Inaugural dissertation

for

obtaining the doctoral degree

of the

Combined Faculty of Mathematics, Engineering and Natural Sciences

of the

Ruprecht - Karls - University

Heidelberg

Presented by

Diplomate Engineer Christina Athanasiou

born in: Athens, Greece

Oral examination: 30 October 2023

Modeling of mechanistic effects of neurotrophin modulators

Referees: Prof. Dr. Rebecca C. Wade

Prof. Dr. Ursula Kummer

«Σα βγεις στον πηγαιμό για την Ιθάκη,
να εύχεσαι να ’ναι μακρύς ο δρόμος,
γεμάτος περιπέτειες, γεμάτος γνώσεις….
Κι αν πτωχική την βρεις, η Ιθάκη δεν σε γέλασε.
Έτσι σοφός που έγινες, με τόση πείρα,
ήδη θα το κατάλαβες οι Ιθάκες τι σημαίνουν.»

ΙΘΑΚΗ, Κωνσταντίνος Καβάφης

“When you set out on your journey to Ithaca,
pray that the road is long,
full of adventure, full of knowledge…..
And if you find her poor, Ithaca has not deceived you.
Wise as you have become, with so much experience,
you must already have understood what Ithaca means.”

ITHACA, Constantine Cavafy

Contents

Abbreviations 3

Summary 7

Zusammenfassung 9

1 Introduction 11

1.1 Neurotrophins and their receptors 11

1.2 Structure and function of neurotrophins and receptors 13

1.3 Neurotrophin signaling 35

1.4 Neurotrophins in neurodegeneration 38

1.5 Neurotrophin mimetics 42

1.6 Computational approaches to study mechanistic effects of neurotrophins and their mimetics 43

1.7 Scope of the thesis 45

2 Theoretical introduction of computational methods 46

2.1 Molecular modeling of biomolecules 46

2.2 Molecular mechanics force fields 46

2.3 Energy minimization 50

2.4 Molecular dynamics simulations 51

2.5 Enhanced sampling methods 53

2.6 Homology modeling 54

2.7 Molecular docking 55

3 Glycosylation effects on the extracellular receptor dynamics 57

3.1 Introduction to the glycosylation of the receptors 57

3.2 Methodology 57

3.3 Results 65
3.4 Conclusions 92
4 Transmembrane domain dimerization mechanism 94
4.1 Introduction to transmembrane helix function 94
4.2 Methodology 95
4.3 Results 103
4.4 Conclusions 123
5 Dynamic behavior of full-length receptors in the neuronal membrane 125
5.1 Introduction to the neuronal membrane composition and its importance for the receptors 125
5.2 Methodology 125
5.3 Results 135
5.4 Conclusions 150
6 Investigation of neurotrophin mimetics' mechanism of action 152
6.1 Introduction to the study of neurotrophin mimetics and its challenges 152
6.2 Methodology 153
6.3 Results 158
6.4 Conclusions 164
7 Final conclusions and future perspectives 166
8 Acknowledgments 172
9 Appendix 174
10 List of Tables 220
11 List of Figures 222
12 References 231

Abbreviations

AA	All-atom
AD	Alzheimer's disease
AKT	Ak strain transforming
ALS	Amyotrophic lateral sclerosis
APBS	Adaptive Poisson-Boltzmann Solver
ARMS	Ankyrin repeat-rich membrane spanning
ASA	Accessible surface area
ATP	Adenosine triphosphate
BB	Backbone
BBB	Blood brain barrier
BDNF	Brain-derived neurotrophic factor
BFCN	Basal forebrain cholinergic neurons
BLAST	Basic Local Alignment Search Tool
BP1	Binding protein 1
CAP	c-Cbl associated protein
CARC	Cholesterol-recognition amino acid consensus
CARV	Carvacrol
CER	Ceramide
CG	Coarse-grained
CNS	Central nervous system
CNTF	Ciliary Neurotrophic Factor
COG	Center of geometry
COM	Center of mass
CRD	Cysteine-rich domain
CREB	cAMP response element binding protein
CV	Collective variable
DAG	Diacylglycerol
DFT	Density functional theory
DHEA	Dehydroepiandrosterone
DHF	Dihydroxyflavone
DPC	Dodecylphosphocholine
DPPC	1,2-dipalmitoyl-sn-glycero-3-phosphocholine
DPPE	1,2-dipalmitoyl-sn-glycero-3-phosphoethanolamine
DPR	Detergent-to-protein molar ratio
DRG	Dorsal root ganglion
EC	Extracellular
ECD	Extracellular domain
ECL	Extracellular linker
EGFR	Epidermal Growth Factor Receptor
ENT	EuroNeurotrophin
ERK	Extracellular signal–regulated kinase
ESP	Electrostatic potential
FES	Free energy surface

FGF Fibroblast growth factor
FKHRL1 Forkhead transcription factor like 1
GDNF Glial - Derived Neurotrophic Factor
GSK Glycogen synthase kinase
HD Huntington's disease
HITS Heidelberg Institute for Theoretical Studies
HPC High Performance Computing
IC Intracellular
ICD Intracellular domain
ICL Intracellular linker
ICV Intracerebroventricular
IFD Induced Fit Docking
IRS1,2 Insulin receptor substrates 1,2
ITN Innovative Training Network
JM Juxtamembrane
KDE Kernel density estimate
LLPE 2-linoleoyl-sn- glycero-3-phosphoethanolamine
LTP Long-Term Potentiation
MAPK Mitogen-activated protein kinase
MC Monte Carlo
MCM Molecular and cellular modeling
MD Molecular dynamics
NGF Nerve Growth Factor
NGFR Nerve Growth Factor Receptor
NMR Nuclear Magnetic Resonance
NPW Non-polarizable water
NSM Neutral sphingomyelin
NT Neurotrophin
PAPC 1-palmitoyl-2-arachidonoyl-sn- glycero-3-phosphocholine
PBK Protein kinase B
PCA Principal Component Analysis
PCSK1,2,3 Proprotein convertase subtilisin/kexin type 1,2,3
PD Parkinson's disease
PDB Protein databank
PDK1 3- phosphor-inositide-dependent kinase-1
PI Phosphatidylinositol
PI3K Phosphoinositide 3-kinase
PIP2 Phosphatidylinositol 4,5-biphosphate
PIP3 P3-phosphorylated phospho-inositide
PKA,B,C Protein kinase A,B,C
PLC Phospholipase
PLPS 1-palmitoyl-2-linoleoyl-sn-glycero-3-phospho-L-serine
PME Particle Mesh Ewald
PNS Peripheral nervous system
POPC 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine
POPE 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoethanolamine
POPI 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoinositol

POPI25 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoinositol 4,5-bisphosphate
POPS 1-palmitoyl-2-oleoyl-sn-glycero-3-phospho-L-serine
PS Phosphoserine
RIP Regulated intramembrane proteolysis
RMSD Root mean squared deviation
RMSF Root mean squared fluctuation
RSK Ribosomal S6 kinase
RTK Receptor tyrosine kinase
SAPI 1-stearoyl-2-arachidonoyl-sn-glycero-3-phosphoinositol
SASA Solvent accessible surface area
SM Sphingomyelin
SNFG Symbol Nomenclature for Glycans
SOS Son of sevenless protein
SP Standard precision
SSM N-stearoyl sphingomyelin
STD Saturation Transfer Difference
TM Transmembrane
TMD Transmembrane domain
TNFR Tumor necrosis factor receptor
TOR Target of rapamycin
TRPC Transient receptor potential channel
WT Wild-type

Summary

Neurotrophins (NTs) are growth factors that are expressed in the central and peripheral nervous systems. They are implicated in different phases of the development and maintenance of the nervous system and they can regulate neuronal survival, development, function, plasticity, as well as neuronal apoptosis. NTs can be used as therapeutics for the treatment of neurodegenerative disorders. However, their poor pharmacokinetic properties and their invasive administration to patients renders them inefficient for use as pharmaceuticals. A solution to this can be offered by small molecule NT mimetics, which can elicit NT mechanisms through binding to NT receptors, which are transmembrane (TM) glycoproteins. The mechanism of activation of NT receptors remains elusive and thus, in this thesis, I have investigated the mechanism of action of NT receptors and mimetics through molecular modeling and molecular dynamics (MD) simulations. I modeled and simulated the glycosylated state of the full extracellular (EC) domains of Tropomyosin receptor kinases A and B (TrkA, TrkB) NT receptors, which revealed that the glycans can shield the accessible surface area of the receptors and participate in the contact area between receptor and NT. Most importantly, glycosylation promoted the extended conformations of the EC domains, which might facilitate NT binding. Then, I performed coarse-grained MD simulations to study the possible arrangements of the TM helical homodimers of TrkA and TrkB receptors in micelles. The results revealed arrangements that could correspond to the active state of the receptors, while metadynamics simulations indicated a stronger binding for the TrkA helices by 10 kJ/mol compared to TrkB. Next, I modeled the full-length structures of the TrkA and TrkB receptors in their homodimeric, glycosylated state bound to their NTs. I embedded the receptors in a realistic model of a neuronal asymmetric membrane. I verified the proper behavior of the membrane and proteins with smaller systems comprising of the TM and intracellular monomers of the receptors. These test simulations revealed interactions between positively charged residues of the kinase domain of TrkA with negatively charged lipids of the inner leaflet of the membrane. These interactions were also formed in the full-length system, and they might stabilize the two kinase domains of the receptor dimer in an orientation that promotes activation, even though the kinase domains were not activated during the simulations. Also, in the full-length systems, the EC domains approached and lay down on the neuronal membrane, while interacting with membrane lipids, such as gangliosides, which are able to activate the NT receptors. Finally, I investigated the binding of small-molecule NT mimetics to the EC and TM domains of TrkA and TrkB receptors, with molecular docking and MD simulations. While plausible poses were

obtained, they were not able to explain the selectivity of the compounds for the receptors and the simulations showed that binding was weak. Due to the cholesterol core of the NT mimetics, I tested the ability of the compounds to enter the cell membrane with MD simulations. The compounds were able to spontaneously penetrate the membranes, indicating that their binding site could also lie in the TM region of the receptors. However, simulations with the compounds bound or close to the TM helices in the membrane environment, showed no specific binding. Further experimental exploration of the binding mechanism of these compounds is required. Overall, this thesis sheds light on the dynamic behavior of TrkA and TrkB NT receptors in membranes and the mechanistic insights provide a basis for future studies to develop NT mimetics.

Zusammenfassung

Neurotrophine (NTs) sind Wachstumsfaktoren, die im zentralen und peripheren Nervensystem vorkommen. Sie sind an verschiedenen Phasen der Entwicklung und Aufrechterhaltung des Nervensystems beteiligt und können das Überleben, die Entwicklung, die Funktion und die Plastizität von Neuronen sowie die neuronale Apoptose regulieren. NTs können als Therapeutika für die Behandlung von neurodegenerativen Erkrankungen eingesetzt werden. Aufgrund ihrer schlechten pharmakokinetischen Eigenschaften und ihrer invasiven Verabreichung sind sie jedoch für den Einsatz als Arzneimittel ineffizient. Eine Lösung für dieses Problem bieten niedermolekulare NT-Mimetika, die durch Bindung an NT-Rezeptoren, bei denen es sich um Transmembran-Glykoproteine (TM) handelt, NT-Mechanismen auslösen können. Der Aktivierungsmechanismus von NT-Rezeptoren ist noch nicht vollständig verstanden. Daher habe ich in dieser Arbeit den Wirkmechanismus von NT-Rezeptoren und -Mimetika durch molekulare Modellierung und Molekulardynamiksimulationen (MD) untersucht. Ich habe die vollständigen extrazellulären Domänen der NT-Rezeptoren der Tropomyosin-Rezeptorkinasen A und B (TrkA, TrkB) im glykosylierten Zustand modelliert und simuliert. Dabei zeigte sich, dass die Glykane die zugängliche Oberfläche der Rezeptoren abschirmen können und an der Kontaktfläche zwischen Rezeptor und NT beteiligt sind. Hauptsächlich aber förderte die Glykosylierung die „extended“ Konformationen der EC-Domänen, was die NT-Bindung erleichtern könnte. Anschließend benutzte ich grobkörnige MD-Simulationen, um die möglichen Anordnungen der helikalen TM-Homodimere von TrkA- und TrkB-Rezeptoren in Mizellen zu untersuchen. Die Ergebnisse zeigten Anordnungen, die dem aktiven Zustand der Rezeptoren entsprechen könnten, während Metadynamics Simulationen eine um 10 kJ/mol stärkere Assoziation der TrkA-Helices im Vergleich zu TrkB ergaben. Als nächstes modellierte ich die vollständigen Strukturen der TrkA- und TrkB-Rezeptoren in ihrem homodimeren, glykosylierten Zustand, gebunden an ihre NTs. Ich bettete die Rezeptoren in ein realistisches Modell einer neuronalen asymmetrischen Membran ein. Ich überprüfte das richtige Verhalten der Membran und der Proteine mit kleineren Systemen, die aus dem TM und intrazellulären Monomeren der Rezeptoren bestanden. Diese Testsimulationen zeigten Wechselwirkungen zwischen positiv geladenen Resten der Kinasedomäne von TrkA und negativ geladenen Lipiden der inneren Membranseite. Diese Wechselwirkungen traten auch im vollständigen System auf und könnten die beiden Kinasedomänen des Rezeptordimers in einer Orientierung stabilisieren, die die Aktivierung fördert, auch wenn die Kinasedomänen während der Simulationen nicht aktiviert wurden. Außerdem näherten sich die EC-Domänen in den vollständigen

Systemen der neuronalen Membran und legten sich auf diese, während sie mit Membranlipiden wie Gangliosiden interagierten, die die NT-Rezeptoren aktivieren können. Schließlich untersuchte ich die Bindung von niedermolekularen NT-Mimetika an die EC- und TM-Domänen von TrkA- und TrkB-Rezeptoren mit molekulare Docking und MD-Simulationen. Obwohl plausible Posen gefunden wurden, konnten sie die Selektivität der Moleküle für die Rezeptoren nicht erklären, und die Simulationen zeigten, dass die Bindung schwach war. Aufgrund der Cholesteringrundeinheit der NT-Mimetika habe ich die Fähigkeit der Verbindungen, in die Zellmembran einzudringen, mit MD-Simulationen getestet. Sie waren in der Lage, die Membranen spontan durchzudringen, was darauf hindeutet, dass ihre Bindungsstelle auch in der TM-Region der Rezeptoren liegen könnte. Simulationen, bei denen die Verbindungen an die TM-Helices in der Membran gebunden waren oder sich in deren Nähe befanden, zeigten jedoch keine spezifische Bindung. Eine weitere experimentelle Untersuchung des Bindemechanismus dieser Verbindungen ist erforderlich. Insgesamt gibt diese Arbeit Aufschluss über das dynamische Verhalten von TrkA- und TrkB-NT-Rezeptoren in Membranen, und die mechanistischen Erkenntnisse bieten eine Grundlage für künftige Studien zur Entwicklung von NT-Mimetika.

1 Introduction

1.1 Neurotrophins and their receptors

Neurotrophins (NTs) are growth factors that are expressed in the central and peripheral nervous systems (CNS, PNS) and they belong to the large family of neurotrophin factors.[1] They are implicated in different phases of the development and maintenance of the nervous system and they can regulate neuronal survival, development, function, plasticity but also neuronal apoptosis.[1–3]

Growth factors are very important in the development and maintenance of the mammalian nervous system. Most of these factors have pleiotropic effects on different systems.[2,3] Other factors exhibit more restricted activity on the nervous system and are therefore classified as "neurotrophic factors." The most intensively studied neurotrophic factors include Ciliary Neurotrophic Factor (CNTF), Glial - Derived Neurotrophic Factor (GDNF), and the neurotrophic factors belonging to the Nerve Growth Factor (NGF) family, known as "neurotrophins" (NTs).[4,5]

The neurotrophin family of molecules includes six molecules, which share a high degree of structural homology; the **nerve growth factor** (NGF), the **brain-derived neurotrophic factor** (BDNF), the **neurotrophin-3** (NT-3), the **neurotrophin-4/5** (NT-4/5), the **neurotrophin-6** (NT-6) and the **neurotrophin-7** (NT-7). Neurotrophins NGF, BDNF, NT-3, NT-4/5 are expressed in mammals and lower vertebrates, whereas NT-6 and NT-7 are only expressed in fish. Neurotrophins can bind to four different receptors: the **tropomyosin receptor kinases A, B and C** (TrkA, TrkB, TrkC), which are receptor tyrosine kinases (RTKs), and the **p75 receptor**, which is a tumor necrosis factor receptor (TNFR).

NGF was discovered in 1951 by Rita Levi-Montalcini, as a molecule that regulates the survival and maturation of developing neurons in the peripheral nervous system (PNS).[6,7] Only later did it become apparent that the NGF actions extend beyond the developmental period, the nerve cells and even beyond the nervous system.[1] NGF and its receptors continue to be produced during adult life and aging by several cell types. NGF expression can be upregulated as a response to injury and plays a role in endogenous neuroprotection.[1]

The four mammalian neurotrophins share approximately 50% amino acid sequence. They are synthesized in the rough endoplasmic reticulum as pre-proproteins. Pre-proneurotrophins have an N-terminal signal peptide, which is necessary for intracellular transport and subsequent secretion.

Cleavage of the signal peptide of pre-proneurotrophins produces immature precursors, the pro-neurotrophins, of approximately 270 residues (30–35 kDa). In the *trans*-Golgi network and in secretory vesicles, pro-neurotrophins dimerize and are proteolytically processed. Pro-neurotrophin precursors of approximately 240–260 amino acids are further cleaved by proteases, to release an N-terminal pro-domain peptide and a C-terminal mature neurotrophin. Mature neurotrophins, which are about 12–13 kDa in size, are secreted and released from the cell, as stable, non-covalent dimers (monomer length: 118 – 129 amino acids).[8]

This cleavage event may occur either within the secretory pathway or following secretion (in secretory vesicles), so that receptors may be exposed to both pro-neurotrophins and mature neurotrophins. The amount of mature neurotrophin secreted from the cell depends on the range of convertases expressed in different cell types; Furin (also known as proprotein convertase subtilisin/kexin type 3, PCSK3) is expressed in most cells, whereas proprotein convertase subtilisin/kexin type 1 (PCSK1) and proprotein convertase subtilisin/kexin type 2 (PCSK2) are mostly expressed in neurons. Pro-neurotrophins may also be cleaved in the extracellular space by the serine protease Plasmin.[9]

Trk receptors are activated only by mature neurotrophins.[10] On the other hand, p75 not only binds both mature and pro-neurotrophins, but also is more effectively activated by pro-neurotrophins (Table 1.1). The binding of pro-neurotrophins to p75 is dependent on the association of p75 with Sortilin or SorCS2, which bind a conserved motif in pro-neurotrophin pro-domains.[11,12] These activation properties of p75 and Trk receptors are indicating that a tight regulation of cleavage is important for neuronal functions and survival.[13] Importantly, in the nervous system, neurotrophin secretion increased after injury, thus indicating a key role of neurotrophins in axonal regeneration.[14]

It has been shown experimentally that neurotrophins activate Receptor Tyrosine Kinases (RTKs). The pathways regulated by tyrosine kinases within neural precursors and neurons include proliferation and survival, axonal and dendritic growth and remodeling, assembly of the cytoskeleton, membrane trafficking and fusion, synapse formation and function. Studies on the neurotrophins have shown that they are participating in the regulation of each of these functions and therefore have increased our understanding of the molecular mechanisms underlying each of them.[15]

As mentioned before, the Trk receptors are activated only by mature neurotrophins, while p75 is activated by both mature and pro-neurotrophins. Specific neurotrophins bind to specific Trk receptors with affinity at the nanomolar range ($K_d \approx 10^{-9}$ M). NGF binds to TrkA; BDNF to TrkB; NT-3 to TrkC, but also to TrkA and TrkB with lower affinity compared to NGF and BDNF; while NT-4/5 binds to TrkB (Table

1.1). All four neurotrophins can also bind to p75 receptor with a similar nanomolar affinity, while pro-neurotrophins bind to p75 with higher affinity. Initially Trk receptors were considered as the high-affinity neurotrophin receptors and p75 was known as the low-affinity nerve growth factor receptor (NGFR). However, this description is misleading because p75 and Trks independently bind to neurotrophins with a similar affinity although with different kinetics.[16] Kinetic studies of neurotrophin binding to p75 indicate that the rates of dissociation are distinct for each neurotrophin, with NGF>NT-3≫BDNF.[17] Interestingly, in the presence of p75, the affinity of the Trks for NTs increases 100 times ($K_d \approx 10^{-11}$ M).[18] Although both the Trks and p75 bind NTs with similar dissociation equilibrium constants in the nanomolar range, the kinetics of binding are different. Trks bind and release NGF slower than p75.[19]

Table 1.1: Neurotrophins as ligands for neurotrophin receptors. Ligands listed in bold have greater affinity and they are more important for receptor activation *in vivo*.

Receptor	Neurotrophin
p75	**pro-NTs**, NGF, BDNF, NT-3, NT-4/5
TrkA	**NGF**, NT-3
TrkB	**BDNF**, **NT-4/5**, NT-3
TrkC	**NT-3**

1.2 Structure and function of neurotrophins and receptors

Structure of neurotrophins

All neurotrophins form non-covalently associated head-to-head homodimers, while neurotrophin monomers of NT-3 and NT-4/5 have been also found to form heterodimers with other neurotrophins like BDNF. X-ray crystal structures of NGF have been known since the early 90s.[20] Structures of BDNF, NT-3, and NT-4 homodimers and of BDNF/NT-3 and BDNF/NT-4/5 heterodimers have also been solved.[21–24]

Neurotrophins share a highly homologous structure. NGF structure is consisting of an elongated, twisted beta sheet core, a cysteine-knot motif, a reverse turn at one end of the molecule (loop III), and

three beta-hairpin loops at the other (loops I, II, and IV), as shown in Figure 1.1. The cysteine knot[1] consists of three disulphide bonds that form a true knot of the polypeptide chain. The dimer interface, consisting of hydrophobic core residues, is highly conserved and maintains the conformation of the dimer. In contrast to the β-strands, the β-loops are highly variable. Characteristic features of NGF are the long unstructured loops (I, II, III, and IV), important for the interactions with the binding sites of its receptors.[20]

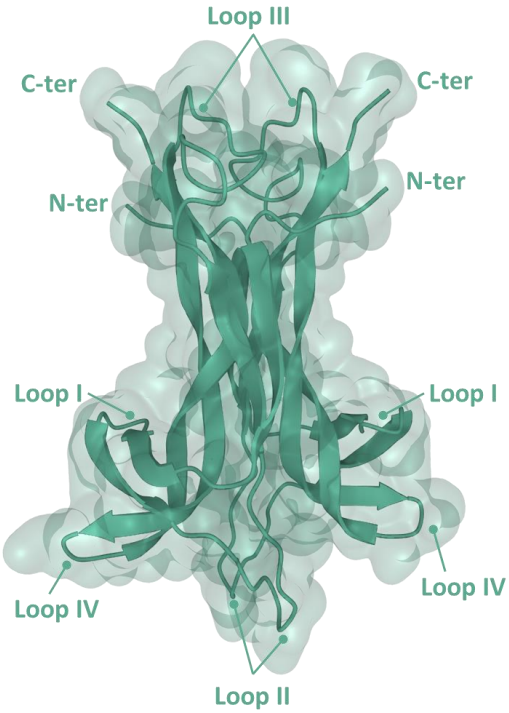


Figure 1.1: Cartoon with molecular surface model of the NGF homodimer X-ray structure (PDB ID: 1BET)[20] with the position of the loops indicated.[25]

[1] A cysteine knot is a protein structural motif containing three disulfide bridges (formed from pairs of cysteine residues). The sections of polypeptide that occur between two of them form a loop through which a third disulfide bond passes, forming a rotaxane substructure. It occurs in many proteins across many species and provides considerable structural stability. There are three types of cysteine knots, which differ in the topology of the disulfide bonds: The Growth Factor Cysteine Knot (GFCK), Inhibitor Cysteine Knot (ICK) and Cyclic Cysteine Knot, or cyclotide. The growth factor cysteine knot (GFCK) was first observed in the structure of NGF.[20] All GFCK structures that have been determined are homodimeric.

Structure of neurotrophin receptors

Neurotrophins recognize and activate two different classes of transmembrane receptor proteins: The Tropomyosin receptor kinase receptors (Trks), and the p75 neurotrophin receptor (p75). Both receptor groups exhibit independent signalling properties, which can initiate individual physiological responses dependent on the downstream signal transduction pathways that are activated upon neurotrophin binding.[26]

Trk receptors

Trks are Type I single-pass transmembrane glycoproteins, of approximately 140 kDa, and they are members of the receptor tyrosine kinase (RTK) superfamily of proteins. Trk receptors consist of three segments: an ***extracellular (EC) domain,*** a ***single transmembrane (TM) domain,*** and an ***intracellular (IC) tyrosine kinase domain*** (Figure 1.2). The extracellular domain has characteristic subdomain organization. On the N-terminus, is the so-called signal peptide composed of 32 amino acids, and this is usually cleaved. After that, there are two cysteine-rich regions (also known as D1 and D3) flanking a leucine-rich repeat (known as D2) (Figure 1.2). The C-terminus of the extracellular domain has two immunoglobulin (IgG)-like domains near the plasma membrane (also termed: Ig-C1 and Ig-C2 or D4 and D5 respectively) (Figure 1.2). Following the D5 domain, which is also the neurotrophin binding domain as will be later discussed, there is a disordered linker, which connects the extracellular segment with the transmembrane helical domain (Figure 1.2). Subsequently, the transmembrane domain is connected to the intracellular kinase domain through another disordered linker, called the juxtamembrane (JM) domain.

Studies of domain exchange between TrkA, TrkB, and TrkC identify the D5 subdomain as the ligand-binding site and functional activation domain (also termed a "***receptor hot spot***").[27] This was confirmed by crystallography.[28,29] Co-crystallization of NGF with D5 subdomain of TrkA has shown characteristic interactions with the β-sheets that form the core of the NGF homodimer and with the N-terminus of NGF.[28–30] The second leucine-rich D2 subdomain also seems to have a modulatory, perhaps indirect, role in ligand interaction.[31,32] The D1 subdomain is an allosteric receptor hot-spot which is exposed when p75 receptor is co-expressed on the cell.[33] Cleavage of the whole extracellular domain of TrkA leads to phosphorylation of the kinase domain.[34] Another study showed that deletion of the first or the second Ig-like domains (D4 and D5 in Figure 1.2) can induce activation of the TrkA receptor, with the D5 domain

deletion showing stronger activation than that of D4.[35] This suggests that these domains play an inhibitory role in receptor activation.

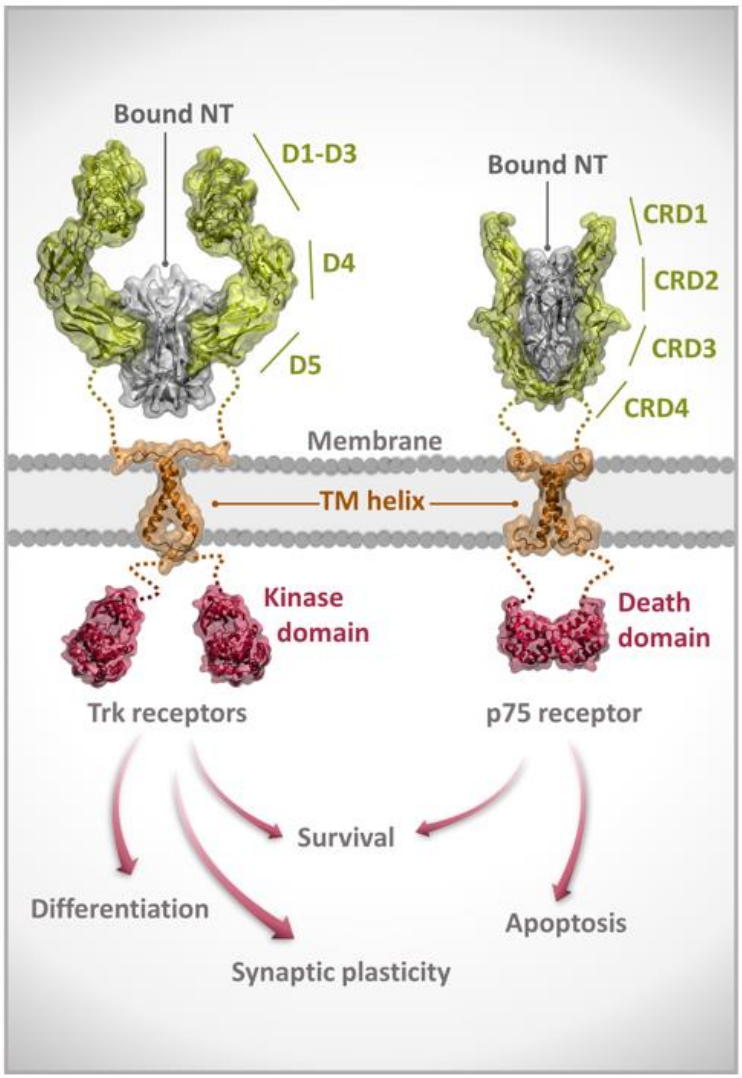


Figure 1.2: Schematic representation of the full-length Trk and p75 receptor homodimers with neurotrophins (NTs) bound to the extracellular domains. Trk receptors consist of five extracellular domains (D1–D5), an α-helical transmembrane (TM) domain, and an intracellular segment containing a kinase domain. The p75 receptor has four cysteine-rich domains (CRD1-CRD4) in the extracellular part, an α-helical transmembrane domain, as the Trks, and an intracellular death domain. Both Trk and p75 receptors can also exist in preformed inactive dimers prior to neurotrophin binding. Neurotrophin binding leads to activation of the receptors and initiation of signaling cascades related to cell survival or apoptosis, differentiation and synaptic plasticity.

