for interaction with the formyl peptide receptor 1. *J. Biol. Chem.* **285**, 14338–14345.
- 431 Walther A, Riehemann K & Gerke V (2000) A novel ligand of the formyl peptide receptor: annexin I regulates neutrophil extravasation by interacting with the FPR. *Mol. Cell* **5**, 831–840.
- 432 Mills JS, Miettinen HM, Barnidge D, Vlases MJ, Wimer-Mackin S, Dratz EA, Sunner J & Jesaitis AJ (1998) Identification of a ligand binding site in the human neutrophil formyl peptide receptor using a site-specific fluorescent photoaffinity label and mass spectrometry. *J. Biol. Chem.* **273**, 10428–10435.
- 433 Manglik A, Kruse AC, Kobilka TS, Thian FS, Mathiesen JM, Sunahara RK, Pardo L, Weis WI, Kobilka BK & Granier S (2012) Crystal structure of the  $\mu$ -opioid receptor bound to a morphinan antagonist. *Nature* **485**, 321–326.
- 434 Granier S, Manglik A, Kruse AC, Kobilka TS, Thian FS, Weis WI & Kobilka BK (2012) Structure of the  $\delta$ -opioid receptor bound to naltrindole. *Nature* **485**, 400–404.
- 435 Wu H, Wacker D, Mileni M, Katritch V, Han GW, Vardy E, Liu W, Thompson AA, Huang X-P, Carroll FI, Mascarella SW, Westkaemper RB, Mosier PD, Roth BL, Cherezov V & Stevens RC (2012) Structure of the human  $\kappa$ -opioid receptor in complex with JDTic. *Nature* **485**, 327–332.
- 436 Thompson AA, Liu W, Chun E, Katritch V, Wu H, Vardy E, Huang X-P, Trapella C, Guerrini R, Calo G, Roth BL, Cherezov V & Stevens RC (2012) Structure of the nociceptin/orphanin FQ receptor in complex with a peptide mimetic. *Nature* **485**, 395–399.
- 437 Murphy PM, Ozçelik T, Kenney RT, Tiffany HL, McDermott D & Francke U (1992) A structural homologue of the N-formyl peptide receptor. Characterization and chromosome mapping of a peptide chemoattractant receptor family. *J. Biol. Chem.* **267**, 7637–7643.
- 438 Ye RD, Cavanagh SL, Quehenberger O, Prossnitz ER & Cochrane CG (1992) Isolation of a cDNA that encodes a novel granulocyte N-formyl peptide receptor. *Biochem. Biophys. Res. Commun.* **184**, 582–589.
- 439 Quehenberger O, Prossnitz ER, Cavanagh SL, Cochrane CG & Ye RD (1993) Multiple domains of the N-formyl peptide receptor are required for high-affinity ligand binding. Construction and analysis of chimeric N-formyl peptide receptors. *J. Biol. Chem.* **268**, 18167–18175.
- 440 Kufareva I, Rueda M, Katritch V, Stevens RC & Abagyan R (2011) Status of GPCR modeling and docking as reflected by community-wide GPCR Dock 2010 assessment. *Structure* **19**, 1108–1126.
- 441 Boulay F, Tardif M, Brouchon L & Vignais P (1990) The human N-formylpeptide receptor. Characterization of two cDNA isolates and evidence for a new subfamily of G-protein-coupled receptors. *Biochemistry* **29**, 11123–11133.
- 442 Lala A, Sharma A, Sojar HT, Radcliff SJ, Genco RJ & De Nardin E (1993) Recombinant expression and partial characterization of the human formyl peptide receptor. *Biochim. Biophys. Acta* **1178**, 302–306.
- 443 Prossnitz ER, Quehenberger O, Cochrane CG & Ye RD (1993) Signal transducing properties of the N-formyl peptide receptor expressed in undifferentiated HL60 cells. *J. Immunol.* **151**, 5704–5715.
- 444 Mills JS, Miettinen HM, Cummings D & Jesaitis AJ (2000) Characterization of the binding site on the formyl peptide receptor using three receptor mutants and analogs of Met-Leu-Phe and Met-Met-Trp-Leu-Leu. *J. Biol. Chem.* **275**, 39012–39017.
- 445 Vanni S, Neri M, Tavernelli I & Rothlisberger U (2011) Predicting novel binding modes of agonists to  $\beta$  adrenergic receptors using all-atom molecular dynamics simulations. *PLoS Comput. Biol.* **7**, e1001053.
- 446 Haga K, Kruse AC, Asada H, Yurugi-Kobayashi T, Shiroishi M, Zhang C, Weis WI, Okada T, Kobilka BK, Haga T & Kobayashi T (2012) Structure of the human M2 muscarinic acetylcholine receptor bound to an antagonist. *Nature* **482**, 547–551.
- 447 Steiner H, Winkler E & Haass C (2008) Chemical cross-linking provides a model of the gamma-secretase complex subunit architecture and evidence for close proximity of the C-terminal fragment of presenilin with APH-1. *J. Biol. Chem.* **283**, 34677–34686.
- 448 Fluhrer R, Grammer G, Israel L, Condron MM, Haffner C, Friedmann E, Böhlend C, Imhof A, Martoglio B, Teplow DB & Haass C (2006) A gamma-secretase-like intramembrane cleavage of TNF $\alpha$  by the GxGD aspartyl protease SPPL2b. *Nat. Cell Biol.* **8**, 894–896.

- 449 Jensen LJ, Kuhn M, Stark M, Chaffron S, Creevey C, Muller J, Doerks T, Julien P, Roth A & Simonovic M (2009) STRING 8—a global view on proteins and their functional interactions in 630 organisms. *Nucleic acids research* **37**, D412–D416.
- 450 Biosciences Y (2010) YASARA: Yet another scientific artificial reality application. .

## 12 Appendix

### List of crystallization screens used in the CrystalMation automated robotic parameter screening of CF expressed Pen-2

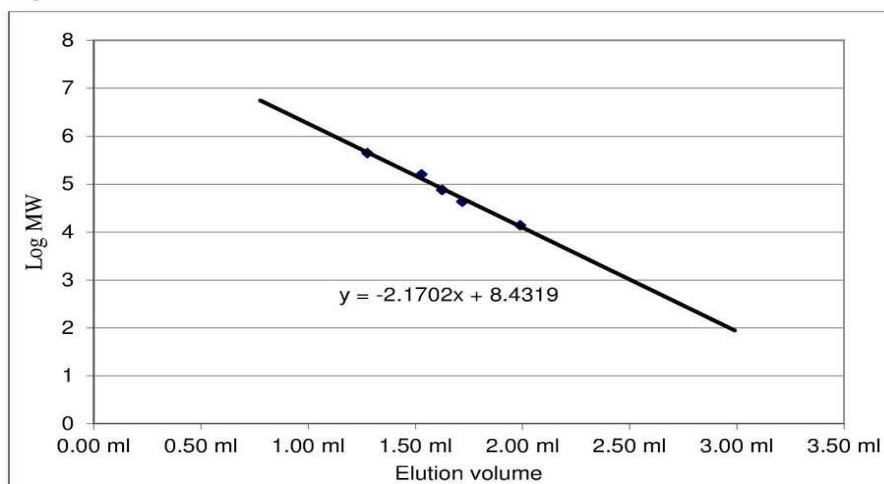
- The CubicPhase I Suite (Qiagen GmbH ,Germany), 22 °C
- The CubicPhase II Suite (Qiagen GmbH ,Germany), 22 °C
- The MbClass Suite (Qiagen GmbH ,Germany), 4 °C
- The MbClass II Suite (Qiagen GmbH ,Germany), 4 °C
- MemGold™ (Molecular Dimensions Ltd., UK), 4 and 18 °C
- MemStart and MemPlus™ (Molecular Dimensions Ltd., UK), 4 and 18 °C
- MemSys™ and Sigma Membrane Kit (Molecular Dimensions Ltd., UK and Sigma-Aldrich Chemie GmbH, Germany), 4 and 18 °C
- PGA Screen™ (Molecular Dimensions Ltd., UK), 4 °C
- JBScreen Pentaerythritol HTS (Jena Bioscience GmbH, Germany), 4 °C

### Sequencing primers for pET-Vectors

pET - T7-promoter: 5' TAA TAC GAC TCA CTA TAG GG 3'

pET - T7-terminator: 5' CTA GTT ATT GCT CAG CGG T 3'