Interaction of neurotrophins with their receptors is influenced by the splicing variants of Trk receptors. In fact, receptor molecules with deletions in the extracellular domains or in the intracellular kinase domain have been identified. A TrkA spliced variant with an 18 base-pairs deletion in the extracellular domain shows decreased activation by NT-3, and a TrkB molecule lacking exon 9, which comprises 33 base-pairs and encodes the juxta-membrane domain, displays decreased affinity for NT-4/5 and NT-3, while NGF binding is unaffected compared to full-length receptor.[36,37]

p75 receptor

p75 is a type I single-pass transmembrane glycoprotein receptor (with N-terminus outside of the cell), of approximately 75 kDa. It is a member of the Tumour Necrosis Factor Receptor (TNFR) superfamily of proteins. After cleavage of its 28-amino acid signal peptide, human p75 is a 399-amino acid protein. The extracellular domain contains four cysteine-rich repeated modules of 40 amino acids each, designated CRD1 to CRD4 in Figure 1.2. The cysteine-rich repeat is a distinguishing structure of the TNFR-superfamily members.[38] The extracellular domain of p75 also contains one N-glycosylated site, and O-linked glycosylated putative sites.

Unlike Trk receptors, p75 is a non-catalytic receptor[2] that is expressed when neurons are injured and its main activity leads to the death of the neurons and related cells. p75 lacks an associated domain with catalytic activity. Similar to other members of the TNFR superfamily, signalling by p75 proceeds via ligand-induced recruitment and release of cytoplasmic effectors to and from its intracellular domain.[39]

The p75 receptor frequently is co-expressed and linked with Trk receptors acting as a co-receptor for Trks. Co-expression of p75 can increase the affinity of TrkA for NGF, leading to high-affinity sites for NGF ($K_d \approx 10^{-11}$ M), and it enhances the specificity of Trks for cognate neurotrophins.[26] As a result, increased ligand selectivity can be conferred on the Trk receptors by interaction with the p75 receptor. This interaction increases Trks affinity for neurotrophins while reducing ligand induced Trk receptor internalization and degradation.[40,41]

[2] Typically, cell surface receptors can bind to specific molecules on the outside of the cell. This binding activates the receptor and then the activated receptor triggers some biochemical changes inside the cell. For many receptors, the portion of the receptor inside the cell is essentially an enzyme that can trigger a biochemical change by itself. However, some receptors lack an intracellular enzyme domain and any enzymatic activity, and it is often unclear how these 'non-catalytic receptors' trigger changes inside a cell. For the case of p75, this happens by attachment of adaptor proteins in the death domain.

Glycosylation of neurotrophin receptors

An important structural characteristic of the neurotrophin receptors is the existence of covalently attached sugar moieties, called glycans, in their extracellular domains. There are two types of glycosylation; the N-glycosylation and the O-glycosylation, which differ in the atom types to which they are attached. N-glycans are bound to asparagine residues, at the N atom, while O-glycans can be attached to serines or threonines, in the O atoms of their side-chains.[42] The minimal amino acid sequence to receive an N-glycan is Asn-X-Ser/Thr in which "X" is any amino acid except Pro. However, not all Asn residues in this sequon are N-glycosylated, because of conformational or other constraints during glycoprotein folding. Thus, Asn-X-Ser/Thr sequons encoded by an mRNA are referred to as potential N-glycan sites.[42] Although defined amino acid sequons that accept O-glycans have not been identified, certain amino acids are preferred near the serine/threonine attachment residues, such as proline.[42]

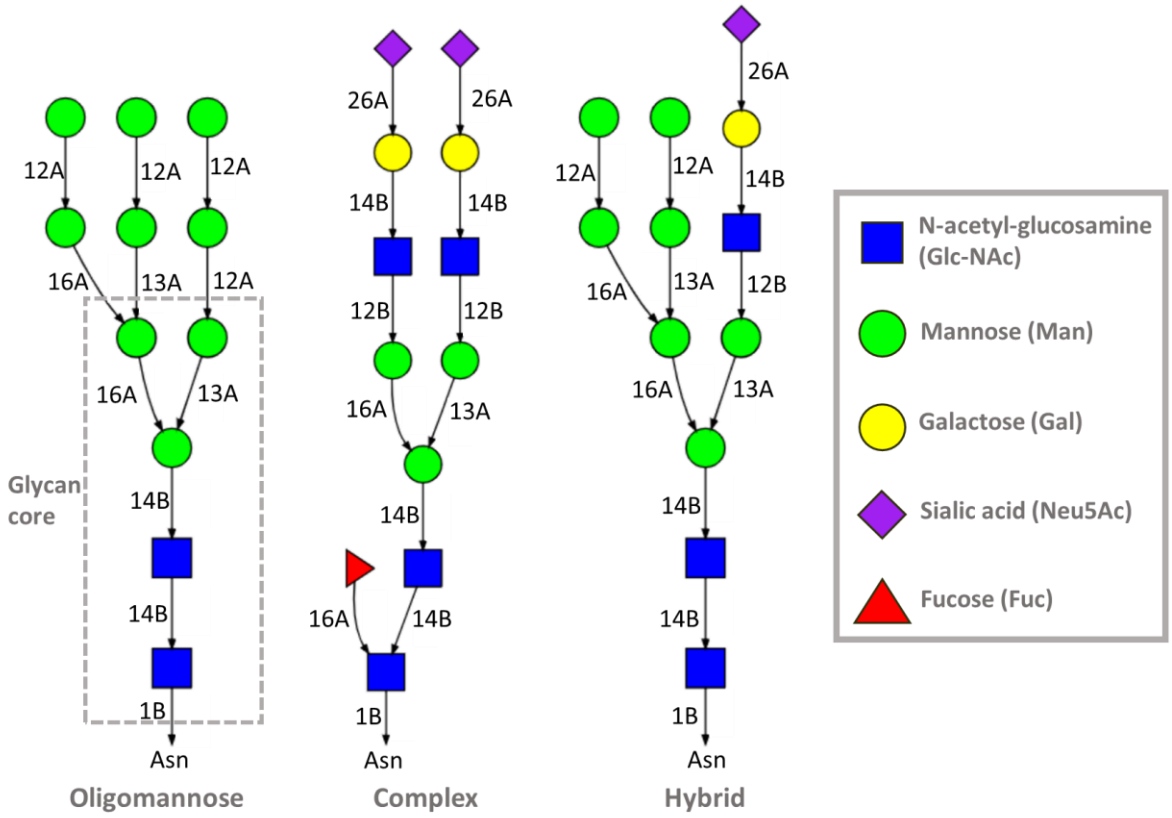


Figure 1.3: Structure of oligomannose, complex and hybrid N-glycans in Symbol Nomenclature for Glycans (SNFG) representation. The types of sugars that correspond to each shape and color are explained within the legend. The numbers in each linkage denote the positions of the two atoms from

the two sugars that participate in the glycosidic bond. The letters A and B refer to the configuration (alpha or beta) of the anomeric carbon (C1) that forms the glycosidic bond. The glycan core, which is common in all N-glycans, is shown within a dashed-line box.

All eukaryotic N-glycans share a common core sequence, Manα1-3(Manα1-6)Manβ1-4GlcNAcβ1–4GlcNAcβ1–Asn-X-Ser/Thr, and are classified into three types: (1) oligomannose, in which only Man residues extend the core; (2) complex, in which "antennae" initiated by GlcNAc extend the core; and (3) hybrid, in which Man extends the Manα1-6 arm of the core and one or two GlcNAc-initiated antennae extend the Manα1-3 arm (Figure 1.3).

The Trk receptors have only N-linked glycans in their EC segments, while p75 has one N-glycan in its N-terminal domain and several O-glycan sites in the EC linker to the TM helix.[43–50] Specifically, TrkA has 13 potential N-glycosylation sites, based on the Asn-X-Ser/Thr sequon, out of which at least 9 are actually glycosylated.[50] Six of these 9 sites are known from the crystal structure of the full-EC TrkA domain bound to NGF.[28] For the case of TrkB, out of the 12 potential N-glycosylation sites, 10 are glycosylated and their exact positions are known.[44,48] The p75 receptor has a complex N-glycan at Asn 60, while it also contains O-glycoside–linked oligosaccharides, clustered in the serine/threonine-rich stalk EC linker domain immediately adjacent to the membrane spanning segment.[45] The mature NTs do not have any attached glycans, although there is evidence that the pro-domain of the pro-neurotrophins is glycosylated.[51–53]

The types of glycans in the TrkA and TrkB receptors are not known, although neuraminidase digestion of the full-glycosylated TrkA receptor, suggests the existence of sialic acid in the sugar moieties of the receptor (complex and hybrid glycans in Figure 1.3).[54,55] The type of the single N-glycan in the EC domain of p75 receptor is known to belong in the complex category of N-glycans (Figure 1.3).[43] The O-glycans in the EC linker are known to contain sialic acids, while it has a core of Galβ1-3GalNAc-O-Ser/Thr.[56] The O-glycans are suggested to be important for proper folding of the linker.[46,56]

TrkA glycosylation is very important for receptor localization to the plasma membrane and ligand-dependent activation.[50] When the receptor is not glycosylated, it is trapped intracellularly and it is constitutively active. Thus, it cannot trigger the Ras/Raf/MAP kinase cascade, which initiates in the plasma membrane, and consequently it cannot promote neuronal survival.[50] This is not the case of the p75 receptor, as removal of its N- and O-glycosylation does not impair receptor transportation to the cell membrane.[45] Instead, inhibition of the O-glycosylation of the EC linker of p75 receptor affects the

apical sorting of the receptor.[45–47] The role of the single N-glycan in the EC domain of the p75 receptor is not known.

Structure of neurotrophin/Trk-receptor complex

The current amount of knowledge hampers the complete understanding of how the biological signal triggered by neurotrophin binding is transformed into the ordered set of biochemical cascades. In 1999 the crystal structure of the extracellular domain D5 (IgG-C2) of TrkA complexed with NGF was solved at resolution 2.2 Å.[29] Two ligand-receptor contact regions have been unveiled. The first, known as the ***common patch***, comprises the four central β-strands and the loop I from NGF, and the AB, C'D, EF loops and the C-terminus from TrkA (Figure 1.4). The common patch comprises residues well conserved in both the neurotrophin and Trk receptors family. The second region known as the ***specific patch***, (specific for this complex), includes residues 2-13 from the N-terminus of NGF and from the ABED sheet of TrkA (Figure 1.4).[29]

Molecular dynamics simulations and Poisson-Boltzmann electrostatic calculations have shown that the ligand - receptor interface is highly hydrated by the presence of about 20 ordered water molecules and 35 non - ordered water molecules, which can also affect the large-scale motions of the complex.[57] Calculation of the NGF-TrkA binding affinity suggests that the free energy of binding is not achieved by direct protein-protein interactions, but rather by exploiting the water molecules located at the interface, which act as sort of a "glue" between ligand and receptor. Principal component analysis of the large-scale motions has shown that a rotation of the TrkA subunits around the NGF homodimer takes place. This information suggests that the TrkA/NGF complex undergoes a well-defined motion on the nanosecond time scale of the performed simulations. The rotation involves residues of the NGF homodimer, particularly from the N-terminus and from the L1 and L4 loops (Figure 1.4) and of TrkA (residues from 348 to 355). The latter also form persistent, direct and water-mediated H-bond interactions, which might provide the contact-surface flexibility that is needed for the rotational motion.[57]

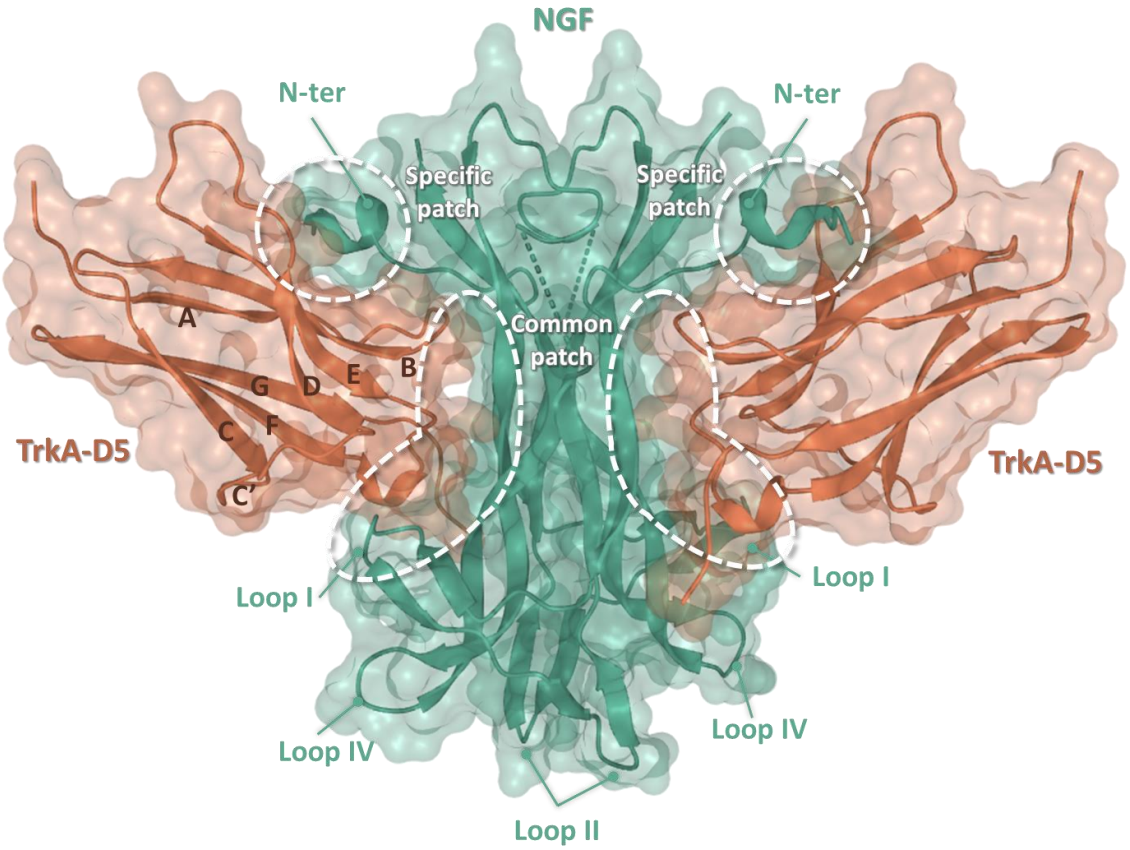


Figure 1.4: First crystal structure of the NGF homodimer (green) in complex with the TrkA-D5 homodimer (orange) (PDB ID: 1WWW).[29] The proteins are shown in cartoon and molecular surface representations. The dashed-circled regions indicate the *common* and *specific patches* of interactions between NGF and TrkA-D5.

The crystal structure of the complete extracellular (EC) domain of TrkA complexed with NGF was solved in 2007 at resolution 3.4 Å.[28] In this study, crystals of the complex between human TrkA-EC domain (residues 33–423) and human NGF were prepared and used for data collection. Techniques of TrkA de-glycosylation to improve diffraction were avoided, since there have been reports that TrkA N-glycosylation can regulate its bioactivity.[50] As mentioned before, the TrkA-EC structure in the full 2:2 NGF/TrkA-EC complex, contains the leucine-rich regions (D2) capped by two cysteine-rich domains (D1, D3 domains) and the Ig-C1 and Ig-C2 domains (D4 and D5, respectively) of the receptor (Figure 1.5).[28] The interfaces of the D3/D4 and D4/D5 domains appear highly structurally constrained, suggesting that

large-scale conformational changes are not likely to occur upon binding of NGF, as has been hypothesized in an allosteric model of TrkA activation.[58]

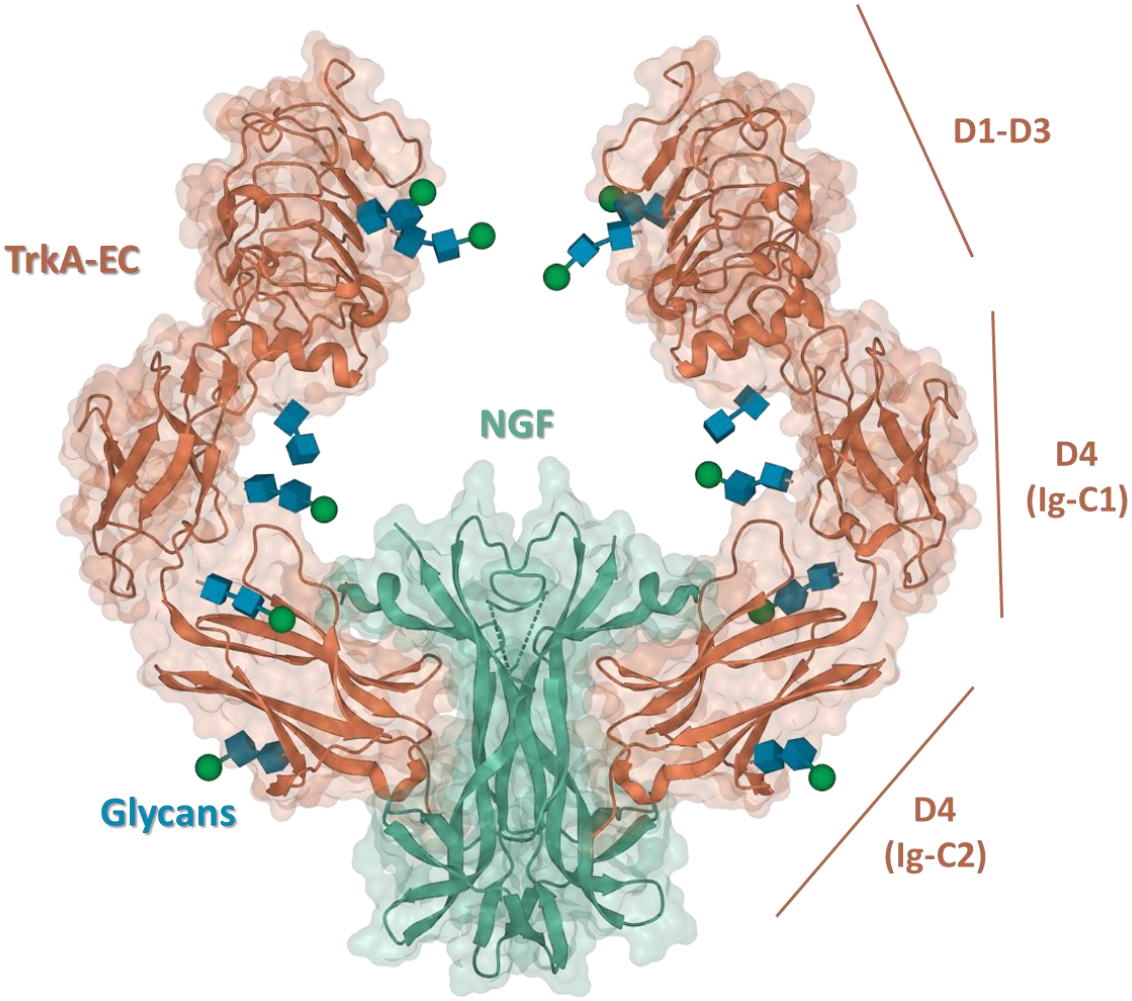


Figure 1.5: Crystal structure of the NGF homodimer (green) in complex with the full extracellular TrkA homodimer (orange) (PDB ID: 2IFG).[28] The proteins are shown in cartoon and molecular surface representations. The different domains of the full extracellular TrkA are indicated in the figure. N-glycans that were solved in the structure are shown in 3D-Symbol Nomenclature For Glycans (3D-SNFG). The blue cubes correspond to N-acetyl glucosamine (GlcNAc) and the green spheres to mannose (Man).

The region of complexation between NGF homodimers and TrkA-EC domains (Figure 1.5) shows no substantial deviations from the previously proposed truncated complex structure (Figure 1.4).[29] The TrkA-EC domains, which contain the previously missing N-terminal D1-D4 domains, are projected upward and away from the centrally located NGF and they do not form any interactions with it. The C-

terminal D5 (Ig-C2) domain of TrkA engages the sides of the NGF homodimer, as previously shown (Figure 1.4).[29] Each copy of TrkA-EC binds to each side of NGF homodimer, confirming the originally proposed 2:2 stoichiometry of the NGF/TrkA-D5 complex.

The crystal structure of the TrkB-D5 domain bound to NT-4/5 was also solved in 2001 at resolution 2.7 Å.[59] The structure displays high similarity with that of the TrkA-D5 bound to NGF (Figure 1.6).[29] In the complex, the N-terminal loop of NT-4/5 is stabilized by interactions with the receptor and forms an ordered 3_{10} helix. This loop is usually disordered in unliganded structures of NT-4/5.[23] As for the case of the TrkA-D5/NGF complex, here as well the NT and the receptor interact in two regions, the common, or conserved, and the specific patch, accounting for NT specificity (Figure 1.6). The consistency in structure and interacting regions for the two complexes (TrkA-D5/NGF and TrkB-D5/NT-4/5) is thought to underline the significant cross-reactivity observed between certain Trk receptor-NT pairs.

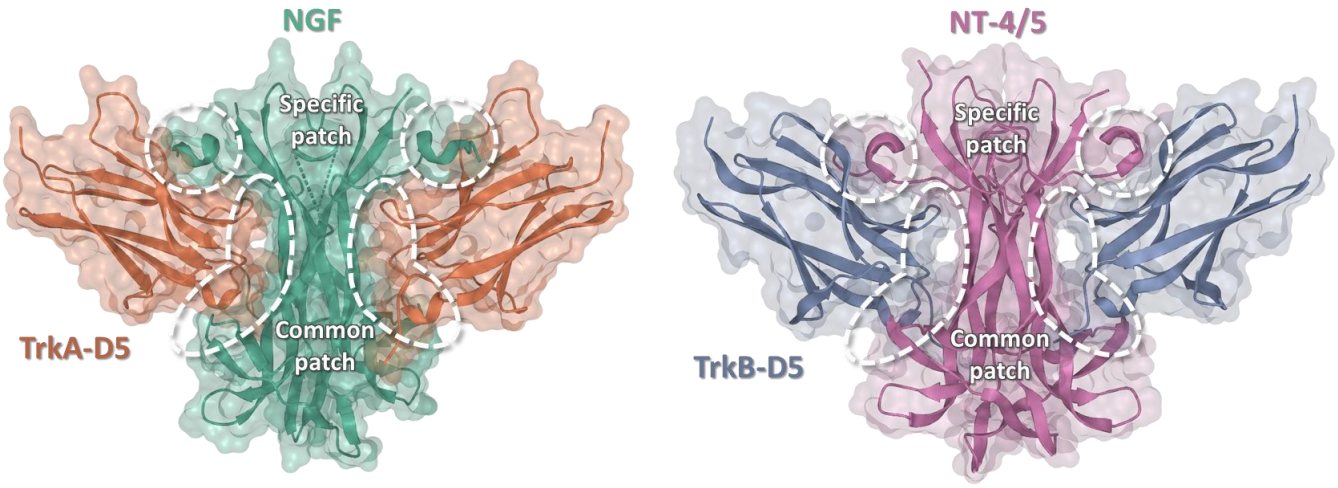


Figure 1.6: Crystal structures of the NGF homodimer (green) in complex with the TrkA-D5 homodimer (orange) (PDB ID: 1WWW),[29] and of the NT-4/5 homodimer (pink) in complex with the TrkB-D5 homodimer (grey) (PDB ID: 1HCF).[59] The proteins are shown in cartoon and molecular surface representations. The dashed-circled regions indicate the *common* and *specific patches* of interactions between NTs and receptors.

Function of neurotrophin/Trk-receptor complexes

Trk receptors display intrinsic catalytic activity. Mature neurotrophins bind to the extracellular domains of their preferred Trk receptors (Table 1.1) and induce ***receptor dimerization****, or* the ***rearrangement of preformed inactive receptor dimers***.[60–63] As a direct consequence the Trk-kinase domains of the two Trk-monomers obtain a suitable orientation which allows them to come into close proximity. In this way the kinase domain of one Trk monomer catalyzes the ***trans-autophosphorylation*** (i.e. phosphorylation of one kinase by the other of the receptor dimer) of several evolutionarily conserved tyrosine residues on the kinase domain of the other Trk monomer. This phosphorylation leads to major structural reorientations within the kinase domain, enabling adenosine triphosphate (ATP) binding and complete catalytic activation.[64–66]

The kinase domain of the Trk receptors consists of two linked lobes. The N-terminal lobe (***N-lobe***) comprises five β-sheets and one αC helix, which is important for the catalytic activity of the enzyme. The C-terminal lobe (***C-lobe***) is considerably larger and encompasses a number of β-sheets and α-helices (Figure 1.7).[64] The N-lobe and C-lobe are linked by a hinge region, which is crucial for the functionally inactive structure of kinase domain. Specific residues and motifs vital for the function of the enzyme are found in each one of the two lobes. A large loop between the 8th β-sheet and the 9th β-sheet of the C-lobe is the ***activation loop***, in which specific tyrosine residues undergo phosphorylation during kinase activation.[64]

As mentioned before, phosphorylation of specific tyrosine residues in the intracellular segment of TrkA is necessary for kinase activity. In human TrkA, three of these residues (Y676, Y680, and Y681, based on the UniProt ID: P04629-1) are in the activation loop of the intracellular kinase domain (Figure 1.7). Phosphorylation of these residues is synonymous to tyrosine kinase activity.[67] Although dimerization is necessary for receptor activation, it is not enough to induce it. In fact, preformed, yet inactive, Trk receptor dimers can also occur by spontaneous dimerization in the cell membrane.[62,63,68] Upon receptor activation, additional tyrosine residues are phosphorylated and they act as docking sites for adaptor proteins that propagate signalling within the cell. For the case of TrkA, these tyrosines are Y496 and Y791, and they are located in the juxtamembrane (JM) domain and the C-terminal linker, respectively (Figure 1.7).[8,69]

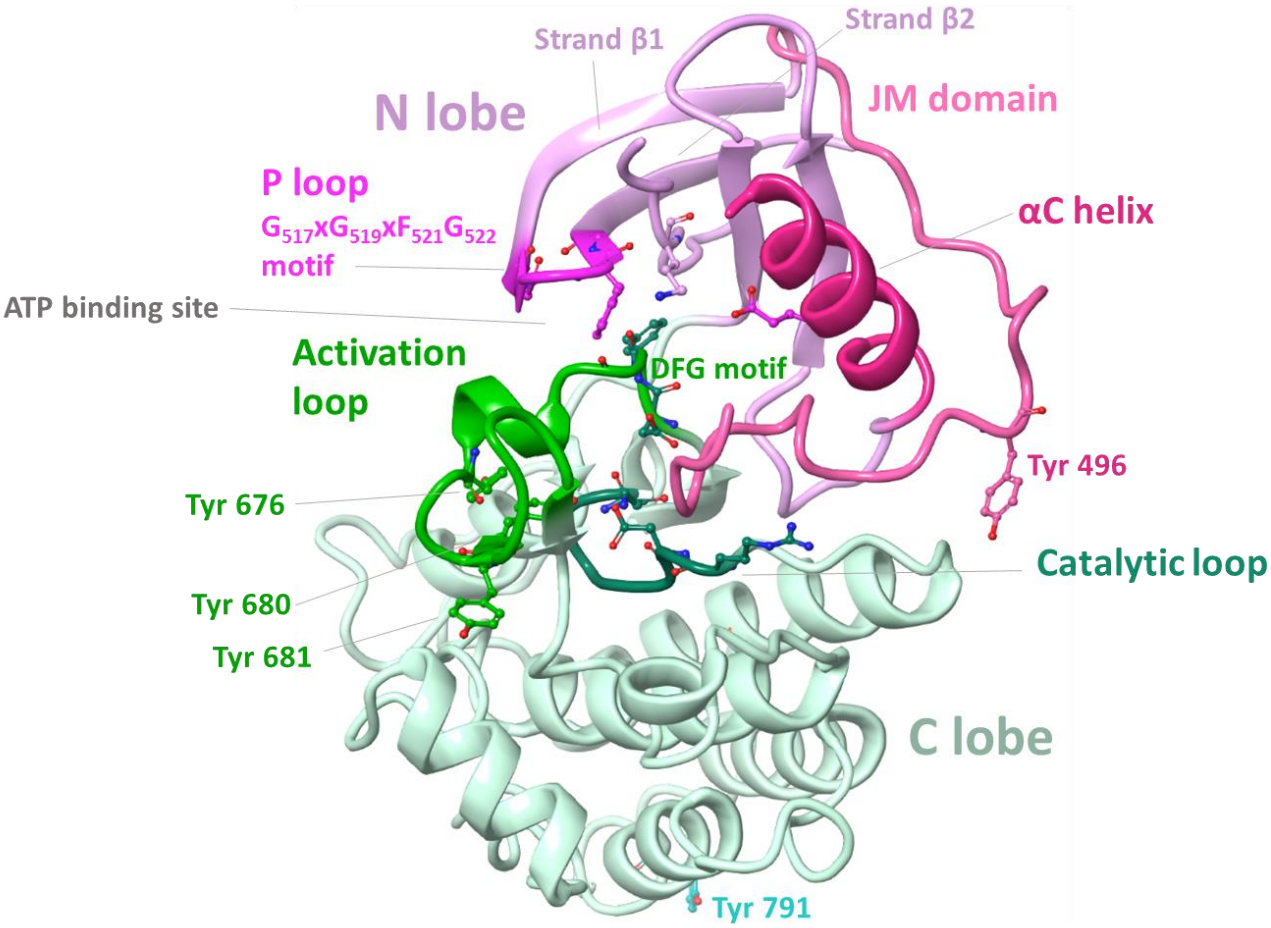


Figure 1.7: Crystal structure of the inactive TrkA kinase domain in cartoon representation (PDB ID: 5H3Q).[70] The kinase domain structure can be divided in two parts, the N-lobe (light pink) and the C-lobe (light green). The ATP binding site is located in a deep cleft at the interface of the two lobes and sits beneath a highly conserved loop (P loop in magenta) connecting strands β1 and β2. This phosphate-binding loop, or P loop, contains a conserved glycine-rich sequence motif (GXGXϕG), where ϕ is usually tyrosine or phenylalanine. The glycine residues allow the P loop to approach the phosphates of ATP very closely and to coordinate them via backbone interactions. The conserved aromatic side chain caps the site of phosphate transfer. The glycine residues make the P loop very flexible in the absence of ATP, a fact that facilitates the binding of small molecule inhibitors. Phosphorylation of the activation loop (green) tyrosines (Y676, Y680, and Y681) stabilizes it in an open and extended conformation that is permissive for substrate binding. In the unphosphorylated state, the activation loop collapses into the active site, blocking the binding of the substrate, as shown here. Optimal phosphotransfer requires the precise spatial arrangement of several catalytic residues that are absolutely conserved among all known kinases. An important role is played by a conserved loop structure at the base of the active site, called

the catalytic loop (dark green). A highly conserved to all kinases DFG motif is situated at the base of the activation loop. The structure of this motif, and in particular the conformation of Asp 668, is tightly coupled to phosphorylation of the activation loop. After the phosphorylation of the activation loop tyrosines, two more tyrosines are also phosphorylated by the other kinase monomer through trans-autophosphorylation. These are Y496 and Y791, which are located in the juxtamembrane (JM) domain and the C-terminal linker, respectively.