Superdex 200 3.2/30



Superdex 200 10/300

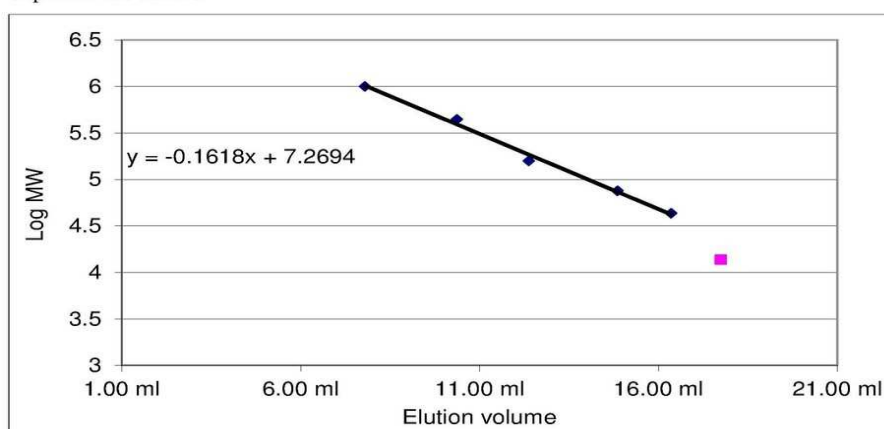


Fig 12.1 Calibration curves of the size exclusion chromatography columns (linear relationship between  $\log MW$  of the protein and elution volume)

## 13 Acknowledgements

First of all I would like to convey my sincere regards to Prof. Dr. **Volker Dötsch**, Prof. Dr. **Slawomir Filipek** and Dr. **Frank Bernhard** for giving me the opportunity for PhD studies and continuous supervision, support and encouragement throughout the project.

In course of the PhD studies, I have received numerous support from many scientists, coworkers and friends. Especially I would like to thank Prof. Harald Steiner for continuous guidance on gamma secretase and performing activity assay of the complex, Dr. Jan Hoffmann for mass spectrometry, Dr. Thomas C. Marlovits for electron microscopy, Dr. Frank Löhr for NMR spectroscopy, Dr. Dorota Latek, Krzysztof Mlynarczyk, Dr. Bartosz Trzaskowski, Wojciech Pulawski for help with modeling, molecular dynamics simulations and constructive discussions.

Also I would like to take the opportunity to say sincere thanks to Yi Ma, Christian Roos, Lei Kai for help with cell-free expression, Dr. Sebastian Richers, Dr. Ratnesh Kumar Srivastav, Dr. P. Shaik Syed Ali, Susanne Stefer, Alexis Rozenknop, Stefan Haberstock, Vladimir Rogov, Birgit Schäfer for fruitful discussions. And Shuguang Yuan, Szymon Niewieczerzal, Rongliang Wu, Michal Kolinski for discussions on modeling and simulations, Paweł Pasznik and Manfred Strupf for IT consultations. Many thanks to Dr. David Popp, Dr. Akihiro Narita and Dr. Robert C. Robinson for initial training, continuous support and collaboration. I would not forget the contribution of Sigrid Oğuzer-Fachinger, Magdalena Powierża, Monica Nowicza, Beata Thakz for taking care of administrative issues. I would like to apologize to those whose names were omitted due to space constraints. Special thanks to Ivan Corbeski for help with the German translation.

I have no words to thank my parents and brother for supporting me since childhood. Also I am indebted to all my teachers. Finally I would like to say 'thank you all' to my family, friends and patrons.

**Funding:** The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 211800. I am also thankful to FEBS (Federation of European Biochemical Societies) for 3 months Short Term Fellowship.

**Hardware:** I would like to acknowledge CSC-FUCHS at Goethe University, Frankfurt and the supercomputer at Technical University Darmstadt, Darmstadt for providing simulation time.

---

**Eidesstattliche Versicherung**

Ich erkläre hiermit an Eides Statt, dass ich die vorgelegte Dissertation über “Cell-free expression and molecular modeling of the  $\gamma$ -secretase complex and G-protein-coupled receptors” selbständig angefertigt und mich anderer Hilfsmittel als der in ihr angegebenen nicht bedient habe, insbesondere, dass alle Entlehnungen aus anderen Schriften mit Angabe der betreffenden Schrift gekennzeichnet sind.

Ich versichere, nicht die Hilfe einer kommerziellen Promotionsvermittlung in Anspruch genommen zu haben.

Frankfurt am Main, den

Umesh GHOSHDASTIDER

# Curriculum Vitae of Mr. Umesh GHOSHDASTIDER

## PERSONAL DATA

---

**Born:** 15<sup>th</sup> Sep, 1985 at Kolkata, India **Citizenship:** Indian **Marital Status:** Single  
**Address:** Max-von-Laue Str. 9, Biocenter N230/1OG, Frankfurt 60438, Germany  
**Mobile:** +49 15734 919755 **Email:** ugd1985@gmail.com



## DEGREES

---

**2008 Jul:** 2<sup>nd</sup> best in college in 2-year **M.Sc. in Human Physiology**, *University of Calcutta* (ranked 3<sup>rd</sup> in India by India Today)

**2008 Jan:** 1-year of **M.Sc in Computer Application**, *Indira Gandhi National Open University*, New Delhi

**2006 May:** 3-year **B.Sc. in Physiology (Honours)** with Physics and Chemistry, *University of Calcutta*

**2003 Apr:** 2<sup>nd</sup> best in school in final **Matura Exam**, *West Bengal Council of Higher Secondary Education*

**2001 Mar:** 2<sup>nd</sup> best in school in final **Secondary Exam**, *West Bengal Board of Secondary Education*

## RESEARCH EXPERIENCE

---

### 2009 Aug - 2012 Nov: PhD Studies

Prof. **Volker Dötsch** and Dr. **Frank Bernhard**, *Institute of Biophysical Chemistry, Centre for Biomolecular Magnetic Resonance, Goethe University*, Frankfurt am Main, Germany

Cell-free expression and structural studies of  $\gamma$ -Secretase complex involved in Alzheimer's disease.

Prof. **Slawomir Filipek**, *Laboratory of Biomodeling, International Institute of Molecular and Cell Biology*, Warsaw, Poland and *Faculty of Chemistry, University of Warsaw*, Poland

Molecular modeling of the components of  $\gamma$ -Secretase complex and the rhodopsin, cannabinoid, opioid, formyl peptide GPCRs

### 2008 Nov - 2009 Apr: Postgraduate Research Fellow

Dr. **David Popp**, *RIKEN Harima Institute at Spring-8*, Hyogo, Japan and Dr. **Akihiro Narita**, *Nagoya University Graduate School of Biological Sciences*, Nagoya, Japan

Structural biology of the bacterial cytoskeleton and DNA segregation proteins AlfA, SopA, ParM, MreB etc.

### 2008 Sep - 2008 Oct: Research Assistant

Dr. **Annalisa Perna**, *Dept. of Biostatistics* and Dr. **Marina Noris**, *Dept. of Molecular Medicine, Mario Negri Institute for Pharmacological Research*, Bergamo, Italy

Analysis of Genetic Linkage of Type-2 Diabetes and Clinical Trials Data

### 2006 Aug - 2007 Dec: Research Internship

Dr. **Banani Saha**, *Dept. of Computer Science and Engineering* and Dr. **Subhashis Mukherjee**, *Dept. of Biophysics, Molecular Biology and Genetics, University of Calcutta*

DNA and Protein Sequence Analysis Algorithms and Protein Structure Prediction

## EXPERIMENTAL SKILLS

---

PCR, Protein Expression (cell-free and *in vivo*) and Purification, SDS-PAGE, Western Blotting, Size-exclusion Chromatography, NMR, X-ray Crystallography, Electron Microscopy, Polymerization and Enzyme Kinetics, Basic Biochemistry and Physiology techniques.

## COMPUTATIONAL SKILLS

---

**Computational Chemistry:** protein modeling (Modeller, Rosetta, HHPred, Yasara, ProQ), visualization (PyMol, VMD, UCSF Chimera, Jmol), molecular dynamics (Gromacs 4.5.5, CHARMM, NAMD, Desmond), Ligand docking (AutoDock, DOCK, Ligplot), Schrödinger Suite 2012

**Bioinformatics:** ClustalW, Blast, JalView etc

**Programming:** Bash Shell Scripting, C, **Python** (Numpy, Scipy, Matplotlib, Flask (web), Biopython, Statistics: Pandas, Mlpy, RPy, scikits etc), R (Basic), L<sup>A</sup>T<sub>E</sub>X

## GRANTS

---

2012 Sep: FEBS (*Federation of European Biochemical Societies*) Short Term Fellowship (3 months)

2009 Aug: Fellowship from *Marie Curie Initial Training Network* (MC-ITN) of EU's 7<sup>th</sup> Framework Programme (FP7) (3 years)

2008 Nov: Research Stipend, *Japanese Science and Technology Agency* (JST) (6 months)

**Language Proficiency:** English (Excellent), German (Beginner), Bengali (Native Speaker), Hindi (Very Good), Sanskrit (Basic)