Another important motif, designated as the ***DFG motif*** (Aspartic acid–Phenylalanine–Glycine), at the end of the activation loop, participates in the formation of the ***ATP-binding pocket*** and is located at the hydrophobic interface between the N-lobe and C-lobe (Figure 1.7). This DFG motif is adjacent to a flexible ***phosphate-binding loop***, or ***P loop*** (also called Glycine-rich loop) of the N-lobe, which contains a conserved glycine-rich sequence (GXGXϕG), where ϕ is usually a tyrosine or phenylalanine. In the presence of ATP, the P loop interacts with the adenine-ribose sugar of ATP, resulting in the coordination of the nucleotide.[66] At the same time, the αC helix of the N-lobe is rotated into an "active" position and is stabilized by a salt-bridge formed between the αC helix and the 3rd β-sheet of the N-lobe, which in turn facilitates αC helix coordination with the β-phosphate of the ATP molecule.[71] Hence the main features of the ATP-binding pocket are: (1) the P loop (2) the DFG motif and (3) the αC helix of the N-lobe.

In the inactive state of the TrkA-kinase domain, the DFG motif adopts an inactive "DFG-out" conformation in which the DFG motif is positioned into the ATP-binding pocket blocking access to the nucleotide.[72] The "DFG-out" conformation is stabilized by a T-stacking interaction between the phenylalanine residue of DFG, and another phenylalanine residue at position 592 in the hinge region of the N- and C-lobes. F592 residue is also known as the "gatekeeper" residue. In this state, a salt bridge between residues K547 (of the 3rd β-sheet) and E563 (of the αC helix) within the N-lobe mentioned above, cannot be formed. As a result, the N-lobe acquires an inactive conformation where the αC helix is positioned away from the ATP-binding pocket.[73] In the activation loop of the TrkA-kinase domain there are also three sites of regulatory phosphorylation, the three tyrosine residues: Y676, Y680, and Y681. Trans-autophosphorylation of these residues, requires the close proximity of the partner TrkA-kinase domain.

In the active state of the kinase domain, the activation loop adopts a new "active" conformation, leading to the repositioning of the DFG motif in such a way that the interaction between the phenylalanine residue of DFG and the gatekeeper residue is disrupted and the ATP-binding pocket becomes accessible to the nucleotide.[66] Upon ATP binding, the αC helix is rotating into an "active**"** position where interactions are now permissible between ATP, the P loop, and the αC helix.[66] The kinase domain is then considered to have adopted a fully catalytic configuration prone to binding and activation of adaptor proteins.[73] This active conformation is also stabilized by coupling of the negatively charged phospho-tyrosine residues with residues in their backbone vicinity, and specifically in the ***catalytic loop*** (Figure 1.7), in a conformation that potentiates kinase activity and the ability of both Y496 and Y791 to be accessed by several adaptor proteins.[8]

The conformational switch from an inactive to an active Trk receptor following ligand binding involves a sophisticated process of key residue repositioning and overcoming of various bonding constraints, which typically stabilize the inactive state. Cross-talk of Trk receptors with p75, which leads to changes in Trk-ECD that result in a high affinity site (which will be presented later in more detail), may also enable the formation of an inactive but more readily activated repositioning of the Trk-kinase domain.

The observation that the three tyrosine residues (Y676, Y680, and Y681 for human TrkA) in the activation loop participate in Trk activation, suggest that ligand binding and consequent phosphorylation of these tyrosine residues, leads to removal of an inhibitory interaction known as ***cis-autoinhibition***. Several research groups have supported that inactive TrkA exists in a monomer–dimer equilibrium, but there is a disagreement on whether the majority of inactive receptors are monomeric or dimeric.[60,74] Irrespective of the oligomerization state though, the inactive state of the Trk kinase domain is maintained through *cis*-autoinhibition of the enzyme which is an intramolecular regulatory mechanism of interactions between different motifs of the kinase domain. The most crucial question is how ligand-induced dimerization of the extracellular regions of inactive Trk monomers leads to activation of the intracellular tyrosine kinase domain, and release of the *cis*-autoinhibitory interactions.

In general, in the receptor tyrosine kinases (RTKs) superfamily the structures of inactive kinase domains differ significantly from one receptor to another, and this variation reflects the diversity of the regulatory mechanisms in this superfamily. Each kinase domain is uniquely *cis*-autoinhibited by a sequence of intramolecular interactions specific for its receptor. Release of *cis*-autoinhibition, following neurotrophin binding, is the key event that triggers RTK activation. Until now the most studied mechanisms of *cis*-autoinhibition are: (1) The kinase autoinhibition by the activation loop; this is the case

of the Insulin and the fibroblast growth factor (FGF) receptors. (2) The autoinhibition mechanism by the juxtamembrane domain, in which RTKs are *cis*-autoinhibited by elements outside the kinase domain itself. (3) The autoinhibition mechanism by C-terminal sequences.[75]

Binding of neurotrophins to Trk receptors appears to be the principal mechanism for activation of these receptors. But the affinity and specificity of Trk receptor for neurotrophins is regulated by the p75 receptor. The presence of the p75 is required for the exposure of high-affinity binding sites of TrkA to NGF, as mentioned before.[37,40,76] Kinetic studies of NGF interactions with TrkA and with p75 proved that, although dissociation constants for each receptor are very similar, the kinetics are quite different. NGF associates with and dissociates from p75 much more rapidly than from TrkA, and the presence of p75 increases the rate of NGF association with TrkA.[77] The inevitable question of how neurotrophin binding converts a Trk receptor from an "off" state to a fully functional active "on" state remains.

The mechanism of *cis*-autoinhibition of Trk receptors is yet unknown. Nevertheless based on models proposed for other members of RTK receptors and on early functional studies, it seems most likely that transmembrane (TM) helical domain rotations are involved in Trk family *cis*-autoinhibition.[60,78–80] Rotation of TM helices, propagated intracellularly, as a requirement for kinase domain activation have been also previously described in multiple cases of kinase receptors.[81–86]

Structural studies suggest that inactive TrkA homodimers are stabilized by interactions of the two TM helical domains. Particularly, these interactions involve an LxxFAxxF motif (residues L424-F431), as indicated by the nuclear magnetic resonance (NMR) structure of the TrkA helical homodimer (Figure 1.8A,B).[60,74] Mutagenesis studies showed that a different dimerization interface is involved in the active state of the TrkA receptor, with residues located on the opposite side of the α-helices.[60] In Figure 1.8C, the inactive interface, in which residues from the NMR interface lie, is indicated with a red curved line, while the active interface residues are indicated with the green curved line. Rotation from the inactive to the active interface takes place upon NGF binding. Interestingly, both the inactive and active interfaces, involve hydrophobic residues, with the more polar side of the helix not participating in any of the two interfaces.

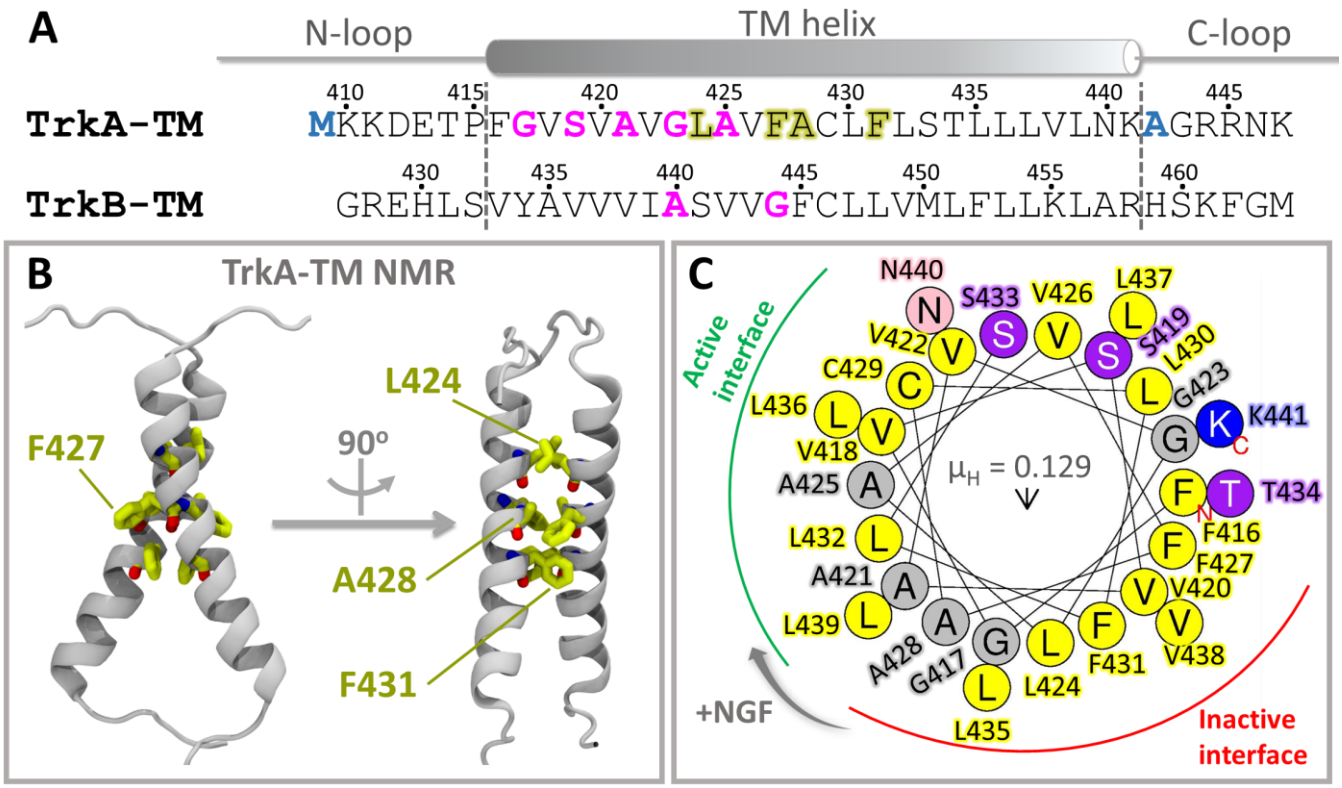


Figure 1.8: (A) Sequences of the transmembrane (TM) domains of TrkA and TrkB receptors. The helical part from the TrkA NMR structure (PDB ID: 2N90)[60] is indicated with the dashed lines, while the N- and C-terminal loop sequences in the solved NMR structure are also shown. Mutated residues in the NMR sequence compared to the human TrkA are colored blue. Residues participating in the GxxxG-like motifs are colored magenta, while the TrkA NMR interface LxxFAxxF motif is highlighted in yellow. (B) Cartoon representation of the TrkA NMR structure (PDB ID: 2N90), with the interface LxxFAxxF motif shown in yellow licorice representation. (C) Helical wheel representation of the TrkA TM helix. The arrow indicates the magnitude and direction of the hydrophobic moment (μ_H). Nonpolar and aromatic residues are yellow, glycines and alanines are grey, charged residues are blue, polar uncharged residues are purple, and asparagine is pink. The letters N and C in the projection indicate the N-and C-terminals, respectively. The active and inactive interfaces, as determined from mutagenesis studies,[60] are indicated with green and red lines, respectively. Rotation from the inactive to the active interface takes place upon NGF binding. The helical wheel was created with the HeliQuest webserver.[87]

It was proposed that the alanine residue at position 428 is essential for the interaction between the dimeric Trk-TM interfaces during receptor activation.[60] As the TrkA dimer is transitioned from an inactive

to an active state, the TrkA-TM regions pivot around A428 to allow the two monomers to interact more closely and form an active interaction interface. When NGF binds to TrkA-D5, a conformational change in the TrkA-EC domain is induced. Concomitantly, the extracellular proximal juxtamembrane subdomains of the dimer are brought into close proximity. This leads to the pivoting movement around A428 and propagates an allosteric shift of the TrkA-IC domain equivalent to a 100° rotation of the kinase domains. Then specific tyrosines are exposed and trans-autophosphorylation occurs.[60] However, whether this 100° rotation is enough to overcome the *cis*-autoinhibition of TrkA, requires further investigation.

Another interesting point is that the NMR interface of TrkA does not involve residues participating in GxxxG-like motifs, which are usually found at the interface of TM helices.[88–90] GxxxG-like motifs consisting of small residues, such as glycine, alanine and serine, permit close packing of TM helices, by allowing H-bond interactions between the backbone atoms of the two helices. The N-terminal TM region of TrkA encompasses three GxxxG-like putative dimerization motifs, $G_{417}xxxA_{421}$, $S_{419}xxxG_{423}$, and $A_{421}xxxA_{425}$, while the TM sequence of the TrkB receptor has only one motif, the $A_{440}xxxG_{444}$ motif (magenta residues in Figure 1.8A). Molecular dynamics simulations of the TrkB-TM helical dimer have shown that the homodimer interacts stably in the $A_{440}xxxG_{444}$ motif, while the crossing angle between the two helices varies based on cholesterol concentrations in the membrane.[91] Antidepressant drugs, such as fluoxetine, were found to bind at the interface of the TrkB-TM helices, stabilizing the helices in a crossing angle that is not dependent on the membrane cholesterol levels.[91]

Apart from the TM domain, the TrkA-EC domains have also been identified as crucial for the regulation of the *cis*-autoinhibition of the receptor.[35] In the absence of ligand and upon deletion of either one or both of Ig-C subdomains (D4 and D5), active Trk dimers are promptly formed. This paradox is not observed in wild-type (WT) receptors or full EC-truncated mutants. Evidently, the relative conformation of the EC domain influences the *cis*-autoinhibition of the TrkA kinase domain. Alterations of ligand binding site accessibility interferes with intracellularly transmitted information and affects the subsequent transition of the receptor from an “off” state to a fully functional active “on” state.[92] It also suggests that extracellular and intracellular structural changes do not occur independently of each other.

Cross-talk between Trk and p75 receptors

As mentioned before, it is well-known that p75 enhances Trk signaling events. Several studies have supported that co-expression of p75 is vital to observe high-affinity binding sites to TrkA for NGF and that p75 is able to modulate TrkA such that it has increased sensitivity to lower concentrations of NGF.[15,40,93,94] At least two decades of research have revealed that p75 increases the affinity and specificity of TrkA and TrkB for their cognate ligands and has a profound impact on cell fate *in vivo*. Despite the accumulated amount of knowledge related to the importance of this signaling system for the development, and maintenance of neuronal networks, and the response to diseases, the mechanism by which p75 modulates Trk–neurotrophin binding remains unknown. The questions that must be answered are: (1) how does p75 augments this mechanism, (2) what is the exact mechanism by which high-affinity binding sites are attained, and (3) is it the same or a different mechanism that promotes Trk-ligand specificity?

Several alternatives for the mechanism by which p75 modulates Trk activity have been proposed. The most interesting are: (1) the formation of hetero-complexes stabilized by p75, (2) the ligand-passing model, (3) allosteric modulation of Trk by p75 for high-affinity receptor formation, and (4) the inside-out activation model of TrkA by p75. These mechanisms are briefly presented below.

Stabilized 1:2:1 hetero-complex model: This model involves the formation of the TrkA, neurotrophin dimer and p75 in a 1:2:1 hetero-complex (Figure 1.9B),[95] that is highly stable and facilitates prolonged signaling.[28,96,97] Each receptor would bind the NGF dimer in antiparallel orientations, as structural evidence indicates that NGF binds to TrkA and p75 in a reversed way.[28] Although there are no steric clashes in this complex, the model includes only one TrkA monomer. Since catalytic activation of the tyrosine kinase domain of TrkA requires the existence of two Trk molecules in close proximity,[66] the functional value of this model is questionable.

The ligand-passing model: This model, is based on affinity kinetics and is known as the "ligand-passing theory".[98] It suggests that p75 is the first of the two receptors that bind NGF at the plasma membrane. In this way, the local concentration of NGF, which is available to bind to TrkA, is increased. Then, p75 forms a temporary complex with TrkA by sandwiching the NGF dimer between the two receptors as it is passed from one receptor to the other. The model steps end with NGF interacting with TrkA only (Figure 1.9C).[98] However, this model is inconsistent with data which show that the p75 extracellular domain, where NGF binds, is not required for the formation of the high-affinity receptor complex.[99,100]

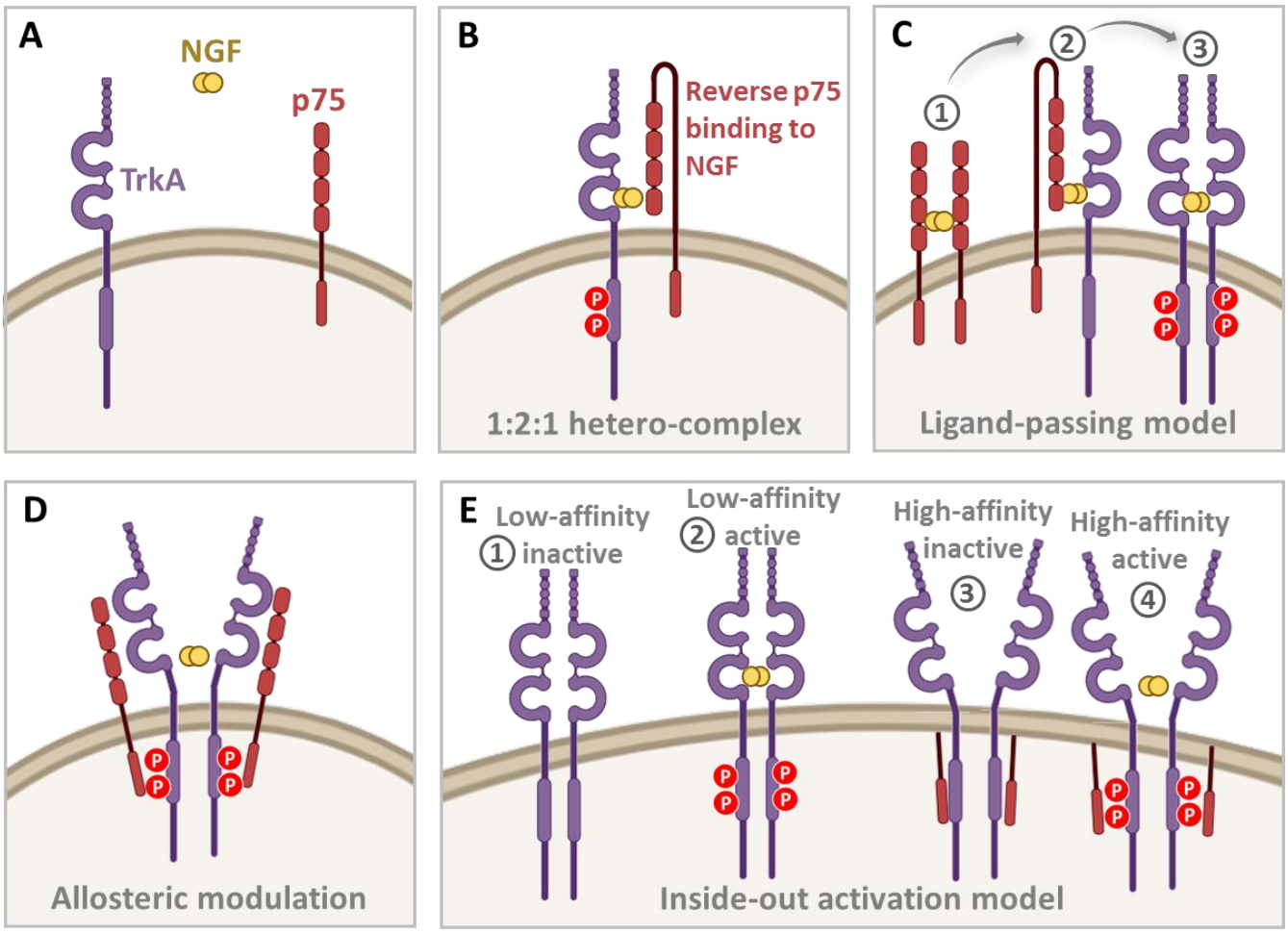


Figure 1.9: Proposed mechanisms for creation of high-affinity NGF binding sites in the TrkA receptor by the p75 receptor. (A) Schematic representation of the NGF homodimer (yellow), the TrkA receptor monomer (purple) and the p75 receptor monomer (red). (B) Stabilized 1:2:1 hetero-complex model: p75 binds in a reversed way to NGF. TrkA activation is depicted by red phosphorylation sites in the intracellular domain. (C) Ligand-passing model presented in three steps. (D) Allosteric modulation model: p75 interacts directly with TrkA, inducing an allosteric conformational change in TrkA that allows better access of NGF to the TrkA binding site. (E) Inside-out activation model: TrkA exists in four states; low-affinity inactive form, a low-affinity active form, a high-affinity inactive, and high affinity active form. The last two are induced by binding of the intracellular segment of p75 to TrkA. The figure was created with BioRender.com.

The allosteric modulation model: Based on biochemical evidence, the allosteric modulation model supports a direct interaction between the two neurotrophin receptors TrkA and p75 , which causes

allosteric modulation of the TrkA receptor such that NGF affinity and specificity is increased.[40,99–101] This modulation may improve the access of NGF to the ligand-binding site in the D5 domain of TrkA or may trigger the participation of other residues of the TrkA-EC domain in the interaction with the ligand (Figure 1.9D). This approach explains the inability of an NGF triple mutant (K32A, K34A, E35A) to bind TrkA with high affinity. As high-affinity binding may be mediated through additional or alternate residues in either TrkA-EC domain or the NGF dimer, the high-affinity binding should be prevented by the NGF mutation.[102,103] However, this model is not yet confirmed from attempts of solving the structure of TrkA bound to NGF in the presence of p75.

The inside-out activation model of TrkA by p75: According to this more recent model, after ***regulated intramembrane proteolysis (RIP)*** of p75, a cleaved p75 intracellular fragment can interact with the TrkA kinase domain to induce an allosteric conformational change to the TrkA-EC domain, promoting the formation of inactive receptors (which are primed for high-affinity binding), without initiating constitutive kinase activation.[104] This model suggests that a TrkA receptor may be found in four possible states: a low-affinity inactive form, a low-affinity active form, a high-affinity inactive, and high affinity active form (Figure 1.9E).[26] Since p75 does not directly affect the catalytic activity of the TrkA kinase domain,[104–106] further structural evidence is required to elucidate the mechanism by which the p75-IC domain could facilitate an outward propagated conformational change of the TrkA-EC domain, resulting in high-affinity binding sites for neurotrophins.

Summarizing the knowledge to-date about the structure and function of NTs/Trk-receptor complex it can be mentioned that:

1. In the absence of NGF binding, inactive TrkA receptors exist in a monomer–dimer equilibrium. TrkA-monomers interact through the LxxFAxxF motif in the TM domain and the kinase domain occupies the "*DFG-out*" inactive conformation.

2. TrkA dimers bind NGF with low affinity through the D5 domain and acquire a low-affinity active dimer conformation. TrkA-EC domain juxtamembrane residues are located proximally and the TM domains pivot through A428, inducing a 100° rotation of the kinase domains obtaining a configuration prompted to trans-autophosphorylation. The DFG motif and the activation loop of the kinase domain are repositioned to allow for ATP binding. It seems that the active interaction interface in the TM domain stabilizes the dimer, probably because of the rotation of the entire TM domain downstream to the proximal juxta-membrane residues.

3. TrkA dimers are complexed with p75-IC domain fragments, and adopt a high-affinity binding EC domain conformation. The stoichiometry of the p75-IC–TrkA complex is uncertain. TrkA dimers may be stabilized through the inactive LxxFAxxF motif of the TM domain, or the active interaction interface, or an alternate "high-affinity-receptor" interaction interface. The proximity of TrkA extracellular juxtamembrane residues and the rotational status of the TrkA-kinase domains in the presence of p75-IC fragments are unknown.

4. Upon NGF binding the p75-IC–TrkA active complex obtains a high-affinity EC domain-binding site configuration. Allosteric modulation to the Trk-EC domain may (i) improve ligand access to D5, (ii) reposition of D1 and D4 appropriately to promote NT-D5 interaction, and/or (iii) reveal a probable second NGF-binding site through a conformational change of the EC domain. TrkA extracellular juxtamembrane residues come in close proximity, while an active or alternate TM interaction interface is probably engaged, and the kinase domain obtains a catalytically viable position.

Activity of neurotrophin receptors in lipid rafts

Neurotrophin receptor activity has been often linked to the localization of the receptors in ordered, cholesterol and sphingomyelin-rich membrane microdomains, called lipid rafts.[107–111] Assays have shown that, upon NGF binding, TrkA translocates into lipid rafts, an important step for activation of the extracellular signal–regulated kinase (ERK) pathway.[112] TrkA is recruited and retained within lipid rafts by association with flotillin, which is also localized in these membrane microdomains, via the adapter protein, c-Cbl associated protein (CAP). Mutant forms of CAP that lack protein interaction domains do not allow TrkA localization to lipid rafts and consequently attenuate ERK activation.[112] NGF activity can also be regulated by two caveolins which are located in lipid rafts.[113] Caveolin-1 inhibits NGF-signaling by sequestering TrkA and p75 receptors together with the ribosomal S6 kinase 2 (Rsk2), which prevents cAMP response element binding protein (CREB) phosphorylation. In contrast, caveolin-2 promotes mitogen-activated protein kinase (MAPK), Rsk2 and CREB activation by NGF.[113]

Similarly to NGF, BDNF induces the localization of the TrkB receptor into cholesterol-rich lipid raft domains from non-raft regions of neuronal plasma membranes.[114] This translocation can be hindered by TrkB-kinase inhibitors, suggesting a role for the kinase domain in lipid raft translocation.[114] Disruption of lipid rafts prevented potentiating effects of BDNF on transmitter release in cultured neurons and synaptic response to tetanus in hippocampal slices.[114] Under hypoxic conditions, localization of TrkB in

lipid rafts is reduced, while supplementation of vitamin E can restore TrkB translocation of raft domains and promote neuronal survival.[115]

p75 receptor is also known to localize in lipid rafts, an event that is regulated by the cAMP-dependent protein kinase A (PKA).[116] Additionally, proteolytic cleavage, called shedding, of the extracellular domain of p75 is modulated by lipid rafts, as raft disruption enhances p75 shedding and thus the pathways that are related to it.[117] Overall, one can say that lipid rafts can function as traffic lights modulating and fine-tuning neuronal signaling.[118]

Apart from the general role of lipid rafts in receptor activity, specific lipids found in these domains of neuronal membranes have been identified to bind and activate the neurotrophin receptors. This is the case of the ganglioside GM1, which can activate the TrkA receptor.[119–121] Through the association with GM1, TrkA can enhance neurite outgrowth and neurofilament expression in rat PC12 cells elicited by a low dose of NGF that alone is insufficient to induce neuronal differentiation.[122] In the presence of GM1, there is 3-fold increase of NGF-induced autophosphorylation of TrkA as compared with NGF alone.[122] GM1 interacts specifically with the glycosylated TrkA receptor and not with the unglycosylated one.[123] Additionally, in cells deficient in GM1, TrkA was not localized in the cell membrane, suggesting that GM1 may be necessary for normal TrkA expression in the plasma membrane.[124] It has been suggested that the oligosaccharide portion of GM1 is responsible for TrkA-mediated neurodifferentiation, while photolabeling experiments have verified the direct interaction of GM1 with the TrkA-EC domain.[125] GM1 is also able to activate the TrkB receptor and induce neuroprotection.[126] Another study suggests that GM1 can activate the Trk receptors by increasing the release of neurotrophins.[127] Apart from GM1, exposure of TrkA to ceramide can also activate the receptor by promoting receptor dimerization.[128]

1.3 Neurotrophin signaling

The expression of Trk receptors is developmentally regulated. They are detectable very early in the developing hippocampus, cerebellum, spinal cord, and dorsal root ganglion (DRG). Although Trk receptors distribution changes with the progressive maturation of the central and peripheral nervous systems (CNS and PNS), immunoreactive Trk receptors were detected in neurons of the adult human hippocampus, nucleus basalis of Meynert, cerebellum, spinal cord, and DRG. This pattern of Trks distribution implies that mature Trk-receptor-bearing neurons retain the ability to respond to neurotrophins long after the terminal neuronal differentiation is completed.[129]

Nevertheless, TrkA and TrkC, are largely found on specific neuronal subtypes including sympathetic neurons and basal forebrain cholinergic neurons (BFCNs) of PNS. On the other hand, the TrkB receptor is widely expressed in a large number of neuronal cell populations of the CNS. Certain neuronal mechanisms are modulated by specific receptor–ligand combinations including: BDNF/TrkB-mediated promotion of synaptic plasticity and NT3/TrkC-mediated survival of peripheral proprioceptive neurons.[130,131]

Neurotrophins show similar expression patterns as their related receptors, through the PNS and CNS. The NGF neurotrophic factor is the most dominant neurotrophic factor in the PNS for sympathetic and sensory neurons, in relation to TrkA expression patterns. In contrast, BDNF is considered to be the most abundantly expressed factor in the CNS in relation to TrkB expression patterns.

Signaling pathways regulated by Trk receptors, affect many diverse neuronal functions, including: cell survival and differentiation, axonal and dendritic growth and arborization, synapse formation, and synaptic plasticity. The most studied signaling pathways activated by Trk receptors are: (i) the extracellular signal–regulated kinase (ERK)/mitogen-activated protein kinase (MAPK) pathway, (ii) the phosphoinositide 3-kinase (PI3K)/protein kinase B (Akt) pathway, and (iii) the phosphoinositide phospholipase C-γ (PLC-γ)/protein kinase C (PKC) pathway, which influence proliferation, neuronal survival, differentiation, and neurite outgrowth.[33,132–134]

Signaling through the ERK/MAPK pathway

Phosphorylation of the tyrosine residue Y496 in human TrkA, or Y516 in human TrkB, located in the juxtamembrane intracellular domain right after the TM helix, creates a binding site for the Src Homologous and Collagen-like (Shc) adaptor protein. Phosphorylation of Shc initiates the recruitment of the growth factor receptor-bound protein 2/mammalian son of sevenless protein (Grb2/SOS) complex at the membrane, which activates the Rat sarcoma (Ras) GTPase protein. Consequently, downstream transient activation of the rapidly accelerated fibrosarcoma (Raf)/mitogen-activated protein kinase (MEK)/extracellular signal-regulated kinase (ERK) (c-Raf/MEK/ERK) cascade occurs. Subsequent activation of the p38 mitogen-activated protein kinase (p38 MAPK[3]) and the p90 ribosomal S6 kinase (RSK), and their translocation into the nucleus, stimulate phosphorylation of the cAMP response element

[3] The p38 mitogen-activated protein kinases are a class of mitogen-activated protein kinases (MAPKs) that are responsive to stress stimuli. The p38MAPK is activated by a variety of cellular stresses including osmotic shock, inflammatory cytokines, lipopolysaccharides (LPS), ultraviolet light, and growth factors.

binding protein (CREB) transcriptional factor. The latter regulates transcriptional events leading to the activation of pro-survival genes.[135] ERK signaling also leads to local axonal growth.

This mechanism is a transient but not prolonged activation of ERKs. Prolonged ERK activation is also initiated at the Y490 site, but requires the Kinase D interacting substrate of 220 kDa/Ankyrin repeat-rich membrane spanning (Kidins220/ARMS) protein, which recruits the CT10 Regulator of Kinase (Crk), another adaptor protein. Binding to Crk activates the Crk SH3-domain-binding guanine-nucleotide releasing factor (C3G) and thus initiates the Ras-proximate-1 (Rap1)/Raf-dependent MEK/ERK signaling.[136]

Signaling through the PI3K/Akt pathway

The phosphatidylinositol-3 kinase (PI3K) plays an important role in Trk-dependent survival in many types of neurons. Trk-dependent activation of PI3K can occur through Ras-dependent or Ras-independent pathways.

The Ras-independent pathway involves activation of the Src homology and collagen (Shc)/ growth factor receptor-bound protein 2 (Grb2) complex, which binds the insulin receptor substrates -1, and -2 (IRS1 and IRS2) and the Grb-associated binder-1 (Gab-1), which induces binding to PI3K. That allows the interaction of PI3K with inositol phospholipid (PIP_2) substrates at the inner leaflet of the plasma membrane and their catalytic transformation to P3-phosphorylated phospho-inositides (PIP_3).[137] Phosphorylated lipids generated by PI3K result in a change of the local membrane composition, and induce membrane recruitment of the 3- phosphor-inositide-dependent kinase-1 (PDK1), a member of the serine/threonine kinases family, and translocation of the protein kinase B (PBK or Akt) to the plasma membrane.[138] Together with PIP_3, PDK1 activates the Akt protein kinase (also known as protein kinase B, PKB) another serine/threonine kinase.[139] Once Akt is activated by PDK1, it phosphorylates and deactivates its substrates, which are components of the apoptotic machinery.

Akt activity results in increased protein translation via the mammalian target of rapamycin (mTOR)-p70S6 kinase and the eukaryotic translation initiation factor 4E-binding protein 1 (4E-BP1) pathways, and in enhancing axonal growth through phosphorylation and inactivation of the glycogen synthase kinase-3β (GSK-3β). Additionally, activated Akt promotes neuronal survival by inhibiting the forkhead transcription factor like 1 (FKHRL1), which regulates expression of pro-apoptotic genes, by phosphorylating and therefore inhibiting the pro-apoptotic protein BCL2 associated agonist of cell death

(Bad). Also, Akt induces the phosphorylation and degradation of the inhibitor of nuclear factor kappa B (IκB) kinase, which is an inhibitor of the NFκB pro-survival pathway.[140,141]

Signaling through the PLC-γ/PKC pathway

Phosphorylation of the Y791 residue, which lies in the carboxyl-terminal of the TrkA receptor, with corresponding residues in TrkB and TrkC receptors, serves as a docking site for the phosphoinositide phospholipase C gamma (PLC-γ). Trk phosphorylated PLC-γ hydrolyses PIP_2 (phosphatidylinositol 4,5-biphosphate) and generates inositol triphosphate (IP_3) and diacylglycerol (DAG).[132] Release of IP_3 leads to release of intracellular Ca^{2+}, which in turn activates Ca^{2+}-dependent enzymes (CaM kinases), such as Ca^{2+}-calmodulin-regulated protein kinases and the phosphatase Calcineurin.

Release of Ca^{2+} and generation of DAG, activate the protein kinase C (PKC), which concomitantly stimulates ERK signaling via Raf, as well as the capsaicin VR1 channel and the transient receptor potential channel (TRPC), which contributes to the BDNF-induced rise of Ca^{2+} at growth cones and synapses.[142]

Other activities that are affected include the formation of TrkB-postsynaptic density 95 (PSD95) complexes at synapses, and CREB-dependent transcription. PLC-γ signaling in response to both NGF and BDNF has been implicated in chemoattraction of axonal growth cones. Also, PLC-γ signaling induces the transcription of a sodium channel. Furthermore, mice harboring a targeted mutation at the TrkB/PLC-γ docking site, Y816, have impaired hippocampal Long-Term Potentiation (LTP)[4], along with impaired induction of CREB and CaM kinase signaling.[143]

1.4 Neurotrophins in neurodegeneration

Neurotrophins (NTs) were originally described as innate factors that regulate differentiation and support growth during development of the vertebrate nervous system. NTs are essential for proper development of the embryonic nervous system through their ability to promote survival and stimulate neurite outgrowth from CNS and PNS neurons.[144] After development, NTs play a critical role in maintaining neuronal morphologies and functions as well as providing trophic and tropic activities in the neuronal

[4] Long Term Potentiation (LTP) is a continuing increase in the amplitude of excitatory postsynaptic potentials, as a result of stimulation of afferent axons. It is measured both as the amplitude of excitatory postsynaptic potentials and as the magnitude of the postsynaptic-cell population spike. LTP is most often studied in the hippocampus and is often considered to be the cellular basis of learning and memory in vertebrates.

responses to injury.[1,145] Although most experimental efforts have focused on activities of the mature neurotrophins, several studies have indicated that proneurotrophins are also biologically active and are actually opposing to the growth promoting and survival effects of the mature proteins.[146]

On the other hand, each of the NTs can also express their action through an apoptotic pathway initiated by binding to the p75 receptor. The spatial and temporal balance achieved between neuronal survival and death depends on the overall level of NTs present and the type of receptors that are expressed.[15]

NGF is the dominant neurotrophic factor in the PNS, acting on sympathetic and sensory neurons. On the other hand, BDNF is the predominant neurotrophin, in the CNS, utilized due to the abundant expression of TrkB. However, NGF provides trophic support specifically to the basal forebrain cholinergic neurons (BFCNs) which express TrkA. Studies in animal models expressing reduced levels of NGF and BDNF reveal that these two factors are essential for multiple functions throughout adulthood such as proper memory acquisition, memory retention, LTP, and cholinergic innervation.[147,148]

Since the discovery of NGF in the 1950s a large amount of experimental data has pointed to multiple roles for the NTs. Firstly, most neurotrophins are required during development and differentiation, during a time in which specific synaptic connections are being made and proper circuits are being formed. Secondly, neurotrophin signaling plays an important role in adulthood at a time in which continued maintenance and modulation of those connections is required for normal brain function. The notion that NTs support neuronal survival and function in the adult CNS, generates broad interest in the use of these factors to intervene in neurodegenerative disease.

Given the critical role of NTs in regulating neuronal functions, it is expected that a significant number of neurodegenerative disorders is associated with altered NGF and BDNF levels and with changed expression of their receptors. For example, neurodegenerative phenotypes similar to Alzheimer's disease (AD) are observed in an animal model in which half of the NGF level is neutralized by antibodies.[149] In the brains of AD patients and in animal models, reduced NGF levels in BFCNs have been measured.[150–152]

NGF and Alzheimer's disease (AD)

NGF is mainly synthesized in the cortex and hippocampus and retrogradely transported to the BFCNs where TrkA and p75 receptors are expressed. Given that NGF is the main neurotrophin that affects BFCNs, the expression and protein levels of NGF in this region have been investigated in AD. NGF

unquestionably has a strong impact in preventing cellular death of BFCNs. The key role for NGF was discovered in early studies on transected fimbria in which administering NGF upon transection was able to markedly reduce cholinergic neuron death, which was typically induced by the procedure.[153] BFCNs expressing TrkA and p75 receptors, undergo moderate degenerative changes upon aging, including cell atrophy, downregulation of transmitter-synthesizing enzymes, and mild cell loss both in animal models and humans, and these degenerative changes are markedly intensified in AD.[154–156] Furthermore, atrophic changes can be reversed in studies with aged animal models by infusions of exogenous NGF.[156] These findings suggest that failure of NGF signaling may underlie or contribute to the atrophic neuronal changes observed in aging or degenerative diseases. However, most of these studies agree that NGF levels are not substantially reduced in aging or AD. Other studies showed an increase in NGF in the frontal cortex and hippocampus but not in the parietal cortex, cerebellum, motor cortex, and entorhinal cortex in an AD group.[157] Thus, it is unclear whether or not loss of NGF contributes to AD pathogenesis.

On the other hand, available evidence does suggest that intrinsic neuronal changes might compromise the ability of aged BFCNs to derive neurotrophic support. In aged animal models, BFCNs exhibit a reduced capacity for generalized retrograde transport and a pronounced reduction specifically in the retrograde transport of NGF. Cholinergic neurons that do not transport NGF are severely wasted and downregulate TrkA expression.[152,158] Even though NGF levels are not reduced in the frontal cortex, they are substantially reduced in the BFCNs of aged animals and patients with AD.[159]

Taken together, these observations suggest that an intrinsic, age-related reduction in the ability of BFCNs to sustain retrograde NGF traffic, may contribute to the degenerative changes seen in BFCNs in aging that is markedly exacerbated in AD. BFCNs may be exquisitely sensitive to interruptions in NGF signaling as a by-product of the positive feedback in NGF signaling, such that an initially small failure in NGF signaling may be rapidly exacerbated by downregulation of NGF receptors and other changes.[160]

BDNF in diseased CNS

Two very important areas that control spatial memory and higher cognitive function have been reported to be immunoreactive for BDNF/TrkB: the frontal cortex (FC) and the entorhinal cortex (EC).[161] BDNF probably acts on such cortical neurons via an autocrine loop, because TrkB is expressed simultaneously by neocortical neurons. Alterations in BDNF in these neurons and the overall selective vulnerability of specific areas to degeneration are seen in not only AD, and normal aging, but also other

disorders of the brain, pointing to multiple roles for BDNF in particular. Low levels of BDNF have been observed in various neurological and neurodegenerative disorders such AD, Parkinson's disease (PD), amyotrophic lateral sclerosis (ALS), and Huntington's disease (HD).

BDNF and AD: In patients with AD, low BDNF levels have been reported in the dentate gyrus and in neurons that have neurofibrillary tangles, a hallmark of AD.[157,161–163] Several lines of evidence suggest that BDNF can have therapeutic benefits in AD. Genetic delivery of BDNF in animal models of AD, increased levels of BDNF in the entorhinal cortex and improved synaptic loss as well as learning and memory, thus highlighting the important roles of BDNF in AD.[164] The effect of BDNF on reversing symptomatic effects of neurodegeneration appear to be independent of amyloid beta (Aβ) peptide clearance. Levels of Aβ remain unchanged upon BDNF therapy.[164,165] Thus, BDNF could be exerting its effects by modulating functional synapses through promoting synapse formation and repair.

Alternatively, withdrawal of BDNF in cultured hippocampal neurons results in a substantial decrease in genes involved in synaptic function, vesicular trafficking, endosomal function and MAPK signaling.[166] These changes correlate with previously reported decreases in gene expression in AD and aging where levels of BDNF are significantly reduced.[167] A comparison of the gene classes that change in AD and in BDNF-deprived hippocampal neurons highlight similarities in genes involved in vesicular trafficking and synaptic function.[166,167]

BDNF and PD: BDNF is also important in the pathogenesis of PD. Reduced expression of BDNF has been reported in the substantia nigra pars compacta, a region that is selectively vulnerable to substantial neuronal loss in PD.[168,169] There are also isoform specific alterations in TrkB expression in PD. Levels of truncated TrkB are decreased in striatal axons, and increased in striatal soma and in substantia nigra pars compacta distal dendrites. In comparison, full-length TrkB is decreased in striatal neurites and substantia nigra pars compacta soma and dendrites, whereas it is increased in striatal somata and in substantia nigra pars compacta axons.[170] The localization of TrkB receptors is significant, since the signaling pathways can be enhanced through MAPK and PI3K/Akt enzymatic activities.[171]

BDNF and ALS: NT-based therapeutic approaches can prevent the progressive loss of motor neuron function that manifests muscle atrophy, weakness and spasticity in ALS. BDNF has been reported to slow the progression of motor neuron atrophy in an animal model of ALS.[172] Moreover, the TrkB agonist 7,8-dihydroxyflavone (7,8-DHF) improved motor neuron deficits in an animal model,[173] although efforts to replicate the positive effects of 7,8-DHF upon TrkB activity have not been successful.[174] An alternative approach is to use agonist monoclonal antibodies for TrkB.[174] The truncated form of TrkB, which is found

on motor neurons and glia, as well as other non-neuronal cells is involved in disease onset in an ALS animal model.[175,176]

BDNF and HD: BDNF plays a critical role in the pathogenesis of HD. Striatal medium spiny neurons are susceptible to degeneration in HD and depend on BDNF for their survival through TrkB signaling.[177] BDNF produced in cortical neurons and anterogradely transported to the striatum supports survival of medium spiny neurons.[178] The Huntingtin protein is thought to play a role in the transport and activity-dependent release of BDNF and a mutant Huntingtin can interfere with the transport and release of BDNF to the striatum.[179,180] Mutations in the Huntingtin protein greatly affect BDNF levels in striatal neurons. This is evident in studies that reported low expression of BDNF in mouse models of HD and in post-mortem human HD brains.[180–182]

1.5 Neurotrophin mimetics

Neurotrophins are known for their ability to regulate differentiation and neuronal survival of the nervous system through interactions with Trk receptors, while apoptotic pathways can also be initiated by interactions with the p75 death receptor. The balance between survival and death depends on the levels of expressed neurotrophins and receptors.[15] The low levels of neurotrophins in some neurodegenerative disorders, has led scientists to explore the potential benefits of increasing neurotrophin supply to treat neurodegenerative conditions.[183–185] However, delivering large proteins such as neurotrophins in the brain is very challenging because of the blood brain barrier (BBB). Additionally, targeting selectively only the affected neurons is very difficult due to crosstalk among multiple regions of the brain.[186] Therefore, administration of neurotrophins must be done directly in the CNS through invasive procedures, such as intracerebroventricular (ICV) injection,[187,188] which can also cause several side effects, such as neuropathic pain.[189–195]

A solution to the poor pharmacokinetic properties of neurotrophins can be the development of small organic molecules or peptides that can act as neurotrophic mimetics. These compounds could act as selective agonists of any of the Trk or p75 receptors, and thus avoid the occurrence of side effects. The small molecule **gambogic amide** has been identified as a TrkA agonist, able to attenuate apoptosis in primary hippocampal neurons.[196] Other TrkA agonists, such as **MT2** and **talaumidin**, can prevent Aβ-mediated neuronal death and promote neurite outgrowth in PC12 cells.[197,198] The selective, partial TrkA agonist **peptidomimetic D3** can bind to the TrkA-EC domain and afford survival of cholinergic neurons.[199]

Carvacrol (CARV) is another TrkA agonist that can induce neurite outgrowth.[200] A series of **dehydroepiandrosterone (DHEA)** derivatives are TrkA agonists,[201] with the C17-spiroepoxy derivative, **BNN-27**, being able to promote neuronal survival, without causing hyperalgia.[202] The C17-spirocyclopropyl DHEA derivatives, **ENT-A010** and **ENT-A013**, were recently proven potent TrkA agonists, with ENT-A013 being a selective agonist for TrkA.[203,204] Both compounds were able to initiate TrkA-related signaling pathways and promote cell survival in PC12 cells.[203,204]

A **peptide mimetic** of BDNF has been shown to activate TrkB and promote neuronal survival.[205] The dimeric dipeptide GSB-106 mimetic of the loop 4 of BDNF can activate TrkB and its signaling pathways, while it can promote survival in neuronal-like cells.[206,207] Small molecule mimetics of BDNF have exhibited TrkB specificity at nanomolar concentrations,[208] while compound **LM22A-4** can prevent neuronal death in *in vitro* models of AD, HD and PD.[131] The small molecule **deoxygedunin** is a selective TrkB agonist able to promote axon regeneration in neurons.[209] The flavonoid **7,8-dihydroxyflavone (7,8-DHF)** is another selective TrkB agonist with neuroprotective properties in PD and HD.[210,211] Compound **BNN-20** is a BDNF mimetic, which has anti-apoptotic, antioxidant and anti-inflammatory properties, through the activation of TrkB signaling pathways.[212]

Apart from Trk receptor agonists, some p75 modulators have also been developed. Compounds **LM22A-24** and **LM11A-31**, synthesized by Longo and Massa, have been shown to interact with p75 and protect against apoptosis in cells susceptible to proneurotrophin-induced death.[213] A p75 antagonist, **THX-B**, can inhibit NGF binding to p75 and prevent death of retinal ganglion cells.[214] Finally, compound **EVT901**, a p75 inhibitor, is able to reduce binding of proneurotrophins to p75 and block apoptosis and enhance neurite outgrowth *in vitro*.[215] The DHEA derivative **BNN-27** has also been proven to activate p75 signaling and rescue neurons from apoptosis.[216]

1.6 Computational approaches to study mechanistic effects of neurotrophins and their mimetics

The mechanism of activation of neurotrophin receptors remains elusive, despite the various sources of experimental information about the role and function of receptor sub-domains. Part of the difficulty has been the nature of these transmembrane receptors to be localized in membrane compartments, something that renders them hard to be studied with conventional structural biology methods. Additionally, the complexity of their signaling cascades and their involvement in both survival and

apoptotic pathways has made it difficult to elucidate fully their exact response to exogenous triggers. Similar is the case for existing neurotrophin mimetics since the mechanisms by which they activate or modulate the neurotrophin receptors are still poorly understood. In fact, even their type of action as direct agonists or indirect modulators of neurotrophin receptors can be a matter of debate.[217]

Computational methods are often used with the aim to shed light into processes that are hard to be examined experimentally due to spatiotemporal restrictions. Among the most common computational techniques, used for the study of biomolecules and their dynamical behavior, are Molecular Dynamics (MD) simulations. The TrkA-D5 domain was simulated in complex with peptide mimetics of the N-terminal loop of NGF, TrkA agonists, to investigate the stability of the complexes.[218] The simulations indicated specific residues in the peptides, which form salt bridges with the receptor, to be responsible for the binding.[218] Furthermore, atomistic simulations of the TrkA-D5/NGF complex accompanied by Poisson-Boltzmann electrostatic calculations showed there are water-mediated hydrogen bonds between NGF and receptor, which can affect the large-scale motions of the complex.[57]

In another study, MD simulations of the N-terminal region of the JM domain of TrkA with a membrane showed insertion of three residues into the membrane, in a conformation stabilized by interactions with PIP_2 lipids.[219] This insertion reduces the degrees of freedom of the JM domain and may play a role in receptor activation. MD simulations coupled with free energy calculations were also used for the understanding of the mechanism of action of allosteric TrkA-kinase inhibitors.[220] In another work, simulations were employed to explore potential ATP-binding sites on the TrkA-NGF and p75-NGF complexes.[221] The interaction between the TM domains of TrkA and p75 receptors has also been simulated, showing that binding of p75 to TrkA leaves exposed residues in TrkA that can participate in the active TrkA homodimer interface.[222] This type of interaction may explain the high-affinity binding sites of NGF in TrkA when p75 is present.

Atomistic MD simulations of the TrkB-TM helical homodimer revealed a stabilization of the dimer in 20% cholesterol concentration in the membrane, while it adopted suboptimal configurations in lower and higher cholesterol concentrations.[91] An MD simulation study from the same authors, revealed cholesterol binding to the TM helix of TrkB at a cholesterol-recognition amino acid consensus (CARC) motif in the sequence of the TM helix.[223] Cholesterol binding was verified experimentally with mutagenesis studies. [91] Additionally, docking of the antidepressant drug fluoxetine to the TM helices of TrkB indicated some residues as important for the binding, which were verified experimentally with point mutations.[91] Subsequent simulations of the TrkB-TM/fluoxetine complex in membranes showed

that fluoxetine can stabilize the TM dimer in an arrangement that is independent of the cholesterol concentration in the membrane environment.[91]

Apart from MD simulations, other computational techniques have also been applied to neurotrophins and their receptors. A recent study performed normal mode analysis in tandem with dynamical network analysis on the complex of TrkA-D5 with the NGF R221W mutant, showing that the mutation can impair the collective functional motions of the complex and weaken coupling of the NGF-TrkA interface, which may account for reduced signal transduction linked to nociception.[224] Density functional theory (DFT) calculations have indicated the binding of a zinc ion at the N-terminal of the D1-D3 segment of TrkA, while the free energy landscape of the motions of the TrkA-EC/NGF complex was reconstructed with metadynamics simulations.[225] Finally, docking calculations together with MD simulations were performed to probe the binding of DHEA and BNN-27 compounds in TrkA and p75 receptors.[202,216]

1.7 Scope of the thesis

Neurotrophin mimetics hold great promise in the treatment of neurodegenerative diseases as a non-invasive alternative to direct neurotrophin administration. An important step for the design of novel and potent neurotrophin mimetics is the understanding of the mechanism by which neurotrophins activate their receptors and elicit their function. Since the exact mechanism of activation of the receptors and the function of their sub-domains is still poorly understood, this thesis focuses on the elucidation of the conformational changes that take place upon receptor activation with the use of a variety of computational methods. This work was carried out within the EuroNeurotrophin ITN project.[226] Particularly, the TrkA and TrkB neurotrophin receptors have been investigated here. Specifically, in chapter 2, the theoretical background of the methods applied in the thesis will be introduced, while in chapters 3, 4, 5 and 6, the main work and results of the thesis will be presented. Chapter 3 focuses on the investigation of the effects of glycosylation in the EC domains of TrkA and TrkB receptors, while chapter 4 presents the dynamical behavior of the TM helical domains of TrkA and TrkB receptors in various lipid environments. Chapter 5 then presents the modeling of the full-length receptor structures and their simulation and behavior in a neuronal membrane model. Next, chapter 6 focuses on the mechanistic studies on neurotrophin mimetics. Conclusions about the mechanism of activation of these receptors and the mimetics will be finally discussed.

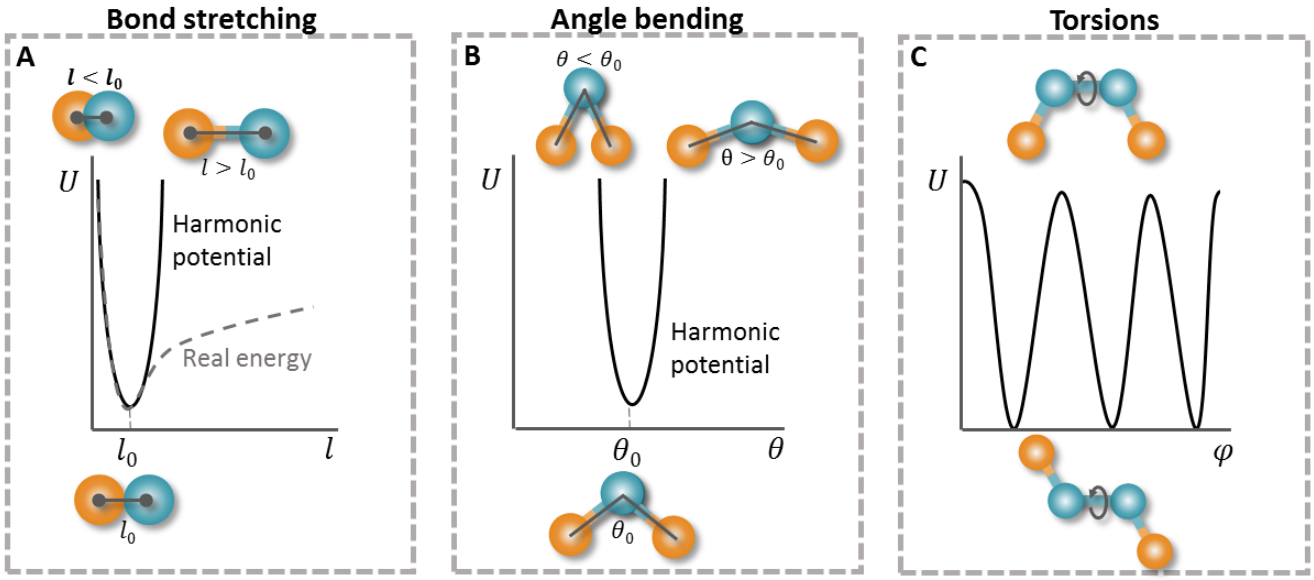
2 Theoretical introduction of computational methods

2.1 Molecular modeling of biomolecules

The term molecular modeling refers to the use of mathematical models to describe the behavior of molecules and molecular systems. Due to the complexity and the large number of calculations required to solve these mathematical models, molecular modeling is intrinsically linked to computer modeling. Molecules consist of electrons and nuclei. Quantum mechanics methods explicitly represent the electrons in a calculation and thus, they make possible the calculation of properties that depend on electronic distribution. However, many of the problems tackled by molecular modeling are too large to be considered by quantum mechanics. A solution to that is provided by molecular mechanics, which use force fields that describe the energy of a system as a function of the nuclear positions. The success of molecular mechanics is because of the validity of some assumptions. The first is the Born-Oppenheimer approximation, which enables the electronic and nuclear motions to be separated, because the much smaller mass of the electrons allows them to rapidly adjust to any nuclear position changes. Consequently, the energy of a molecule can be described as a function of the nuclear coordinates only. Next, molecular mechanics use a simple model of the interactions, such as the stretching of bonds, which can be quite accurately described by simple functions, such as Hooke's law. Another important aspect is the transferability of force fields, i.e. the ability of a set of parameters created for a small number of cases to be applied on a series of related molecules. In the same way, parameters developed for small molecules can be used for larger systems, such as polymers.[227–231]

2.2 Molecular mechanics force fields

Molecular mechanics force fields describe the energy of a system as a function of the nuclei positions and they consist of bonded and non-bonded terms to model intra- and inter-molecular forces. The bonded terms describe the interaction between pairs of bonded atoms, the angle formed by two consecutive bonds, the rotation around a bond, and the planarity of four atoms (Figure 2.1A-D). The non-bonded terms model the electrostatic and van der Waals interactions.