**Hobbies:** Yoga, Reading, Listening to Music, Poetry, Playing Chess, Traveling etc.

**Professional Memberships:** Marie Curie Fellows Association (MCFA), Polish Biochemical Society, Biophysical Society

## RESEARCH ACHIEVEMENTS

---

- Cell-free expression, purification, structural and functional characterization of the  $\gamma$ -Secretase operon and its subunits Pen-2, Presenilin 1 and Aph-1 involved in Alzheimer's disease
- Molecular model of  $\gamma$ -Secretase component interactions with substrate amyloid precursor protein
- Molecular model of GPCR (cannabinoid and formyl peptide receptors) activation mechanisms
- Molecular model of binding between the let-7g microRNA precursor and the human Lin28B protein involved in cell differentiation and cancer
- Defined polymeric structures of bacterial homologs of actin and tubulin like ParM, MreB, Alfa, Alp12 etc. required for cell division, DNA and plasmid segregation using electron and light microscopy, light scattering and by modeling of polymerization kinetics

## CONFERENCE AND WORKSHOP POSTERS AND TALKS

---

1. "2nd Annual Meeting of Structural Biology of Membrane Proteins (SBMPs) Marie Curie Initial Training Network", June 8-11, 2010, Biozentrum, University of Basel, Basel, Switzerland (Talk and Poster)
2. (Invited) EMBO Practical Course "Computational aspects of protein structure determination and analysis: from data to structure to function", Sep 6-10, 2010, European Bioinformatics Institute (EMBL-EBI), Hinxton, Cambridge, UK (Poster, Full expenses grant)
3. (Invited) Marie Curie TranSys ITN International conference "Membrane protein structure, biogenesis and bioinformatics", Oct 25-29, 2010, Maratea, Italy (Poster, Full expenses grant)
4. "3rd Annual Meeting of SBMPs", May 23-25, 2011, Utrecht, Netherlands (Talk and Poster)
5. "The Relationship of Numerical Simulations and Experimental Methods", May 26-28, 2011, Warsaw, Poland (Poster)
6. (Invited) "Expression, Purification, Crystallization and Structure of Membrane Proteins", Bilbao Advanced Courses on Biophysics, Sep 12-18, 2011, Bilbao, Spain (Poster, Full expenses grant)
7. (Invited) "Bacterial Cytoskeleton to  $\gamma$ -Secretase", Dec 20, 2011, Rammohan College, Kolkata, India (Talk)
8. "Final Annual Meeting of Structural Biology of Membrane Proteins (SBMPs) Marie Curie Initial Training Network", May 30-June 2, 2012, Maratea, Italy (Poster)

## PEER-REVIEWED PUBLICATIONS (h-index = 6, citations > 110)

---

1. Roos C, Kai L, Haberstock S, Proverbio D, Ghoshdastider U, Ma Y, Filipek S, Wang X, Dötsch V, Bernhard F. *High Level Cell-Free Production of Membrane Proteins into Nanodiscs*. **Methods Mol Biol** In Press
2. Yuan S, Ghoshdastider U, Trzaskowski B, Latek D, Debinski A, Pulawski W, Wu R, Gerke V, Filipek S. *The role of water in activation mechanism of human N-formyl Peptide Receptor 1 (FPR1) based on molecular dynamics simulations*. **Plos One** In Press IF 4.537
3. Shaik Shaik Ali P, Ghoshdastider U, Hoffmann J, Brutschy B, Filipek S (2012) *Recognition of the human Lin28B by let-7g microRNA precursor*. **FEBS Lett** 586(22):3986-90 IF 3.33 #cit 0
4. Ma Y<sup>§</sup>, Ghoshdastider U<sup>§</sup>, Wang J, Ye W, Dötsch V, Filipek S, Bernhard F, Wang X (2012) *Cell-free expression of human Glucosamine 6-phosphate N-acetyltransferase (HsGNA1) for inhibitor screening*. **Protein Express Purif** 86(2):120-126 IF 1.654 #cit 0 § **Co-first Author**
5. Roos C, Kai L, Proverbio D, Ghoshdastider U, Filipek S, Dötsch V, Bernhard F (2012) *Co-translational association of cell-free expressed membrane proteins with supplied lipid bilayers*. **Mol Membr Biol** Jun 20 (Epub ahead of print) #cit 1
6. Popp D, Narita A, Lee LJ, Ghoshdastider U, Xue B, Srinivasan R, Balasubramanian M, Tanaka T, Robinson RC (2012) *A novel actin-like filament structure from Clostridium tetani*. **J Biol Chem** 287(25):21121-29 IF 5.117, #cit 0
7. Pulawski W, Ghoshdastider U, Andrisano V, Filipek S (2012) *Ubiquitous Amyloids*. **J Appl Biochem Biotech** 166(7):1626-43 IF 1.998, #cit 0
8. Trzaskowski B, Latek D, Yuan S, Ghoshdastider U, Debinski A, Filipek S (2012) *Action of molecular switches in GPCRs - theoretical and experimental studies*. **Curr Med Chem** 19(8):1090-109 IF 4.961, #cit 1
9. Ma Y, Muench D, Schneider T, Sahl HG, Bouhss A, Ghoshdastider U, Wang J, Dötsch V, Wang X, Bernhard F (2011) *Preparative scale cell-free production and quality optimization of MraY homologues in different expression modes*. **J Biol Chem** 286(45):38844-53 IF 5.117, #cit 6
10. Latek D, Kolinski M, Ghoshdastider U, Debinski A, Bombolewski R, Plazinska A, Jozwiak K, Filipek S (2011) *Modeling of ligand binding to G protein coupled receptors: cannabinoid CB1, CB2 and adrenergic beta-2 AR*. **J Mol Model** 17(9):2353-66 IF 2.242, #cit 3
11. Kufareva I, Rueda M, Katritch V, participants of GPCR Dock 2010 (including Ghoshdastider U), Stevens RC, Abagyan R (2011) *Status of GPCR modeling and docking as reflected by community-wide GPCR Dock 2010 assessment*. **Structure** 19(8):1108-26 IF 5.931, #cit 42