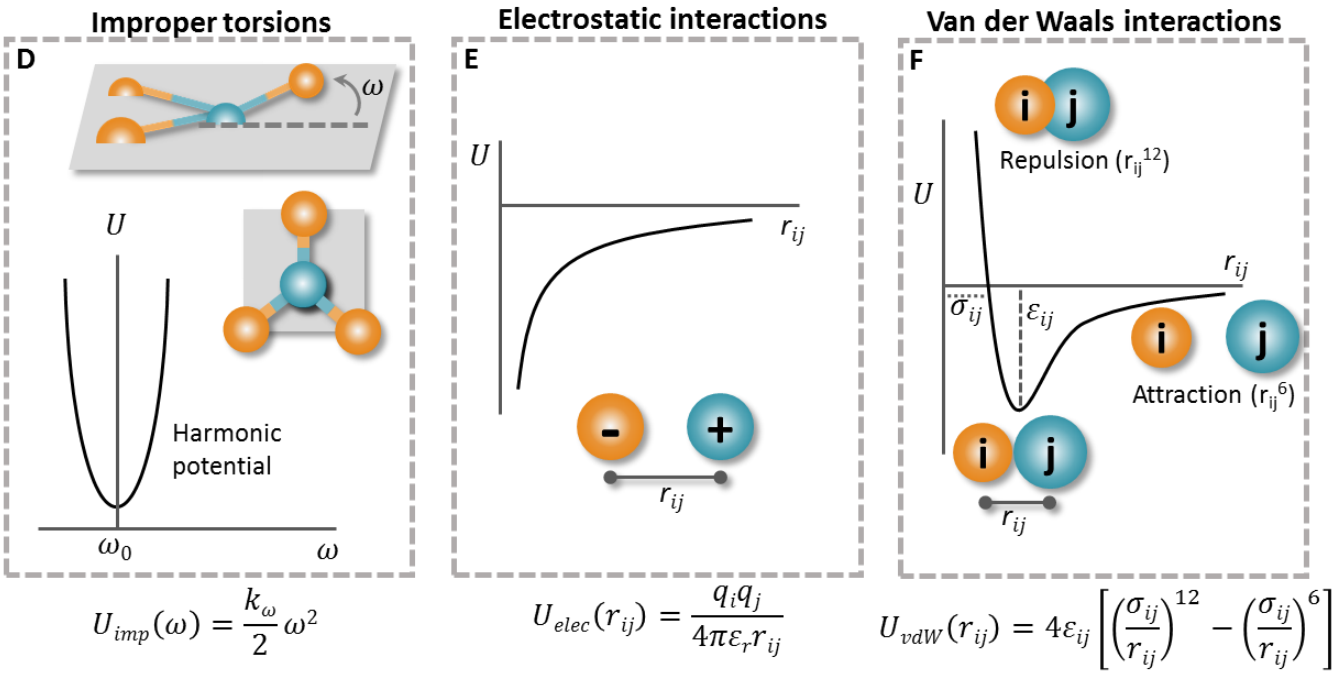


Figure 2.1: Force field terms describing bonded and non-bonded interactions. The bonded terms describe (A) the bond stretching, and the fluctuations of (B) bond angles, (C) torsional angles, (D) improper torsions. The non-bonded terms describe (E) the electrostatic and (F) the van der Waals interactions.

The potential energy curve of a bond stretching is that shown in Figure 2.1A with dashed line. An acceptable approximation to that is Hooke's law or harmonic potential:

$$U_{bond}(l) = \frac{k_l}{2}(l - l_0)^2 \qquad (2.1)$$

where k_l is the stretching force constant and l_0 is the reference bond length. As seen in Figure 2.1A, the Hooke's law is a reasonable approximation, as it mostly deviates from the real potential in high interatomic distances, and it is rare in molecular mechanics calculations for bonds to deviate significantly from their equilibrium values. The potential energy of angle bending is also described by a harmonic potential:

$$U_{ang}(\theta) = \frac{k_\theta}{2}(\theta - \theta_0)^2 \qquad (2.2)$$

where k_θ is the force constant and θ_0 is the reference angle (Figure 2.1B). For the rotation around a bond, dihedral torsional angles are defined by four subsequently bonded atoms and it is modeled by a cosine series expansion:

$$U_{tor}(\varphi) = \frac{V_n}{2}[1 + cos_{(n\varphi - \gamma)}] \qquad (2.3)$$

where V_n is the torsional barrier, n the multiplicity, i.e. the number of minimum points in the potential as the bond rotates through 360° and γ the phase factor, which determines where the torsion angle passes through its minimum value (Figure 2.1C). For planar molecules, out-of-plane bonding terms are used, which enforce that three atoms bonded with another atom in the middle are all in the same plane (Figure 2.1D). A common way to achieve this in a force fields, is to treat the four atoms as an "improper" torsion angle, by calculating the angle (ω) between a bond from the central atom and the plane defined by the central atom and the other two atoms (Figure 2.1D). Then the deviation of the out-of-plane coordinate can be modelled with a harmonic potential, centered at a reference angle ω_0=0:

$$U_{imp}(\omega) = \frac{k_\omega}{2}\omega^2 \qquad (2.4)$$

where k_ω is the force constant.

Considering the non-bonded terms of a force field, these are the electrostatic and van der Waals interactions. In molecular mechanics, the charge distribution in a molecule is represented by partial

atomic charges, assigned to the nuclear centers. The electrostatic interaction between point charges from different molecules is calculated using Coulomb's law:

$$U_{elec}(r_{ij}) = \frac{q_i q_j}{4\pi\varepsilon_r r_{ij}} \qquad (2.5)$$

where ε_r is the relative dielectric coefficient, q_i and q_j are the interacting partial charges and r_{ij} the distance between them (Figure 2.1E). Electrostatic interactions do not account for all of the non-bonded interactions in a system. Even in neutral atoms there are attractive and repulsive interactions, which occur when dipoles are induced by electron fluctuations around nuclei that are influenced by other atoms in proximity. These interactions are modeled by the Lennard-Jones potential (Figure 2.1F), which describes the van der Waals attractive and repulsive forces:

$$U_{vdW}(r_{ij}) = 4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] \qquad (2.6)$$

where σ_{ij} is the collision diameter (the separation for which the energy is zero) and ε_{ij} the well depth. While electrostatic interactions act on a long-range, van der Waals interactions mainly play a role at short distances.[229]

Force fields are usually parameterized to reproduce structural properties, like lipid order parameters for the case of lipid parameters. Bond stretching parameters can be created through fitting to values obtained by electron diffraction, which gives the mean distances between atoms averaged over the vibrational motion at room temperature. Angle and torsion parameters are usually developed with quantum mechanical calculations, which explicitly describe the electrons in a system. Partial atomic charges are calculated in order to reproduce electric moments, for which the exact geometry of the molecule must be known. In some cases, atomic charges can be chosen to reproduce thermodynamic properties calculated using molecular dynamics or Monte Carlo simulations. Quantum mechanics approaches can also be used for the estimation of partial charges. The van der Waals parameters are determined through molecular dynamics simulations and are optimized to reproduce a range of thermodynamic properties.

Most of the force fields used in molecular modeling describe interactions between particles with each particle corresponding to one atom. This can lead to a huge amount of atoms when big systems are of interest to be simulated. A solution to this has been offered with the use of **coarse-grained (CG) force fields**, which offer parameters for particles corresponding to two or more atoms. This can

significantly reduce the number of degrees of freedom of a big system, which can facilitate simulating the same system much faster or for a longer simulation time compared to all-atom force fields.

The most commonly used CG force field for biomolecular simulations is the Martini force field, which models four heavy atoms with one CG particle bead.[232,233] It has been extensively used for systems that comprise lipids, such as membranes, as well as for protein-protein interactions.[234] In this thesis, I have performed CG simulations with the Martini force field for the transmembrane (TM) helices of the neurotrophin receptors. The Martini force field has been successfully used in the past to elucidate the configurational landscape of various TM helical dimers.[235–238] Also, the Martini force field has been used with different types of lipid systems, such as membranes and micelles,[237,239] which is also a reason why it has been employed in the current thesis. Parameters in this force field have been developed to reproduce partitioning free energies between polar and apolar phases of a large number of chemical compounds.[232,233,240]

2.3 Energy minimization

Molecules can adopt various conformations and computationally this is represented by different atomic coordinates. Energy varies when coordinates change and this is referred to as the potential energy surface. Of particular interest are the atom arrangements (or configurations) that correspond to minimum points on the energy surface. These are stable states of the system and any movement away from the global minimum would give a configuration of higher energy. Computationally, these stable states can be identified with energy minimization algorithms. These algorithms minimize the potential energy as a function of the 3N Cartesian coordinates of N atoms in a system. Minima are identified using numerical methods, which gradually change the coordinates to produce configurations of lower and lower energy until a minimum is reached. The most commonly used energy minimization algorithms in molecular modeling are the Steepest Descent and the Conjugate Gradient methods. Both algorithms rely on derivatives of the force field energy function.

In the **Steepest Descent** method, starting from an initial configuration, the potential energy is re-evaluated at every step, after changing the atom positions. The gradient dictates in which direction to move the system and the energy progressively lowers. At every step the vector of the direction is perpendicular to the previous one. This makes the algorithm efficient in approaching the energy

minimum quite fast, even when starting far away from it. However, when reaching close to the minimum, the orthogonal changes in the direction make it slow to land to the final minimum point.

The **Conjugate Gradient** method is often used after the Steepest Descent algorithm, to reach the final minimum, since it is efficient when the system is already close to the minimum. This is achieved by retaining the information of the previous steps, to avoid going back and forth in two subsequent steps.[229]

2.4 Molecular dynamics simulations

Energy minimized configurations of a system can provide information about a particular minimum, or state of the system, usually the one closer to the initial configuration provided by the molecular modeler. However, many macroscopic properties of interest depend on the overall ensemble of configurations corresponding to a vast amount of points on the energy surface. Computationally, these configurations can be produced with simulation techniques, such as molecular dynamics (MD) and Monte Carlo (MC) simulations. In this thesis, I have performed MD simulations, and thus focus will be given to this simulation technique. In MD simulations, an ensemble of configurations is generated by starting from an arrangement of atoms, calculating the force acting on each atom by differentiating the energy function, or force field, and then calculating the acceleration from Newton's second law. Integration of the equations of motion allows the calculation of new atom positions, which eventually yields a trajectory that describes how the positions, velocities and accelerations evolve over time. The time step between trajectory snapshots is usually 1 or 2 femtoseconds, fs, i.e. 10^{-15} s.

There are several algorithms for integrating the equations of motions in MD simulations. The **Verlet** algorithm uses the positions and accelerations at time t, and the positions from the previous step, $\boldsymbol{r}(t-\delta t)$, to calculate the positions at the next time step $t+\delta t$. In this way, the Verlet algorithm lacks an explicit velocity term, which makes it hard to calculate the velocities. The **leap-frog** algorithm is a variation of Verlet, in which the velocities $\boldsymbol{v}(t+1/2\delta t)$ are calculated from the velocities at time $t-1/2\delta t$ and the accelerations at time t. The positions are then deduced from the velocities. This algorithm has advantage of explicitly calculating the velocities at each integration step, but it cannot provide the positions and velocities at the same step. The **velocity Verlet** algorithm gives the positions, velocities and accelerations at the same time step.

Setting up a system for MD simulations consists of several steps, including energy minimization, one or more equilibration steps and then the production simulation. Before everything, an initial

configuration of the system must be established, which is usually close to the desired state. An important consideration for the initial configuration of the system is the periodic boundary conditions, which is a replication of the system in all directions to give a periodic array. This allows for a small number of particles to be simulated, while the particles experience forces as if they were in bulk fluid. A cutoff distance between the most distant particle and the end of the simulation box must be set, so that the system does not interact with its periodic image, which could lead to artificial interactions.

Then, energy minimization takes place, which has already been described in the previous section. Next, the equilibration phase is performed, during which thermodynamic and structural properties of the system are monitored until stability is achieved. It is common to restrain different parts of the system in the various equilibration steps to ensure a gradual relaxation of the system. The first steps are run in the NVT statistical ensemble, in which the number of particles (N), the volume (V), and the temperature (T) of the system remain constant during the simulation. Equilibration at the NVT ensemble allows the heating of the system at the desirable temperature after the energy minimization phase which corresponds to T = 0 K. Next, equilibration at the NPT ensemble takes place, in which the number of particles (N), the pressure (P), and the temperature (T) of the system remain constant during the simulation. This ensemble represents well the physiological conditions of biological systems, and the conditions under which experimental measurements are made. Once the equilibration is completed, the production phase commences, during which properties of the system are calculated to understand the behavior of the system.[227–229]

The convergence of the simulations is assessed by calculation of several parameters. First of all, during the equilibration phase, the temperature and pressure must be monitored to ensure that they are stable. Then in the production, the root mean squared deviation (RMSD) is calculated, which shows the difference at the positions of the atoms during the simulation, $\boldsymbol{r}_i(t)$, from the atom positions in a reference state, which is often the initial structure, $\boldsymbol{r}_i(0)$.

$$RMSD = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\boldsymbol{r}_i(t) - \boldsymbol{r}_i(0)\right)^2}$$

Normally, the RMSD deviates from 0 at the beginning of the simulation and once it is stabilized this is an indication of convergence. Other parameters, like distances between atoms or biological domains, can also be used to asses convergence, as once they reach a plateau, this is usually an indication that

the system does not change anymore. Usually, a combination of parameters are needed to be stable in order to claim convergence.

2.5 Enhanced sampling methods

The goal of performing molecular dynamics simulations is in many cases the calculation of several thermodynamic and structural properties. These properties can be calculated from the simulation trajectories as time averages. However, short simulation times due to computation limitations can lead to incomplete sampling of all the biologically relevant configurations, which can result in inaccurate calculation of properties. A solution to increase the sampling has been offered by enhanced sampling methods, which can work through biasing specific reaction coordinates, restricting the system to sampling a specific region of space, or using conditional probabilities.

In the current thesis, I have used a method that works in the first way, i.e. through biasing of a few collective variables (CVs) of the system, called **metadynamics**.[241] In metadynamics, the sampling is enhanced along the slow the degrees of freedom of the system. This is achieved by adding a time-dependent external bias in the form of a Gaussian potential that counteracts the free energy surface. This allows diffusion of the system in the whole space of the CVs, which can also be used to assess convergence. From the external bias, one can directly estimate the free energy surface as a function of the selected CVs. At the same time, it is possible to calculate the free energy surface along other unbiased CVs by removing the bias, a technique called reweighting.

Many variants of metadynamics exist and in this thesis, I have performed well-tempered metadynamics simulations.[242] In this method, the Gaussian potentials that are deposited at regular time intervals decrease with time, which can be used as a metric to judge when the simulation is converged. In well-tempered metadynamics, a bias factor, γ, is introduced, which can be linked to the effective temperature of CV sampling $T + \Delta T$, with $\gamma = (T + \Delta T)/T$. The free energy surface can then be estimated directly from the reverse of the bias potential. Evaluation of the free energy surface evolution over time is also used to assess convergence.[243]

Metadynamics, combined with the coarse-grained Martini force field, has been successfully used in the past to study the full free energy landscape of the transmembrane helix dimerization of the Epidermal Growth Factor Receptor (EGFR).[238] In that study, the interhelical distance was used as a biasing CV for well-tempered metadynamics, which allowed the exploration of the various

transmembrane helical dimer arrangements and their relative binding free energy. Considering this successful application of metadynamics to a similar receptor tyrosine kinase system to the neurotrophin receptors, I employed metadynamics to study the dimerization of the transmembrane helical dimers of the neurotrophin receptors.

2.6 Homology modeling

For the initiation of MD simulations, a first configuration of the system is required, which is usually a protein model from experimental methods, like crystallography, Nuclear Magnetic Resonance (NMR) or cryo-electron microscopy. In cases where the experimental structure is not available for a protein, homology or comparative, modeling can be employed to derive a structural protein model from a homologous protein structure which has been experimentally determined. Nowadays, people increasingly use machine learning methods like AlphaFold,[244] which can predict the 3D structure of a protein from its sequence, even when no homologous structures exist experimentally. However, in this thesis, traditional homology modeling was performed, as homologous experimental structures existed and because the models presented here were created before the release of AlphaFold.

As a first step in homology modeling, a database search is performed to find potential homologous protein sequences with an experimentally available 3D structure. This is usually accomplished with a Basic Local Alignment Search Tool (BLAST) search, which constructs sequence alignments with the database sequences.[245] In this search, the query and database sequences are split in small fragments which are locally aligned and scored. The score shows the identity of the two sequences, which is calculated based on a matrix that encodes information on how frequently substitutions of one amino acid with another can occur.

The alignment together with the quality of the experimental structures, dictate which structure will be used as template for the homology modeling of the unknown structure. A sequence identity below 40% would denote that the query and the template might not share so many structurally common regions and therefore additional structural information should be explored. Once an adequate template is identified, the modeling of the protein takes place, with the template backbone coordinates being directly adopted as a first step to build the model. Side chains of the conserved residues are also often taken from the template, while non-conserved residues are attached to the backbone and then a conformational sampling can take place followed by energy minimization.

In this thesis, I have used the MODELLER software[246] for homology modeling, which creates the 3D protein structures with satisfaction of spatial restraints. These restraints can be derived from a number of different sources, including related protein structures (comparative modeling), NMR experiments (NMR refinement), rules of secondary structure packing (combinatorial modeling), cross-linking experiments, fluorescence spectroscopy, image reconstruction in electron microscopy, site-directed mutagenesis, intuition, residue–residue and atom–atom potentials of mean force. The restraints can be enforced on distances, angles, dihedral angles, pairs of dihedral angles and some other spatial features defined by atoms or pseudo atoms. The output 3D model is obtained by optimization of a molecular probability density function (pdf). The molecular pdf for comparative modeling is optimized with the variable target function procedure in Cartesian space that employs methods of conjugate gradients and molecular dynamics with simulated annealing.

2.7 Molecular docking

An integral part of molecular modeling is predicting the bound state of a small molecule ligand to a protein target. The most common technique for this task is molecular docking, which searches for a conformation and orientation of the ligand inside the protein binding site, with a low binding free energy. This splits the problem in two tasks: the conformational sampling and the scoring. In the conformational sampling step, many ligand binding poses, i.e. many orientations and conformations of the ligand bound to the protein, are generated either by building the ligand step by step in the protein pocket or by exploring possible arrangements of the full ligand structure inside the protein cavity. This task can be followed by energy minimization and further relaxation of the system under the influence of a molecular mechanics force field. The generated poses are then evaluated for their stability and favorable interactions with a scoring function.

In this thesis, I have used the Glide tool of the Schrodinger Suite[247–250] for docking of the neurotrophin mimetics to the receptors. The default protocol of Glide performs docking by treating the receptor as rigid and the ligand flexible, i.e. rotations around bonds can take place. The ligand can also assume three translational (in the x, y, z directions) and three rotational external degrees of freedom apart from the internal ones. Glide applies a series of hierarchical filters for the pose generation. In this approach, the receptor Coulombic and van der Waals potentials are represented on a grid, while the conformational search of the ligand is performed separately for the core region and the rotamer groups of the ligand.

The shape and size of the ligand are compared with that of the pocket, the properties of which are part of the generated grid. The initial position of the ligand is evaluated and scored with the OPLS-AA force field.[251] Then, an energy minimization with the OPLS4 force field[252] and a Monte Carlo sampling determine the torsional energy minima. Finally, the Glide docking score evaluates the final poses. The scoring function of Glide consists of several terms to describe the lipophilic interactions, the hydrogen bonds, the metal-ligand contacts, the entropic penalty due to loss of rotational flexibility, as well as the Coulomb and van der Waals interactions. Finally, desolvation penalties are included as well as ligand exposure to the solvent.

3 Glycosylation effects on the extracellular receptor dynamics

3.1 Introduction to the glycosylation of the receptors

Glycosylation of cell membrane receptors is an integral part of their structure, which is usually not taken into account due to difficulties in identifying experimentally the type and positions of glycosylation in proteins. Also, glycosylation is usually missing in available crystal structures, because glycans tend to decrease the X-ray diffraction quality of the crystals.[253,254] Molecular dynamics (MD) simulations can be used to model the full glycosylated state of proteins and probe the effects of the glycans on protein dynamics. In fact, computational studies are increasingly considering glycosylation in their protein models, by using mass spectroscopy data about the type and position of the glycans.[255–258]

As has already been introduced in the 1st chapter, the glycosylation of the TrkA receptor is necessary for receptor localization to the plasma membrane and ligand-dependent activation.[50] When the receptor is not glycosylated, it is trapped intracellularly and it is constitutively active. Thus, it cannot trigger the Ras/Raf/MAP kinase cascade, which initiates in the plasma membrane, and consequently promote neuronal survival.[50] Therefore, in the current chapter of the thesis, I have modeled and simulated the glycosylated extracellular (EC) domains of the TrkA and TrkB receptor homodimers, bound to their native neurotrophins (NTs) NGF and NT-4/5, respectively. The main goal is to identify the effects of glycosylation on the conformations of the EC domains of the receptors and their interaction with the NTs.

3.2 Methodology

Protein modeling

For the modeling of the TrkA receptor with NGF, I used the crystal structure of the complex of the TrkA full-EC domain (residues 33-382) homodimer with the mature NGF homodimer (PDB ID: 2IFG).[28] The structure of NGF had the following missing residues: Ser 122, Pro 182 – Ser 187, Ala 237 – Ala 241 (residue numbers correspond to the UniProt sequence of human NGF, UniProt code: P01138), which I

built with the AutoModel class of MODELLER v.9.23,[246] using the default parameters. The linker corresponding to residues Pro 182 – Ser 187 has a conserved conformation in existing crystal structures and therefore I modeled it using the structure of human NGF from PDB ID: 5JZ7[259] as a template. The EC domain of TrkA has Ala 33 – Pro 35 missing (UniProt code of human TrkA-II isoform: P04629-1) and thus I built these residues with AutoModel.

For the modeling of the TrkB receptor with NT-4/5, I modeled the full-EC domain (residues 32-383) bound to the mature NT-4/5. I used the crystal structure of the complex of the TrkB immunoglobulin-like domain 5 (D5) homodimer with the mature NT-4/5 homodimer (PDB ID: 1HCF)[260] together with the crystal structure of the complex of the TrkA full-EC domain homodimer with the mature NGF homodimer (PDB ID: 2IFG) for the modeling of the D1-D4 domains. The first three residues of the TrkB-D5 are cloning artefacts and thus I mutated them to the wild-type sequence. I modeled the rest of the D1-D4 domain residues with AutoModel using the 2IFG PDB structure as a template. The structure of NT-4/5 has the following missing residues: Asn 145 – Gly 150 and Gly 207 – Ala 210 in chain A and Gly 81 – Glu 84, Asn 145, Arg 209 and Ala 210 in chain B (residue numbers correspond to the UniProt sequence of human NT-4/5, UniProt code: P34130). I modeled the missing residues with the AutoModel class of MODELLER v.9.23.[246] Additionally, I built the linker corresponding to residues Asn 145 – Gly 150 using the structure of human NT-4/5 from PDB ID: 1B98[23] as a template.

For each receptor, I modeled three different systems; one with the receptor EC-homodimer bound to neurotrophin, one with one receptor EC-monomer bound to neurotrophin and one with the receptor EC-monomer alone. I modelled all systems with and without glycans.

Glycosylation modeling

The TrkA and TrkB receptors have N-linked glycans in their EC segments.[48,50] Specifically, TrkA has 13 potential N-glycosylation sites, based on the Asn-X-Ser/Thr sequon, out of which at least 9 are actually glycosylated.[50] Six of these 9 sites are known from the crystal structure of the full-EC TrkA domain bound to NGF.[28] These are the asparagine residues at positions 95, 121, 188, 262, 281 and 358 (Table 3.1). For the case of TrkB, out of the 12 potential N-glycosylation sites, 10 are glycosylated and their exact positions are known; Asn 67, 95, 121, 178, 205, 241, 254, 280, 338, 412 (Table 3.1).[44,48] All glycosylation positions in TrkB are found in the EC domains, with the only exception being Asn 412, which is located in the linker between the D5 domain and the transmembrane (TM) helix.

In order to decide the glycan positions in TrkA that were not already know experimentally, I compared the possible glycosylation positions in TrkA structure with the known positions in TrkB. The equivalent positions of TrkA and TrkB glycosylation sites are also shown in Table 3.1. I decided to model glycans in four additional positions, Asn 67, 202, 253 and 401 (residues in bold font in Table 3.1), to get a total of ten glycosylation sites, as for TrkB.

Table 3.1: N-glycosylation positions on TrkA and TrkB receptors. For TrkB, the exact glycosylation positions are known experimentally.[44,48] For TrkA, the glycosylation positions that are known from the PDB ID: 2IFG[28] crystal structure are shown with normal font, while the ones that I added based on comparison with TrkB positions are shown in bold letters.

TrkA	TrkB
Asn 67	Asn 67
Asn 95	Asn 95
Asn 121	Asn 121
Asn 188	Asn 178
Asn 202	Asn 205
-	Asn 241
Asn 253	Asn 254
Asn 262	-
Asn 281	Asn 280
-	Asn 338
Asn 358	-
Asn 401	Asn 412

The next step was to decide the type of glycans that would be covalently attached in the aforementioned positions. Since no experimental information about the type of glycans is available for both TrkA and TrkB, except for maybe the indication of sialic acid presence in the glycans of TrkA,[54,55] I decided to model the glycan root, $Man_3GlcNAc_2$, on all positions of TrkA and TrkB that were present in the EC domains, i.e. all glycans in Table 3.1, except for Asn 401 in TrkA and Asn 412 in TrkB, because these are located in the linker after the EC domain, which I did not model in this case. Glycosylation of all ten positions of TrkA and TrkB, will be presented in chapter 5, where I model the full-length receptors. Also, two cases of longer oligomannose and complex glycans, which I modeled in EC-TM systems, will be presented in chapter 5 as well.

I inserted N-glycans of chemical type $Man_3GlcNAc_2$ into the protein models with CHARMM-GUI.[261] Models of the TrkA/B-EC homodimer with the NGF/ NT-4/5 homodimer bound, the TrkA/B-EC monomer

with the NGF/ NT-4/5 homodimer bound, and the TrkA/B-EC monomer were built with and without glycans (Figure 3.1).

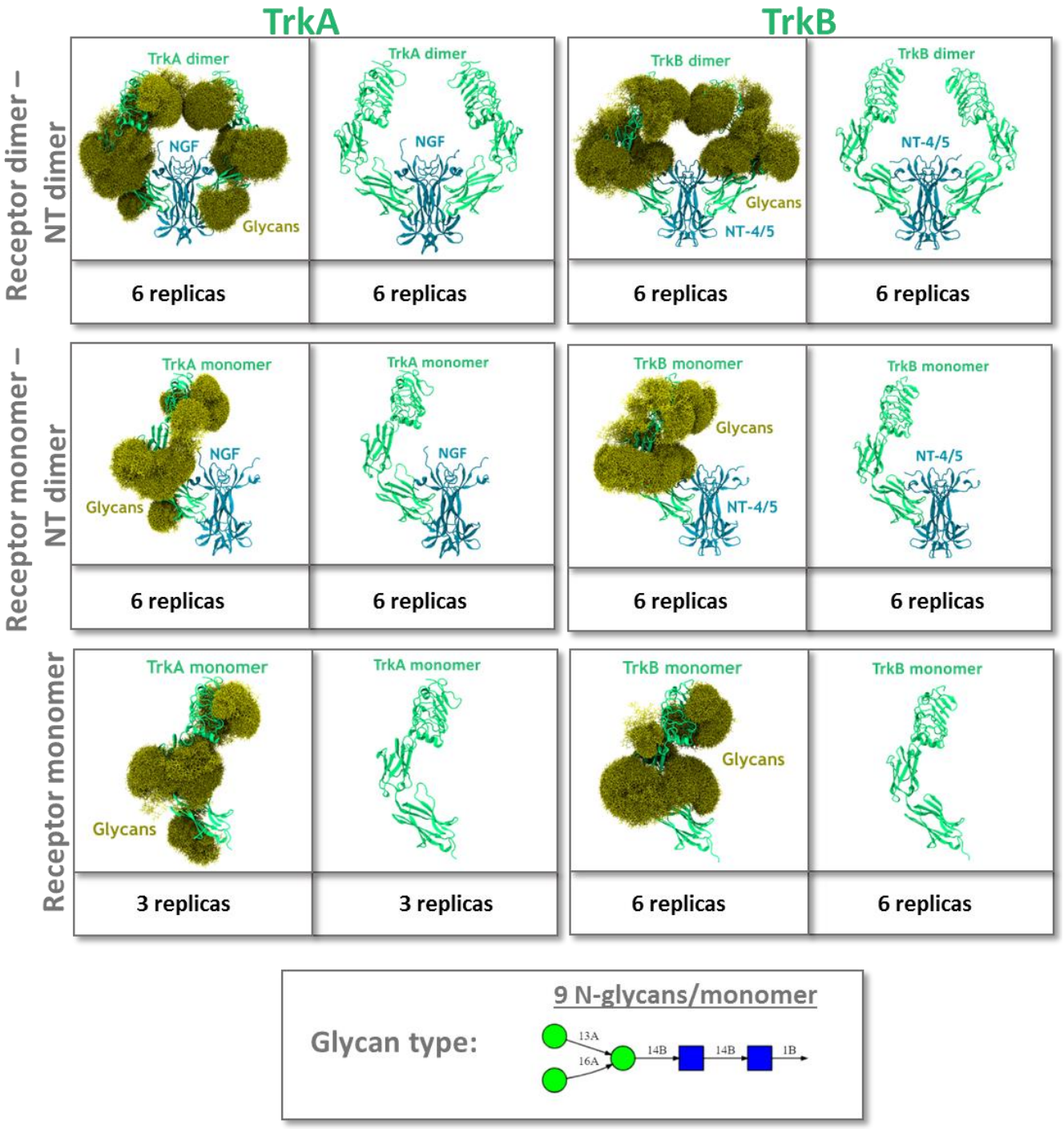


Figure 3.1: Summary of modeled glycosylated and unglycosylated systems. The systems that I modeled to test glycosylation effects on the EC domains of TrkA and TrkB receptors were: the TrkA-EC homodimer with the NGF homodimer bound, the TrkA-EC monomer with the NGF homodimer bound, the TrkA-EC monomer, the TrkB-EC homodimer with the NT-4/5 homodimer bound, the TrkB-EC monomer with the

NT-4/5 homodimer bound, the TrkB-EC monomer with and without glycans. All systems are shown in cartoon representation with the receptors in green and the NTs in blue. The systems with glycans present are shown with all the glycan conformations that were sampled during the simulations from all frames. The glycans are shown in yellow wire representation. The numbers of replica simulations that I ran for each system are also indicated. The last box at the bottom of the figure shows the type of the glycan, $Man_3GlcNAc_2$, which was attached to all nine glycosylation positions of the receptors. This is in Symbol Nomenclature for Glycans (SNFG) representation with the blue boxes denoting the N-acetyl-glucosamine and the green circles the mannose monosaccharides.

Molecular dynamics (MD) simulations

I ran all simulations with the GROMACS v.2020 MD engine,[262] using the CHARMM36m force field.[263] I solvated each system with the *gmx solvate* tool of GROMACS. Then I neutralized the charge of the box by adding Na^+ and Cl^- ions at 150 mM concentration. I started each simulation with a steepest-descent energy minimization and this was followed by one NVT and one NPT equilibration step, in which I removed gradually the protein atom restraints. Specifically, I ran the first NVT equilibration at 310 K for 500 ps with the protein backbone restrained with a force constant of 400 kJ/(mol·nm^2) and the side chains restrained with a force constant of 40 kJ/(mol·nm^2). I ran the subsequent NPT equilibration at 310 K and 1 atm for 500 ps with a force constant on the backbone atoms of 200 kJ/(mol·nm^2) and on the side chains of 20 kJ/(mol·nm^2). I used the Nosé-Hoover thermostat[264,265] and the Parrinello-Rahman barostat.[266,267] I used a coupling constant of 5 ps to maintain isotropic pressure coupling with a compressibility of $4.5 \cdot 10^{-5}$. I applied a time step of 1 fs in the first equilibration step and then I switched it to 2 fs for the second step. I treated the non-bonded electrostatic interactions with PME[268,269] with a cutoff distance at 1.2 nm. Non-bonded van der Waals interactions were calculated with a cutoff of 1.2 nm and a switching distance of 1.0 nm. The LINCS algorithm[270] was employed to fix the length of all hydrogen-containing bonds. Subsequently, I performed production runs of 500 ns for each replica with the protein completely unrestrained. For the production runs, I used the same simulation parameters as for the last equilibration step. I ran six replica simulations for each system, each starting with different initial velocities, except for the TrkA-EC monomer system, for which I ran only 3 replicas (Table 3.2).

Table 3.2: Summary of systems that I simulated with or without glycans.

System No	Protein	Glycosylation	Replicas	Time/Replica (ns)
1	TrkA EC-dimer, NGF dimer	+	6	500
2		-	6	500
3	TrkA EC-monomer, NGF dimer	+	6	500
4		-	6	500
5	TrkA EC-monomer	+	3	500
6		-	3	500
7	TrkB EC-dimer, NT-4/5 dimer	+	6	500
8		-	6	500
9	TrkB EC-monomer, NT-4/5 dimer	+	6	500
10		-	6	500
11	TrkB EC-monomer	+	6	500
12		-	6	500

Analysis of simulations

I post-processed all simulations with the *gmx trjconv* tool of GROMACS to make molecules whole, remove periodic boundary conditions (pbc) jumps and center the proteins in the simulation box. I saved frames every 100 ps. I also aligned the systems to the whole protein backbone throughout the trajectories. I performed the graphical rendering of protein structures from the trajectories with VMD 1.9.3.[271]

Root Mean Square Deviation (RMSD)

I calculated the Root Mean Square Deviation (RMSD) of the protein backbone for all systems. Before the RMSD calculation, I aligned the proteins to their backbone throughout the simulation and then I calculated the RMSD of the backbone during the simulations with the MDAnalysis python library.[272,273] I calculated the RMSD in all frames of the trajectories and visualized the RMSD time evolution with the Matplotlib python library.[274] The distribution densities of the RMSD values were calculated and visualized with the Seaborn python library.[275] Additionally to the whole protein RMSD, I calculated the internal RMSD of all protein subdomains, by aligning the trajectory to the backbone of each subdomain before the RMSD calculation of that particular domain.

Root Mean Square Fluctuation (RMSF)

I calculated the Root Mean Square Fluctuation (RMSF) of the protein residues by aligning first the protein C_α atoms to an average structure from the simulations. I calculated the average structure with the align.AverageStructure class of MDAnalysis, which aligns the trajectory to the first frame and then averages the coordinates. Then I calculated the RMSF of the C_α atoms of each residue using MDAnalysis. Visualization of the results was performed with Matplotlib. I also calculated the RMSF of each glycan by aligning first each EC protein monomer to the first frame and then calculating the RMSF of each sugar in the glycans. Next I visualized the RMSF values of all sugars of each glycan in box plots with Seaborn.

Principal component analysis (PCA)

I performed principal component analysis (PCA) to identify the intrinsic motions of the D13 (D1-D3) domain of the TrkA and TrkB receptors, since during visualization of the simulations, I observed mostly hinge motions between the D13 and D4 domains. For this purpose, I aligned each EC receptor monomer to the D4-D5 domains, which have a stable conformation, in order to retain only the D13 domain motion with respect to the D4-D5 domains. After the alignment, I performed PCA in the D13 coordinates using the built-in class in MDAnalysis. Plotting of the first and second principal component values was performed with Seaborn.

I also performed PCA in the phi and psi dihedral angles (dPCA) of each residue in the linker between the D13 and D4 domains, because the different conformations in this linker are responsible for the hinge motions between the two domains. I calculated the dihedrals during the trajectories with MDAnalysis, and then I performed PCA with the scikit-learn python library,[276] because the MDAnalysis PCA class performs only PCA on coordinates. As with the standard PCA, I plotted the first and second principal component values with Seaborn.

Solvent accessible surface area

I performed solvent accessible surface area (SASA) calculations with the *measure sasa* tool of VMD 1.9.3.[271] I calculated the sasa of the proteins (both receptor and neurotrophin) with and without taking into account the glycans every 100 frames, and by subtracting these two, I calculated the area shielded by the glycans. I used different probe radii of 1.4, 2, 3, 4, 5, 6, 7, 8, 9 and 10 Å to emulate the interaction with molecules of different sizes, from water to small-molecules and proteins.

Contact area between D5 – NT

I calculated the contact area between the D5 domain of TrkA and TrkB receptors with the NTs NGF and NT-4/5, respectively, using the *measure sasa* tool of VMD 1.9.3.[271] Specifically, I calculated the solvent accessible surface area (SASA) of the D5 domain of each receptor twice; with and without the presence of the NT. For this I used a probe radius of 1.4 Å. Then, I subtracted the two areas to retrieve the contact area of the D5 with the NT (Figure 3.2). In the simulations with the glycans, I performed this calculation with and without the glycans as part of the D5, to observe any participation of the glycans in the contact area between D5 and NT. I repeated these calculations for both D5 monomers, in the systems where EC-receptor dimers were present.

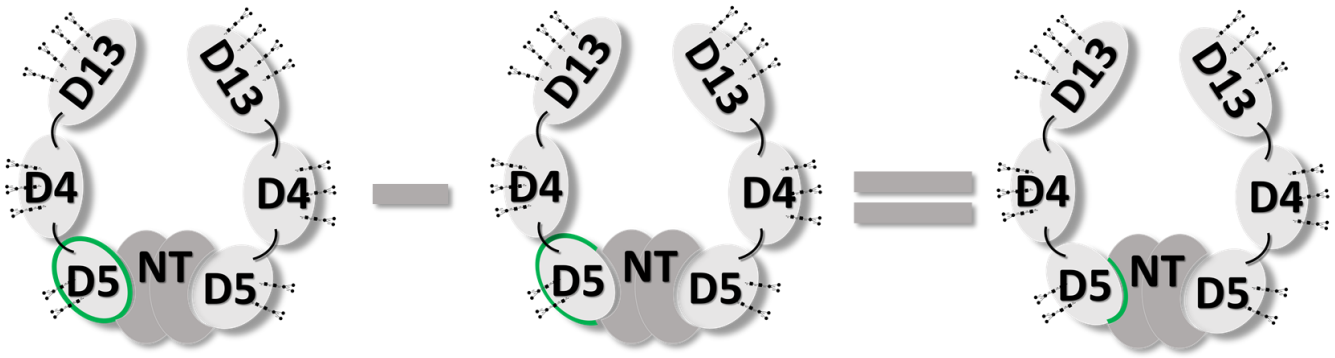


Figure 3.2: Schematic representation of the solvent accessible surface area (SASA) calculations that I performed to estimate the contact area of the D5 domain with the NT. Specifically, I calculated the SASA of the D5 domain without and with the NT present and the difference of these two values gives the contact area of D5-NT.

Radius of gyration

I calculated the radius of gyration of the backbone of each receptor monomer with MDAnalysis throughout the trajectories. Then I calculated and visualized the radius of gyration distribution densities for the glycosylated and non-glycosylated systems with Seaborn.

Distances and angles between domains

I calculated the distance between the centers of geometry (COG) of the two monomers of the D13 domain, the COG of two monomers of the D4 domain and the COG of the two monomers of the D5 domain, as well as the COG of the D13-D5 domains (when both monomers were present in the systems,

i.e. in the dimer receptor systems) with MDAnalysis and the numpy.linalg.norm function of the Numpy library.[277] Using the same python libraries, I calculated the angles that are formed between the centers of geometry of the D13-NT-D13, D4-NT-D4 and D5-NT-D5 domains. Both the distances and the angles are calculated every 100 frames. Then, I post-processed the data with the Pandas library[278,279] and I visualized the results with violin plots in Seaborn.

Hydrogen bond analysis

I calculated the hydrogen bonds (H-bonds) between each glycan and the protein residues with the HydrogenBondAnalysis class of MDAnalysis. I calculated the H-bonds for all frames of the trajectory. Then, I post-processed the results with Pandas and Numpy to make the data suitable for visualization and finally, I plotted with Matplotlib the percentage of the frames in which each glycan forms H-bonds with specific residues. For the plotting, I kept only the glycan-residue pairs which form H-bonds for more than 10% of the frames.

Dihedral angles (χ1) calculation

I calculated the χ1 dihedral angles of the Asparagine residues that get glycosylated in both the glycosylated and non-glycosylated systems. I calculated the χ1 angles of each Asn with MDAnalaysis for all frames of all replica simulations. Then, I plotted the distributions of the χ1 values of the Asn residues in box plots with Seaborn.

3.3 Results

In order to study the effects of glycosylation in the extracellular (EC) domains of the TrkA and TrkB receptors, I modeled and simulated them with and without glycans present and compared the differences between the simulations. Specifically, I performed atomistic unbiased Molecular Dynamics (MD) simulations. I also simulated different oligomerization states of the protein; receptor dimers bound to NT dimers, receptor monomers bound to NT dimers and receptor monomers alone. Several different analyses were conducted and the results are presented here.

The first step was to compare the glycosylation of TrkA and TrkB in their different simulated states. For this, I visualized the several conformations of the glycans from all snapshots that were sampled during the simulations and displayed then on top of the initial models, which were based on available crystal structures. This visualization is presented in Figure 3.3. Comparison of the glycosylation of TrkA

and TrkB receptors showed that TrkA has glycans in all the length of its EC domain, while the glycans in TrkB are mostly accumulated in its N-terminal region. The most important difference that occurs from this visualization is that TrkB does not have a glycan shielding of its D5 domain, as in TrkA where the glycan at residue Asn 358 shields the solvent exposed area of the D5 domain.

Overall, several different glycan conformations were sampled during the simulations, as revealed by the cloud of glycans displayed in Figure 3.3. Also, the glycan conformational sampling does not seem to be influenced by the oligomerization state of the proteins, since no important differences of the glycan clouds are observed among the several TrkA and TrkB systems. After this initial visualization of the glycosylated systems, quantitative analysis of the simulations was performed.

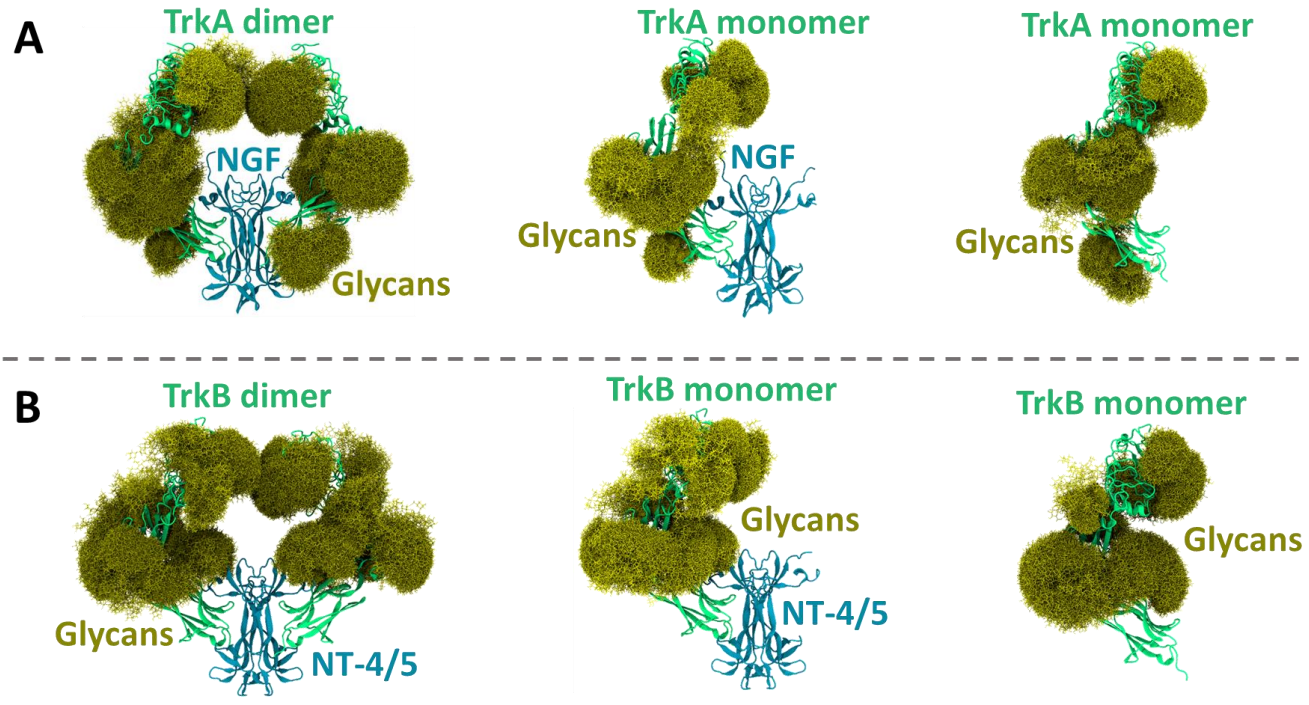


Figure 3.3: Comparison of the several (A) TrkA and (B) TrkB systems with all the glycan conformations sampled during the simulations. The proteins are shown in cartoon green and blue representation, while glycans from the different frames are shown in yellow wire representation. I performed alignment to each subdomain before the visualization of its glycans.

Initially, I performed root mean square deviation (RMSD) calculations compared to the initial protein models for all the different TrkA and TrkB systems (Figure 3.4 and Figure 3.5). The results show mostly similar protein RMSD values for the glycosylated and non-glycosylated TrkA and TrkB systems, with the "TrkA EC monomer-NT" and "TrkA EC monomer" systems showing slightly higher RMSD values for the non-glycosylated states. In contrast to this, the "TrkB EC monomer-NT" system displayed smaller RMSD values in its non-glycosylated state, as revealed from the distribution density peaks. Apart from these small differences between glycosylated and non-glycosylated states, all systems display large RMSD values during the simulations, in the range of 10-30 Å. This indicates large deviations from the initial models, which were created based on crystal structures.

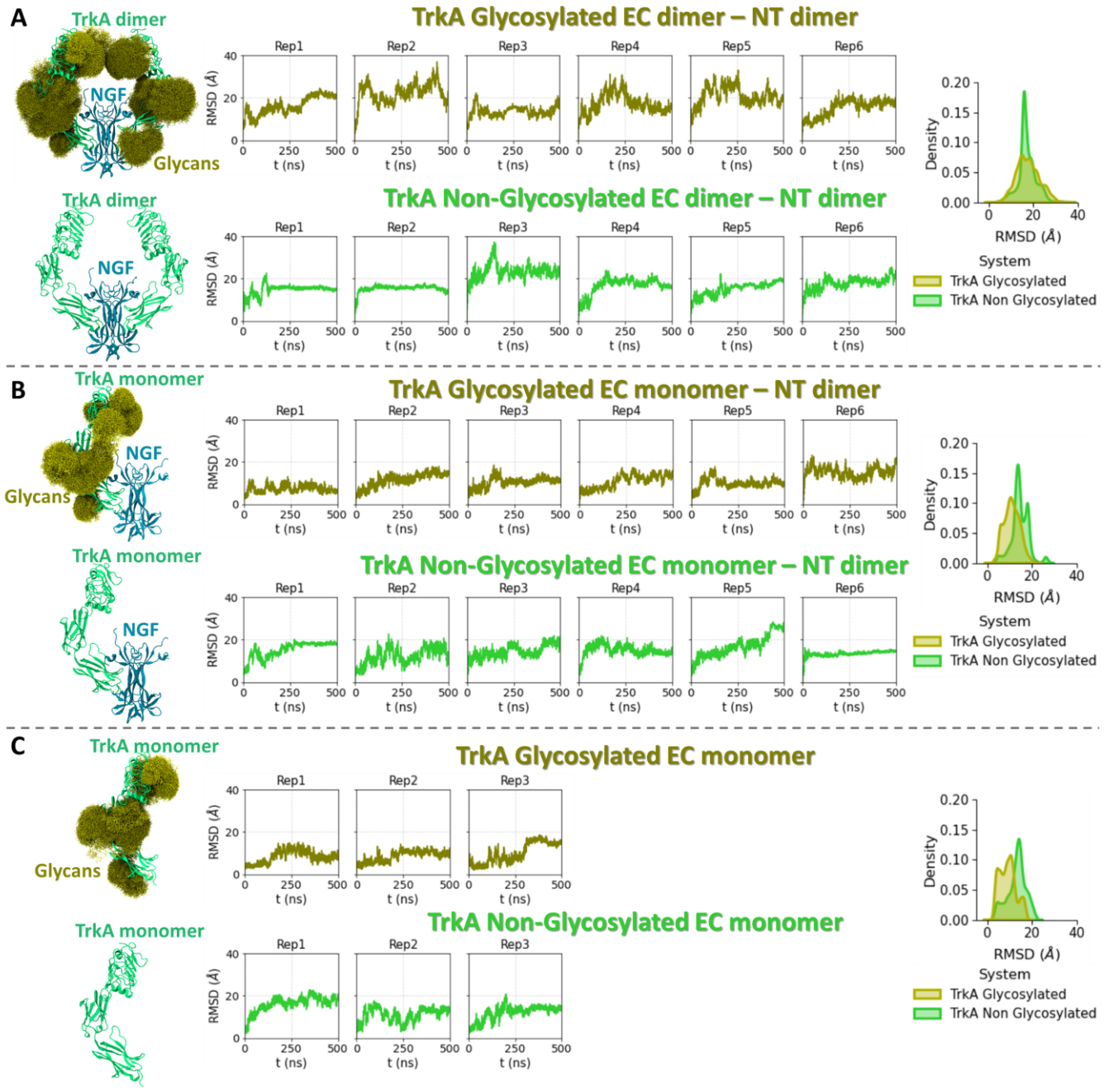


Figure 3.4: RMSD calculations of the glycosylated and non-glycosylated TrkA systems. (A), (B) and (C) show the RMSD for the TrkA-EC homodimer with the NGF homodimer bound, the TrkA-EC monomer with the NGF homodimer bound, the TrkA-EC monomer systems, respectively, with and without glycans. For each system the time-dependence of the RMSD of the protein backbone compared to the first frame (which corresponds to the crystal structure) is shown for each replica, as well as the distributions of

RMSD values across all replicas and all frames of the production runs for the glycosylated and non-glycosylated systems (right plots in A, B and C).

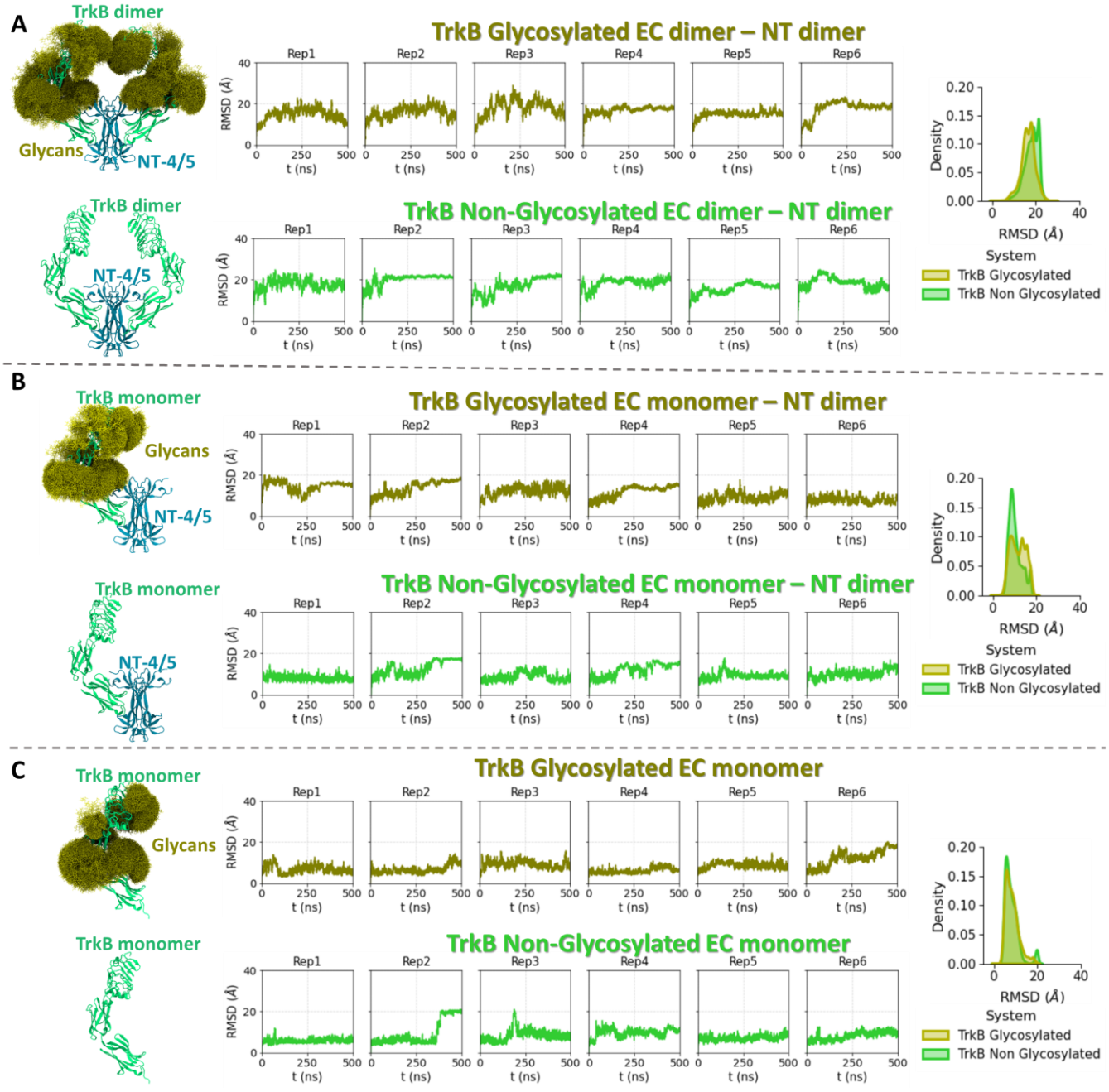


Figure 3.5: RMSD calculations of the glycosylated and non-glycosylated TrkB systems. (A), (B) and (C) show the RMSD for the TrkB-EC homodimer with the NT-4/5 homodimer bound, the TrkB-EC monomer with the NT-4/5 homodimer bound, the TrkB-EC monomer systems, respectively, with and without glycans. For each system the time-dependence of the RMSD of the protein backbone compared to the

first frame (which corresponds to the crystal structure) is shown for each replica, as well as the distributions of RMSD values across all replicas and all frames of the production runs for the glycosylated and non-glycosylated systems (right plots in A, B and C).

In order to examine the origin of these large RMSD values, I calculated the RMSD values for the internal motions of each subdomain of the receptors, by first aligning to the initial frame of each subdomain and calculating afterwards the RMSD (Figure 3.6 and Figure 3.7). The results show much lower RMSD values than the global RMSD calculation; less than 3 Å for most of the domains in different simulations, showing that the domain structures are preserved during the simulations.

In some rare cases higher RMSD values occurred, such as for the D13-A domain in Replica 1 of the "TrkA EC dimer-NT glycosylated" and Replicas 1, 4, 6 of the "TrkA EC dimer-NT non-glycosylated" systems. Especially high RMSD values, of ~10 Å, were observed in Replica 2 of the "TrkA EC monomer-NT glycosylated" system. A closer look at the sampled conformations of these domains was taken to understand the reason for the high RMSD values.