12. \*Sobhanifar S, Schneider B, Löhr F, Gottstein D, Ikeya T, Mlynarczyk K, Pulawski W, Ghoshdastider U, Kolinski M, Filipek S, Güntert P, Bernhard F, Dötsch V (2010) *Structural investigation of the C-terminal catalytic fragment of presenilin 1*. **Proc. Natl. Acad. Sci. USA** 107(21):9644-9 IF 10.472, #cit 22
  13. Popp D, Narita A, Maeda K, Fujisawa T, Ghoshdastider U, Iwasa M, Maéda Y, Robinson RC (2010) *Filament structure, organization, and dynamics in MreB sheets*. **J Biol Chem** 285(21):15858-65 IF 5.117, #cit 17
  14. Popp D, Narita A, Ghoshdastider U, Maeda K, Maéda Y, Oda T, Fujisawa T, Onishi H, Ito K, Robinson RC (2010) *Polymeric structures and dynamic properties of the bacterial actin Alfa*. **J Mol Biol** 397(4):1031-41 IF 3.981, #cit 8
  15. \*Popp D, Xu W, Narita A, Brzoska AJ, Skurray RA, Firth N, Ghoshdastider U, Maéda Y, Robinson RC, Schumacher MA (2010) *Structure and filament dynamics of the pSK41 actin-like ParM protein: implications for plasmid DNA segregation*. **J Biol Chem** 285(13):10130-40 IF 5.117, #cit 11
- \***Recommended by Faculty of 1000**, IF = Journal Citation Reports 5-year Impact Factor

#### CONFERENCE PROCEEDINGS

---

1. Ghoshdastider U, Mlynarczyk K, Pulawski W, Kolinski M, Filipek S (2010) *Molecular Dynamics Simulations of NMR structure of C-Terminal Fragment of Presenilin 1*. Front Neurosci Conference Abstract: Neuroinformatics 2010 doi: 10.3389/conf.fnins.2010.13.00124
2. Ghoshdastider U, Saha B (2007) *GenomeCompress-A novel algorithm for DNA compression*. Proceedings of International Conference on Information Technology, Haldia, WB, India (ISSN 0973-6824) #cit 1

#### SUBMITTED MANUSCRIPTS

---

1. Ghoshdastider U, Xu R, Trzaskowski B, Mlynarczyk K, Miszta P, Viswanathan S, Renugopalakrishnan V, Filipek S (2012) *Modeling of encapsulation of glucose oxidase with graphene flakes*.

#### DRAFT MANUSCRIPTS

---

1. Ghoshdastider U, Latek D, Dötsch V, Bernhard F, Filipek S (2012) *Cell-free Expression, NMR and computational modeling and validation of Pen-2*
2. Ghoshdastider U, Dötsch V, Steiner H, Bernhard F, Elofsson A, Filipek S (2012) *Prediction of the effects of presenilin mutations in Alzheimer's disease by machine learning*
3. Ghoshdastider U, Dötsch V, Steiner H, Bernhard F, Filipek S (2012) *Cell-free synthesis of the  $\gamma$ -secretase operon*