In Figure A.1 and Figure A.2, a comparison between the initial conformations of the domains and the conformations at a time step of high RMSD value are shown in superposition for the "TrkA EC dimer-NT" and "TrkA EC monomer-NT" systems. Most of the high values occur in the D13 domain, which seems to be due to changes in its N-terminal region. The D13 domain has several disordered regions and so does its N-terminal part, which seems to undergo different rearrangements during the simulations. This is in agreement with the high b-factors of this N-terminal region in the crystal structure of the TrkA-EC bound to NGF (PDB ID: 2IFG)[28] as shown in Figure A.3. In Replica 2 of the "TrkA EC monomer-NT glycosylated" system, a large part of the N-terminal region, containing also two strands of the beta sheet, has been unfolded (Figure A.2). Apart from the D13 domain, also the D4 domain seems to sustain some conformational changes, especially in the linker between the D and E strands, leading to an RMSD of 10 Å. Overall, most of the high RMSD values occur for the non-glycosylated systems rather than the glycosylated ones. In general though, the domain RMSD values are much lower than those of the global RMSD, indicating that the high RMSD values originate in the non-domain regions, and more specifically in the linker (or hinge) regions between the subdomains.

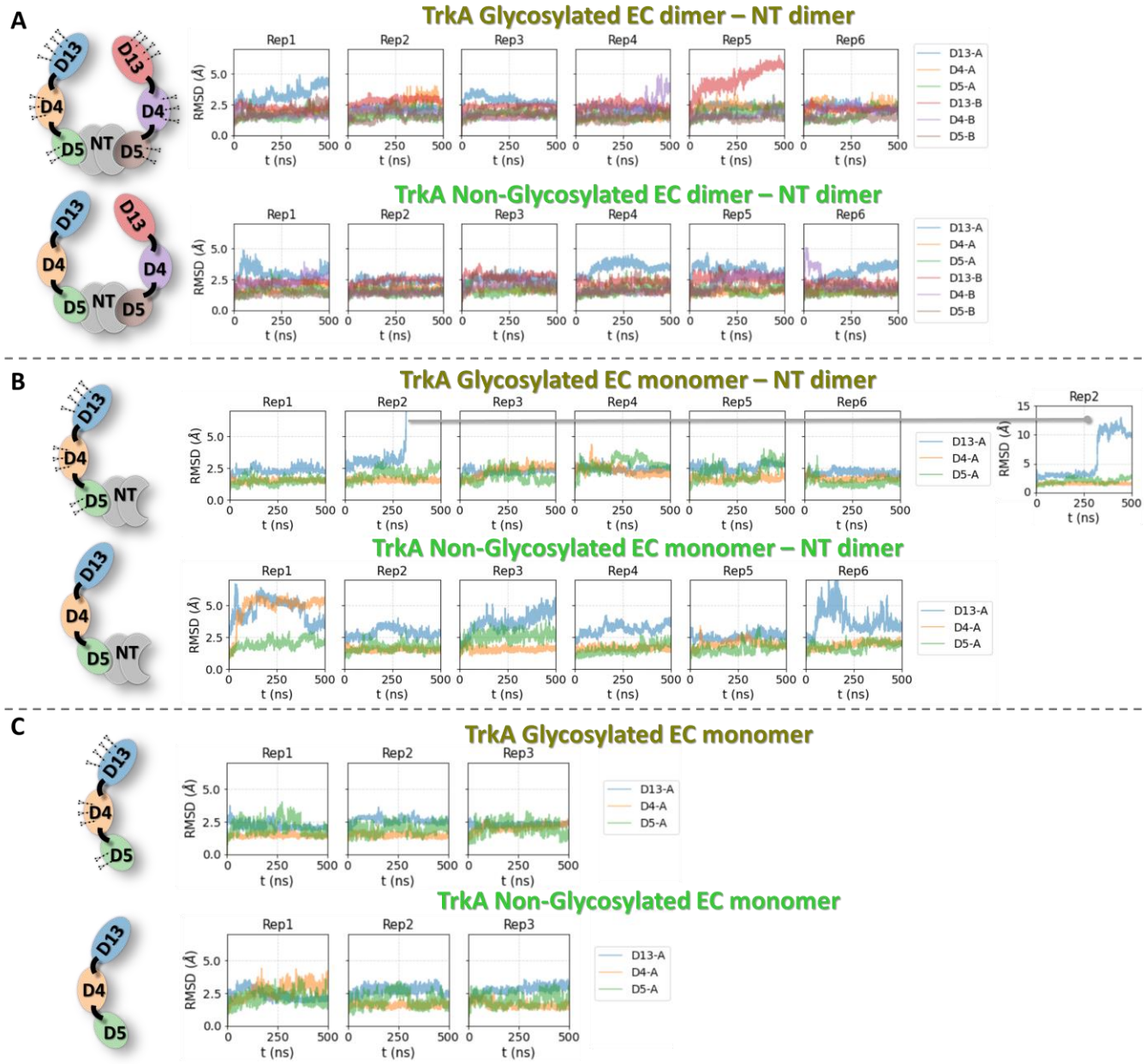


Figure 3.6: RMSD calculations for each domain of the glycosylated or not TrkA systems. (A), (B) and (C) show the internal RMSD for each domain the TrkA-EC homodimer with the NGF homodimer bound, the TrkA-EC monomer with the NGF homodimer bound, the TrkA-EC monomer systems, respectively, with and without glycans. For each system the time-dependent RMSD of the domain backbone compared to the first frame is shown for each replica. Before the calculation of the internal RMSD of each domain, I performed alignment of the respective domain at each simulation snapshot. Schematic representation

of each system, with color coding of the domains is shown at the left. The first and second subunits are indicated with A and B letters in the legends.

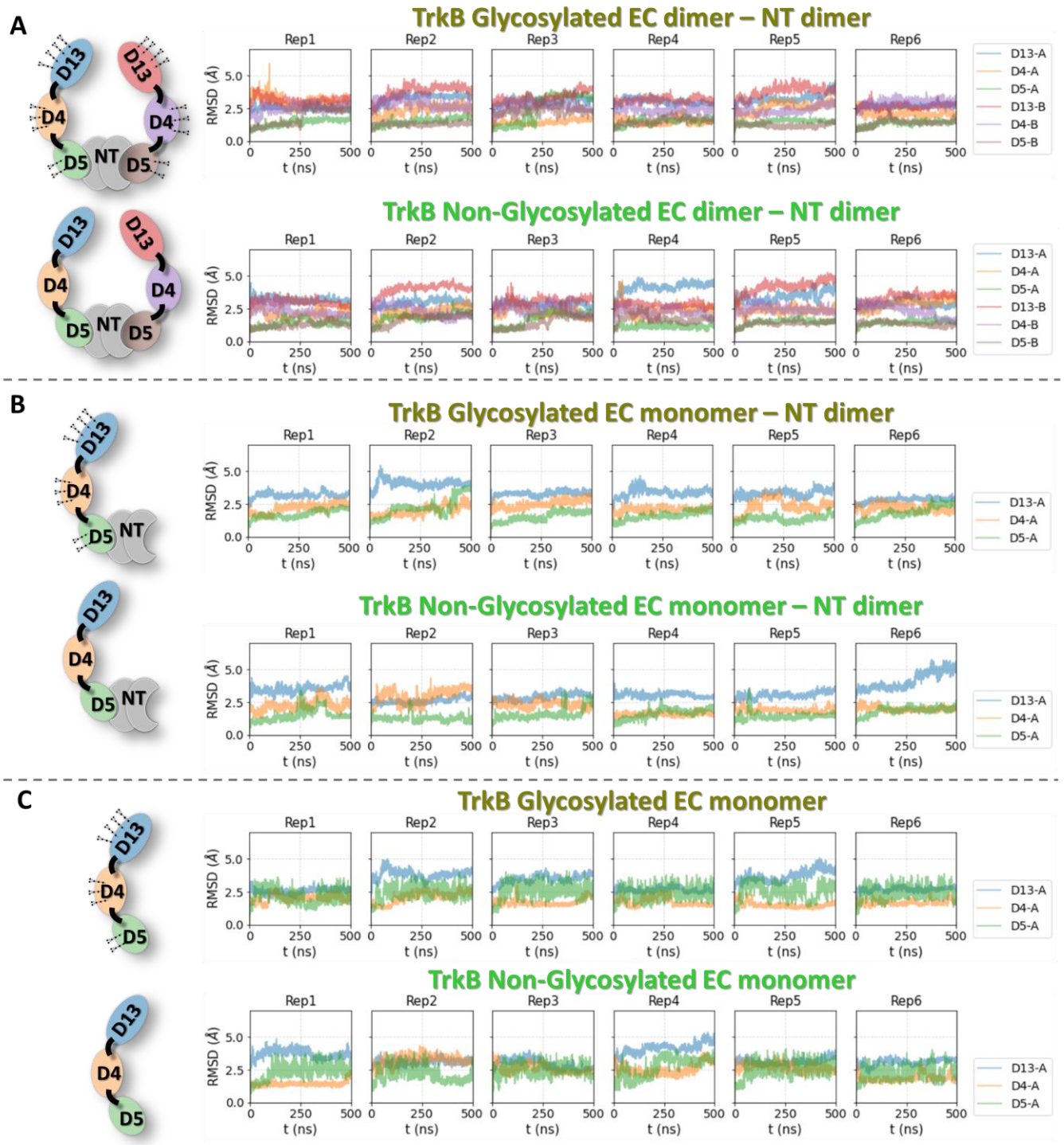


Figure 3.7: RMSD calculations for each domain of the glycosylated or not TrkB systems. (A), (B) and (C) show the internal RMSD for each domain the TrkB-EC homodimer with the NT-4/5 homodimer bound, the TrkB-EC monomer with the NT-4/5 homodimer bound, the TrkB-EC monomer systems, respectively,

with and without glycans. For each system the time-dependent RMSD of the domain backbone compared to the first frame is shown for each replica. Before the calculation of the internal RMSD of each domain, I performed alignment in the respective domain at each simulation snapshot. Schematic representation of each system, with color coding of the domains is shown at the left. The first and second monomers are indicated with A and B letters in the legends.

Subsequent analysis was focused on the calculation of the root mean square fluctuation (RMSF) of each protein residue during the simulations of the several TrkA and TrkB systems. I performed the RMSF calculations, after aligning first the proteins to average structures from the simulations. The results are presented in Figure 3.8. Comparison of the residue RMSF shows no significant differences between the glycosylated and non-glycosylated systems. Only for the case of the "TrkA EC monomer" simulations (Figure 3.8C), the D4 domain seems to fluctuate more in its non-glycosylated state, although there is some overlap between the areas of RMSF standard deviation in the glycosylated and non-glycosylated states. The RMSF values of the neurotrophin residues are smaller than that of the receptor, except of the N- and C-terminal regions, which are more mobile because they correspond to unstructured, flexible loops. Also for the case of the neurotrophin, the glycosylation of the receptor does not seem to affect the residue fluctuations over the simulations.

Apart from the RMSF of the protein residues, I also calculated the RMSF of the glycans. For this purpose, I aligned first each of the EC domains to the first frame and then I calculated the averaged RMSF value of all monosaccharides of each sugar. The box plots showing the variance of the average RMSF for each glycan across the several replicas are presented in Figure 3.9. A general trend of higher RMSF values is observed for the glycans that are located closer to the N-terminus of the EC domain. This seems to follow the RMSF behavior of the EC domain itself, rather than the fluctuations of each glycan. Comparison of the RMSF for the TrkA and TrkB glycans shows no significant differences between the two receptors.

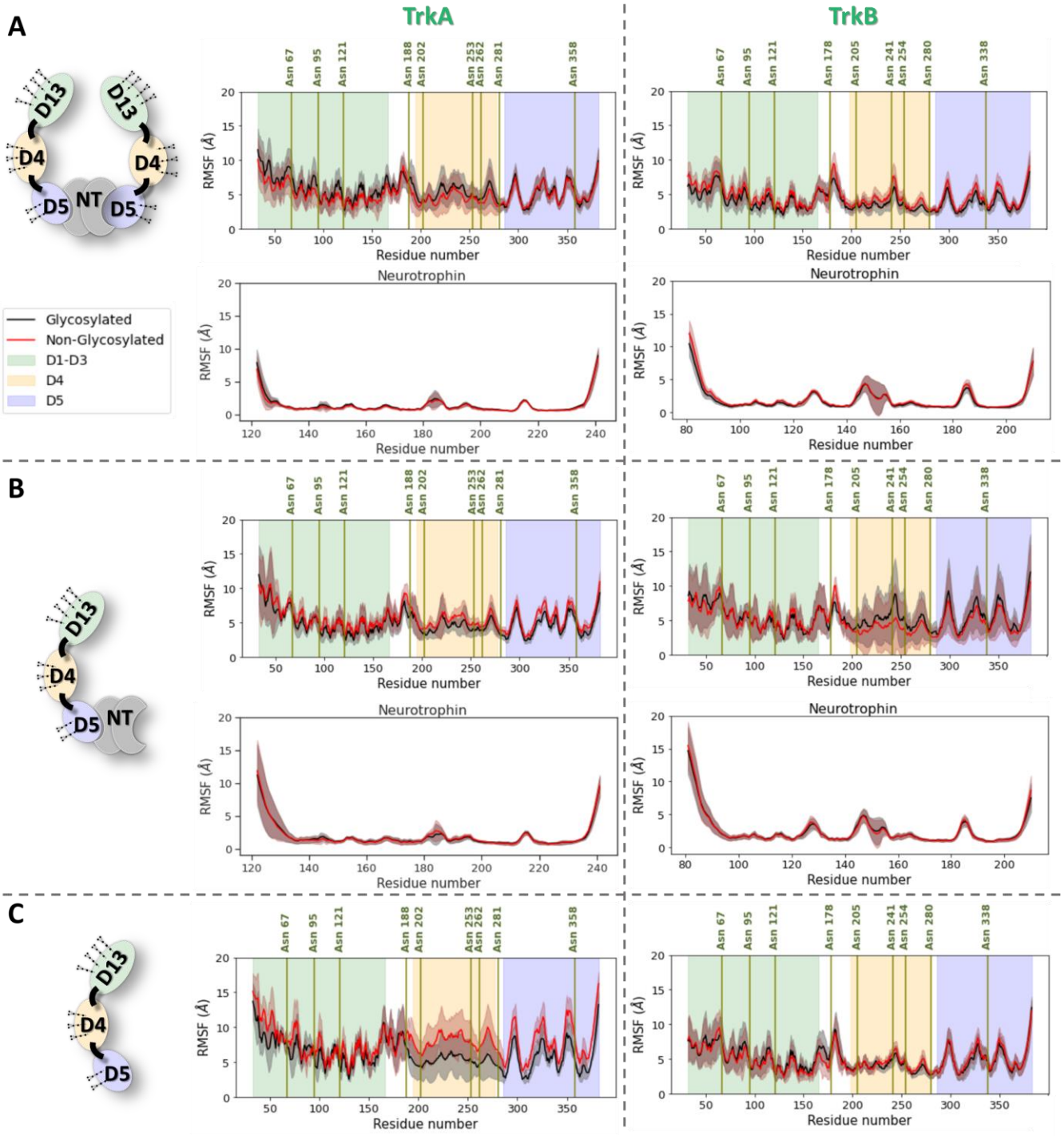


Figure 3.8: RMSF calculations of all the glycosylated or not TrkA and TrkB systems. (A), (B) and (C) show the average RMSF values from all replica simulations for each receptor and neurotrophin residue of the TrkA/B-EC homodimer with the NGF/ NT-4/5 homodimer bound, the TrkA/B-EC monomer with the NGF/ NT-4/5 homodimer bound, the TrkA/B-EC monomer systems, respectively, with or without glycans. The average RMSF values from all replica simulations are shown for each residue in black and red lines for the glycosylated and non-glycosylated systems, respectively. The standard deviation for each residue is

shown with shadowed areas. The residue positions of glycosylated asparagines are shown with vertical lines. Before the RMSF calculation, I aligned the systems to an average structure from the simulations.

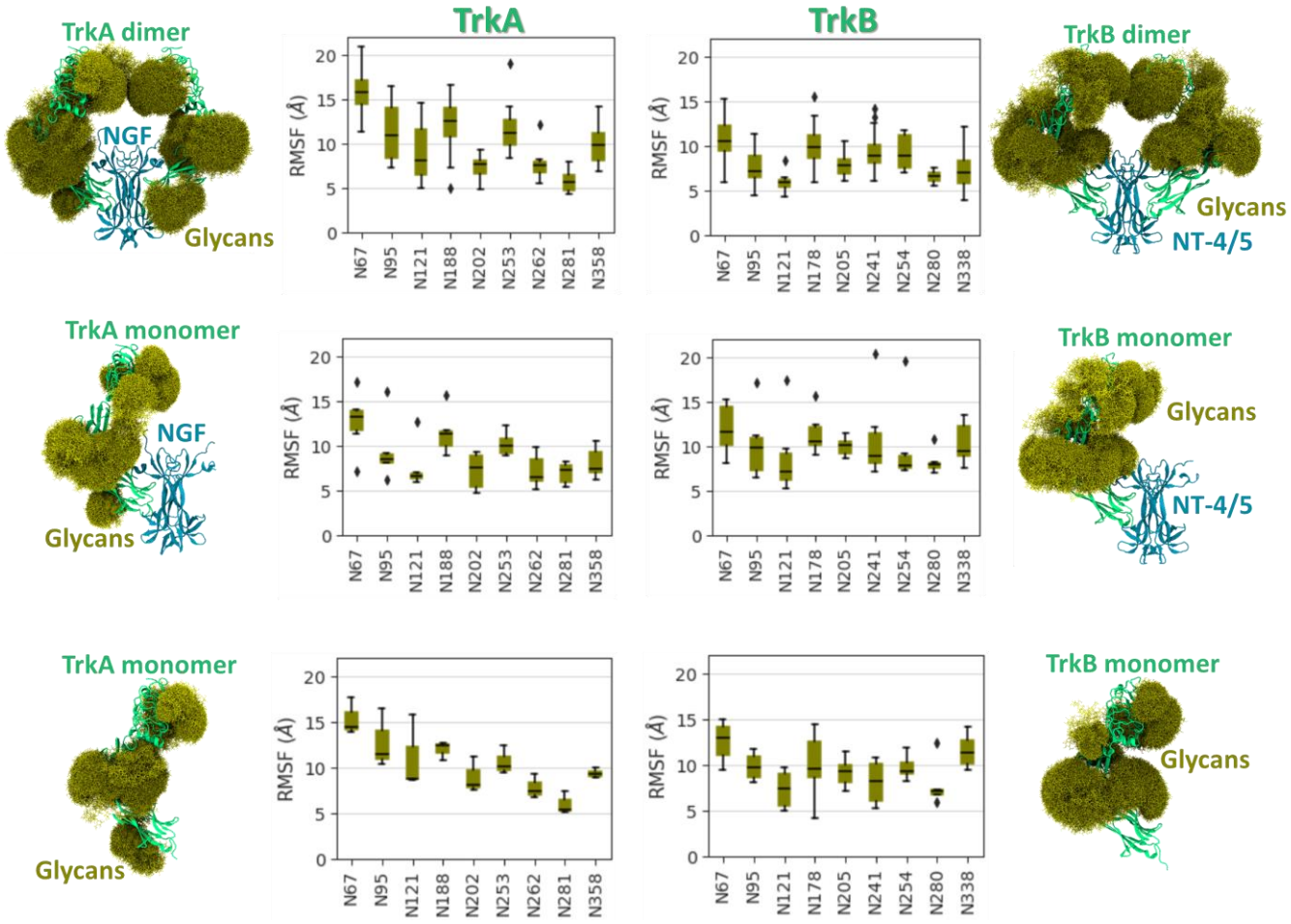


Figure 3.9: Box plots of RMSF calculations of all glycans attached in asparagine residues in all TrkA and TrkB glycosylated systems. Before the RMSF calculation, I performed alignment of each EC segment to its average structure from the simulations.

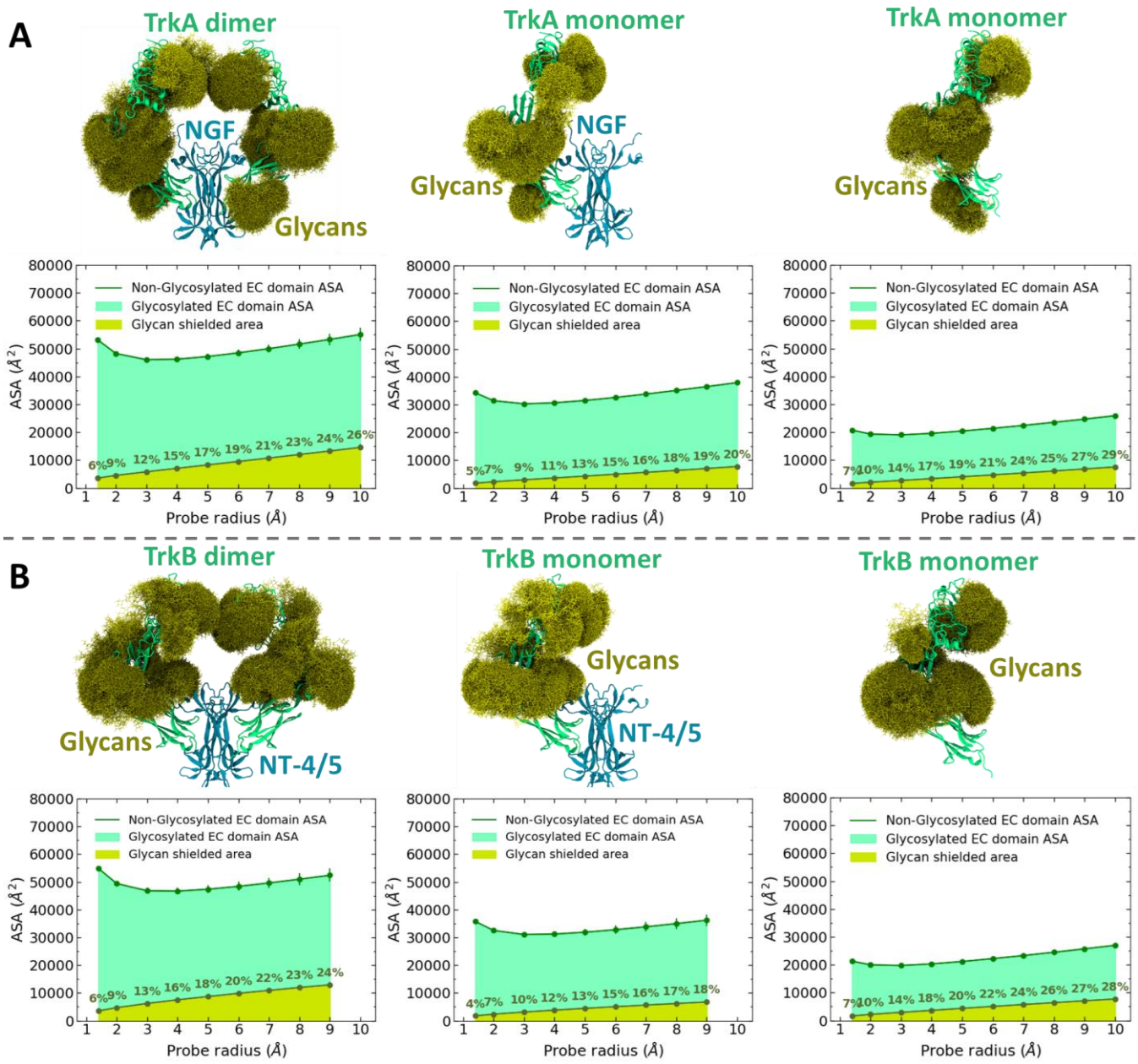


Figure 3.10: Accessible surface area (ASA) of the (A) TrkA and (B) TrkB systems (including the ASA of the NTs when present) and the area shielded by glycans at multiple probe radii from 1.4 Å (water molecule) to 10 Å (protein-sized molecule). The values have been calculated and averaged across all replicas with glycans and are reported with standard deviation bars. The area shielded by the glycans is presented in yellow (rounded % values are reported), whereas the green line represents the accessible area of the protein in the absence of glycans. Highlighted in green is the area that remains accessible in the presence

of glycans. The ASA for the TrkB dimer and monomer systems, which was calculated for probe radius 10 Å, is not shown because of incorrect calculation of the ASA in this high radius for the particular systems.

Subsequent analysis was focused on the investigation of the accessible surface area shielding by the glycans. As has been previously demonstrated by MD simulations, glycans are able to shield large parts of protein area exposed to solvent or other molecules.[255] Due to that, I calculated the accessible surface area (ASA) of the EC parts of the receptors, with and without glycans, for different probe radii (Figure 3.10). The latter range from 1.4 Å, which corresponds to the size of water molecules, to 10 Å, which accounts for larger molecules, such as proteins. Intermediate values correspond to the size of small-molecules (~3-4 Å) or oligopeptides (5-7 Å). The results presented in Figure 3.10 show that increasing the size of the probe has a more apparent effect on the glycan shielding, which can be up to 29% of the receptor ASA, meaning that the binding of larger proteins rather than small molecules may be more disturbed. This may be relevant considering the protection of the receptors from proteolytic cleavage by e.g. secretases, which is something that happens physiologically for the p75 receptor.[280] Comparison of the TrkA and TrkB receptors, does not show any differences, even though the glycans are attached in different positions on the receptors. This means that the size of the glycans and the total number of them, which are the same for both receptors, determine the amount of shielding. The small bars for standard deviation show that all replicas sample configurations with the same ASA, providing confidence about the presented results.

In the same context of accessible surface area, I also calculated the contact area of the D5 domain of the receptors with the bound NT and how this is affected by the glycans (Figure 3.11). The contact area was calculated by measuring the ASA of the D5 domain with and without the NT present and then by subtracting the two values. In the systems where glycans were present, the calculation was performed with and without considering the glycans attached to the D5 domain as part of it (Figure 3.11 A, C). Differences in the contact area from those two calculations (with and without glycans) demonstrates the participation or not of the glycans in the D5-NT interface. As shown in Figure 3.11 A and C, the contact area is higher in most cases where the glycans are taken into consideration. This means that some of the glycans that are attached in the D5 domain interact directly with the NT and participate in the D5-NT interface. Especially for the TrkB systems, the difference with and without glycans is even higher. This is probably due to the glycan attached in Asn 338, which is in close proximity to the NT and which is present only in TrkB and not TrkA. This glycan seems to interact with the NT during

the simulations and to increase the contact area of the D5 with the NT. The same analysis was performed also for the systems without glycans (Figure 3.11 B, D), showing similar contact areas with the glycosylated systems when glycans were not considered in the calculation. However, for the "TrkA EC dimer-NT non-glycosylated" and "TrkB EC monomer-NT non-glycosylated" systems, the contact area seems to be slightly higher than the equivalent glycosylated systems. This could be due to tighter packing of the D5-NT contact area in the absence of the glycans.

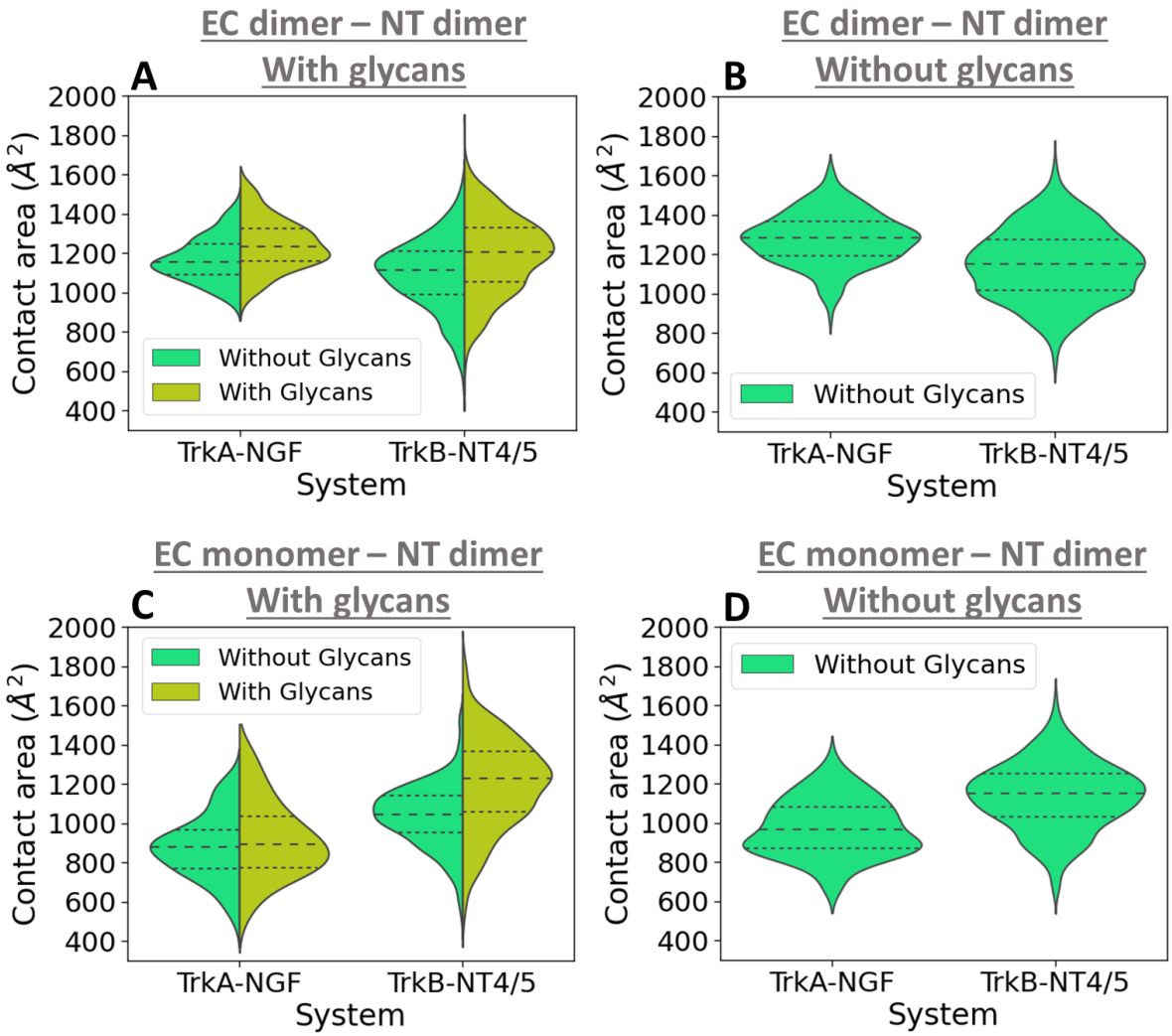


Figure 3.11: Violin plots with the contact area between the D5 domain of TrkA and TrkB receptors with the NTs NGF and NT4/5, respectively. The systems with the EC receptor dimer bound to NTs with and without glycans are shown in (A) and (B), while the systems with the EC receptor monomers bound to NTs with and without glycans are shown in (C) and (D). For the systems with the glycans (A) and (C), I calculated two contact areas: one between the D5 domain together with the glycans and the NT (yellow) and one between the D5 domain without the glycans and the NT (green). I calculated the contact area

for each monomer separately and thus included two values (for the two monomers) from each frame. This comparison shows the participation of the glycans in the interaction of the D5 domain with the NT. For the systems without the glycans (B) and (D), only one contact area between the D5 domain and the NT could be calculated and thus there is only one distribution in these violin plots.

The main reason of modeling and simulating the glycosylation of the TrkA and TrkB receptors was for the investigation of the glycan effects in the conformations of the EC domains of the receptors. To study this, I calculated the radius of gyration of the EC domains during the simulations of all TrkA and TrkB systems (Figure 3.12 and Figure 3.13). The radius of gyration is a metric of the distribution of the atoms of a protein around its axis. For the case of the EC domains here, high values of protein radius of gyration indicate extended conformations of the domains, while smaller values correspond to more compact and globular conformations. Comparison of the radius of gyration distributions from the glycosylated and non-glycosylated systems reveals that the non-glycosylated systems can sample conformations of smaller radius of gyration, indicating that the EC domain conformations become less extended. This effect is particularly obvious for the TrkA systems. However, these smaller values have low density because they are sampled in just a small number of frames during the simulations. This happens probably due to the initially extended conformations of the domains, which originate from the available crystal structure of TrkA with NGF[28] that was used to model the receptors. The systems starting from this extended conformation do not have enough time in the simulated 500 ns to properly equilibrate and explore all possible conformations in the glycosylated and non-glycosylated states. However, the current simulations show that there is a trend for the glycosylated systems to remain extended. Besides, the crystal structure of the whole EC domain of TrkA dimer bound to NGF is solved with the presence of the first three oligosaccharides of the glycans, and thus these are perhaps enough to promote the extended conformation of the EC domain. However, simulations of this crystal structure have shown that also less extended conformations of the EC domains can occur in this state.[225] Comparison of TrkA with TrkB shows that the effect of glycosylation is less apparent on TrkB. In fact, there is quite a lot of overlap in the distribution densities of the radius of gyration for the glycosylated and not TrkB systems. This seems to happen because TrkB samples smaller radius of gyration values even in its glycosylated state.

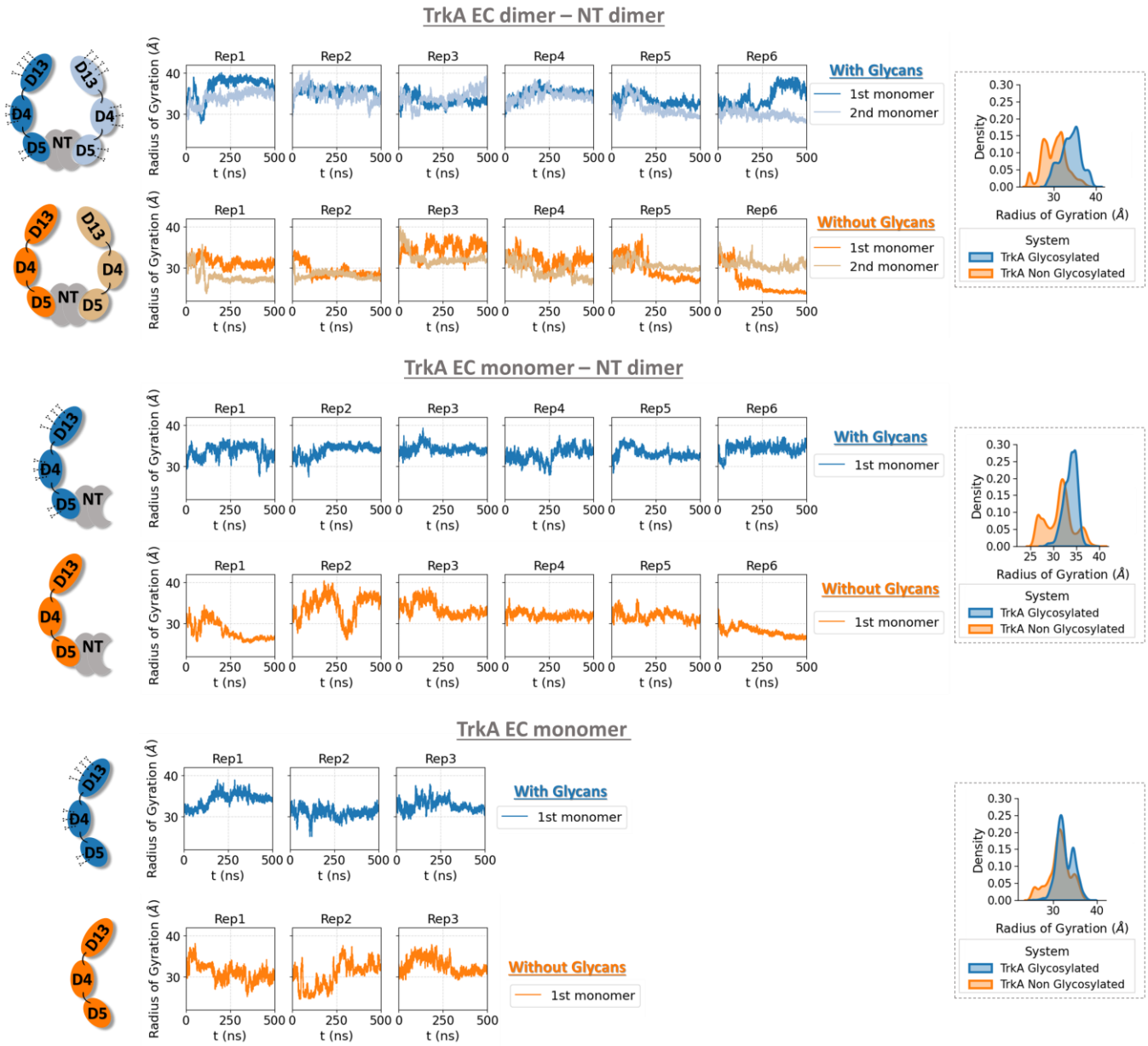


Figure 3.12: Radius of gyration of the receptor backbone of the glycosylated and not TrkA systems. The color coding of each domain is shown in the schematic representations of the systems at the left of the plots. The time-evolution of the radius of each receptor monomer is shown for each replica, as well as the distributions of the radius of gyration values across all replicas for the glycosylated and non-glycosylated systems (right plots).

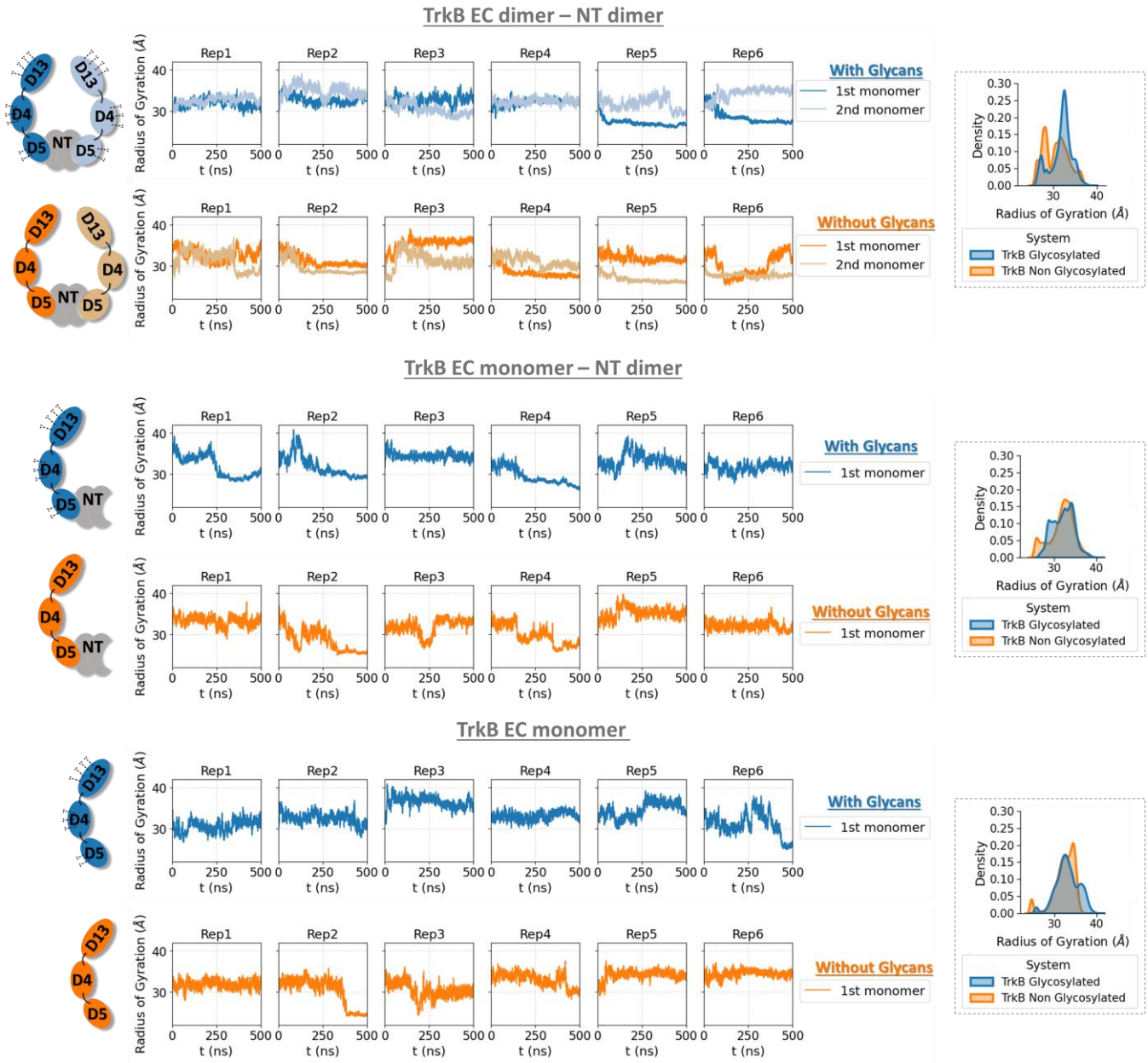


Figure 3.13: Radius of gyration of the receptor backbone of the glycosylated and not TrkB systems. The color coding of each domain is shown in the schematic representations of the systems at the left of the plots. The time-evolution of the radius of each receptor monomer is shown for each replica, as well as the distributions of the radius of gyration values across all replicas for the glycosylated and non-glycosylated systems (right plots).

In order to better understand the protein conformations that correspond to the aforementioned low radius of gyration values, I visualized the trajectories at the time steps where those values occurred. The conformations are shown in Figure 3.14 and Figure 3.15 for TrkA and TrkB, respectively. The figures show that the EC domains bend and interact with each other or with the NT, while for the case of the EC monomers, the D13 domain interacts with the D5. If the glycans were present in these closed conformations, they would sterically clash with the proteins. Similar closed conformations of the EC domains have been reported before from enhanced sampling simulations.[225] These were performed on the crystal structure of the TrkA EC domain bound to NGF,[28] and they included the glycans which were already solved in the crystal structure; i.e. the first three oligosaccharides. Consequently, longer glycan structures appear to be important for keeping the extended conformations of the EC domains.

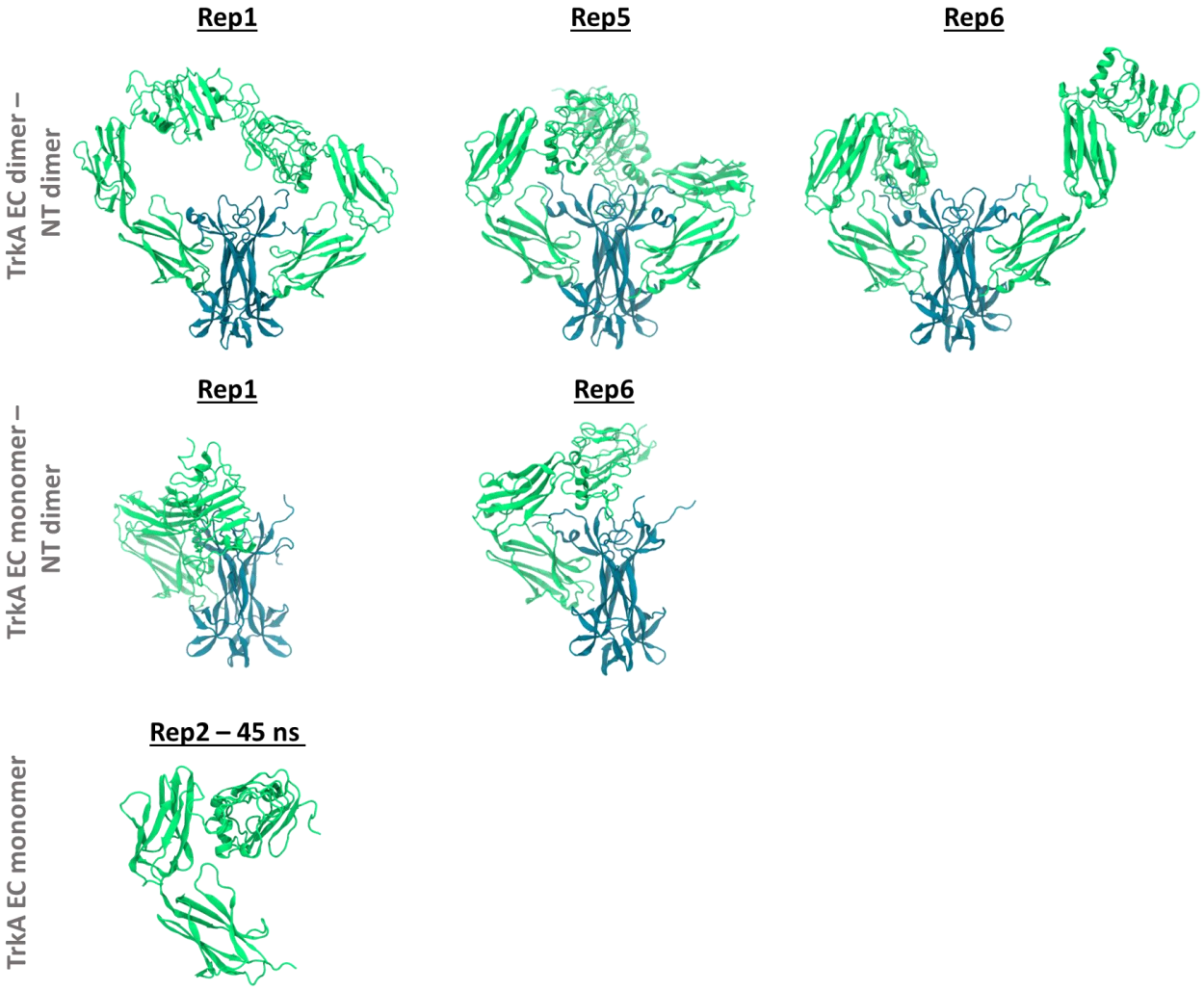


Figure 3.14: Conformations of the different non-glycosylated TrkA systems, which correspond to low radius of gyration values presented in Figure 3.12. All the conformations are taken from the last frame

at 500 ns of the indicated replica simulations, except for the last one which is a rendering at 45 ns of the simulation.

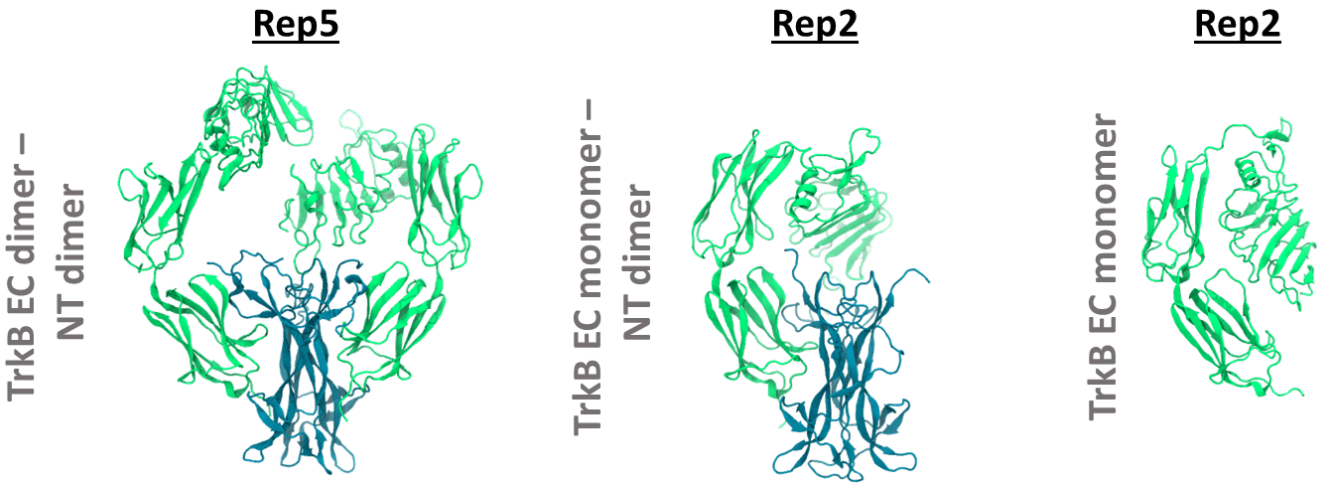


Figure 3.15: Conformations of the different non-glycosylated TrkB systems, which correspond to low radius of gyration values presented in Figure 3.13. All the conformations are taken from the last frame of the indicated replica simulations.

To better understand and quantify the conformations that were sampled in the glycosylated and non-glycosylated states of the receptors, I calculated the distances and angles between the domains and the NTs. Specifically, I analyzed the distances between the D13, D4 and D5 domains from the two receptor monomers, as well as the angle that is formed between these domains and the NT in the middle of them (Figure 3.16). Comparison of the glycosylated and non-glycosylated states shows that smaller D13-D13 distances are sampled in the non-glycosylated systems of both TrkA and TrkB (Figure 3.16A). A smaller average distance is also observed between the D4-D4 domains of the non-glycosylated TrkA system (Figure 3.16B). The distributions for the angles (Figure 3.16 D, E, F) follow a very similar behavior to that of the distances (Figure 3.16 A, B, C). The smaller distances between the D13-D13 domains is in agreement with the smaller radius of gyration for the non-glycosylated systems (Figure 3.12 and Figure 3.13) and with the closed conformations of the receptors that are presented in Figure 3.14 and Figure 3.15.

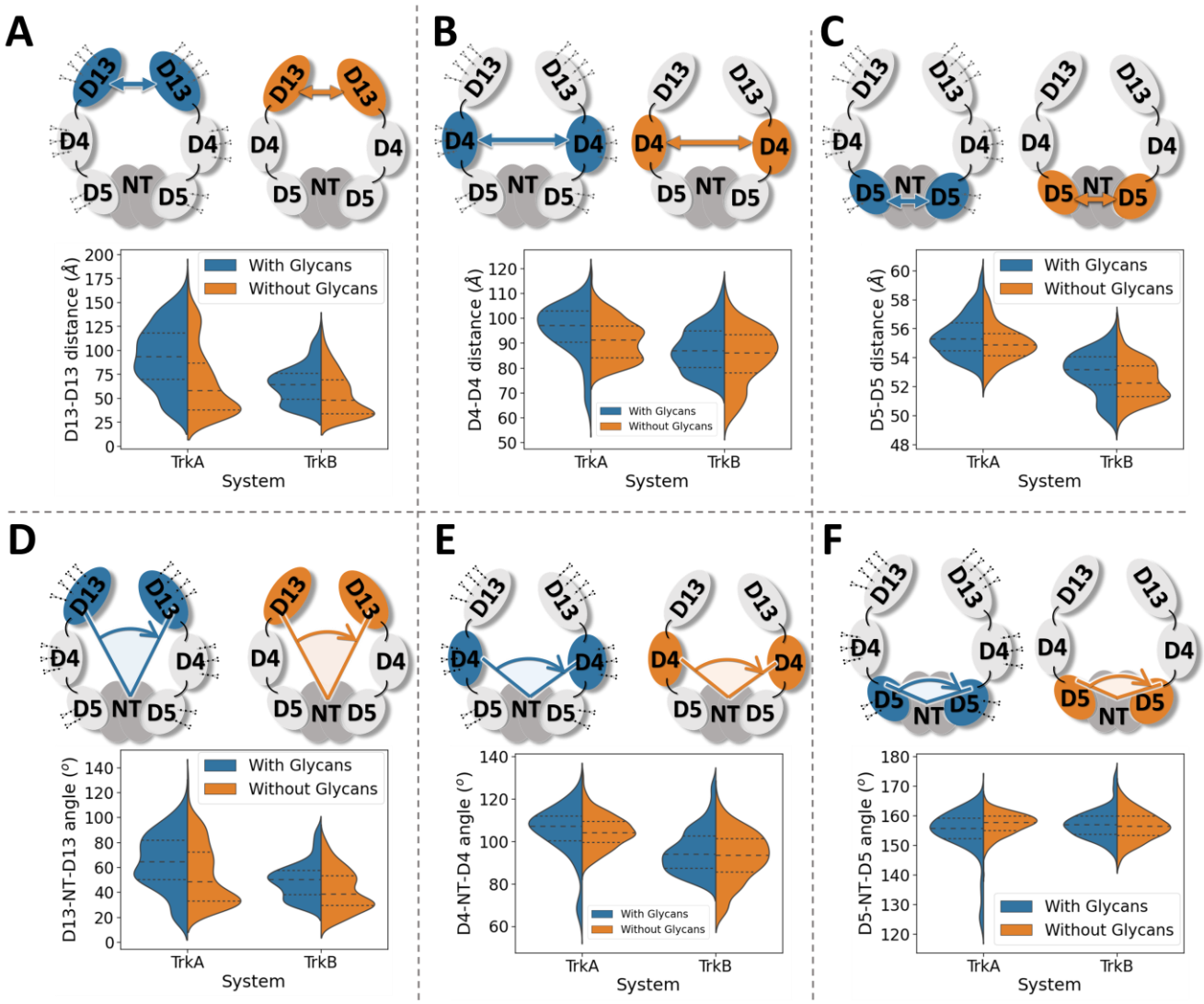


Figure 3.16: Distance and angle analysis between the receptor monomer domains. (A, B, C) Calculated distances between the centers of geometry (COG) of the D13, D4 and D5 domains from the two different receptor monomers. (D, E, F) Calculated angles formed between the COG of one of the same domains from the first receptor monomer, the COG of the NT and the COG of the same domain from the second monomer. In all cases, the distances and angles were calculated every 100 frames and the final distributions from the glycosylated and non-glycosylated TrkA and TrkB systems are shown in violin plots. The quartiles of the distributions are indicated with dashed lines.

A more suitable metric for estimating how extended the EC domains of the TrkA and TrkB receptors are in their glycosylated form is the distance between the D13-D5 domains. According to the results in Figure 3.17, smaller D13-D5 distances are sampled for almost all non-glycosylated systems compared to the glycosylated ones. This seems to be more apparent for the TrkA systems, but also the TrkB non-glycosylated systems seem to explore more conformations of smaller distance compared to when glycans are present.

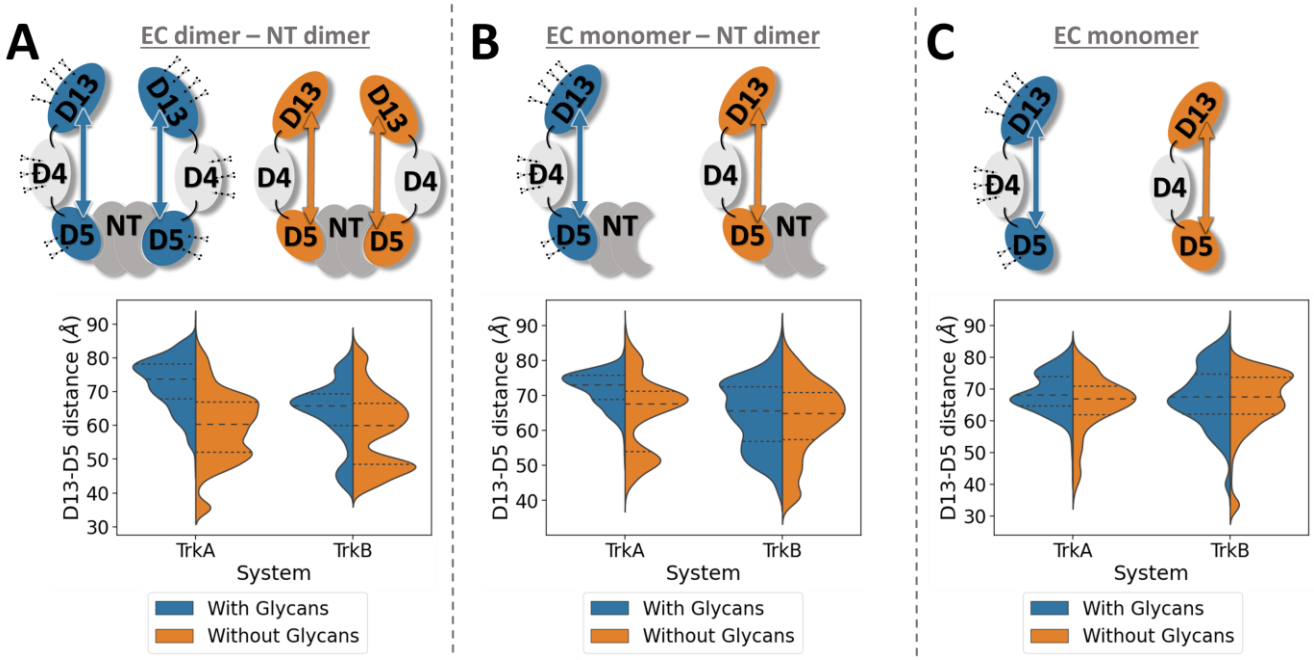


Figure 3.17: Distance between the centers of geometry of the D13-D5 domains in the different TrkA and TrkB systems (A, B, C). In all cases, the distances were calculated every 100 frames and the final distributions from the glycosylated and non-glycosylated TrkA and TrkB systems are shown in violin plots. The quartiles of the distributions are indicated with dashed lines.

The differences caused by glycosylation in the several TrkA and TrkB systems were also examined with Principal Component Analysis (PCA) which can reveal the most dominant motions of the proteins. Specifically, since most fluctuations in the systems originated from the movement of the D13 domain, I calculated the Principal Components of this domain after aligning the trajectories to the D4-D5 domains. The results for the different TrkA and TrkB systems are presented in Figure 3.18 and Figure 3.19, respectively. Comparison of the principal motions of the glycosylated and non-glycosylated systems shows that they cover mostly the same first (PC1) and second (PC2) principal component space. It should

be mentioned that the PC1 and PC2 do not correspond to the same vectors in all replica simulations, as they have been separately calculated for each replica. In order to examine the relationship between the PC1 and PC2 from the separate replicas, I calculated the dot products for each vector pair from glycosylated and non-glycosylated replicas (Figure A.4 and Figure A.5). The results revealed dot products close to zero, and thus not so similar motions between the glycosylated and non-glycosylated TrkA systems, whereas for TrkB, stronger similarities were observed with parallel or anti-parallel principal components, especially for the systems with receptor monomers. This behavior seems to coincide with the bigger differences observed between glycosylated and non-glycosylated forms for the TrkA systems compared to TrkB.

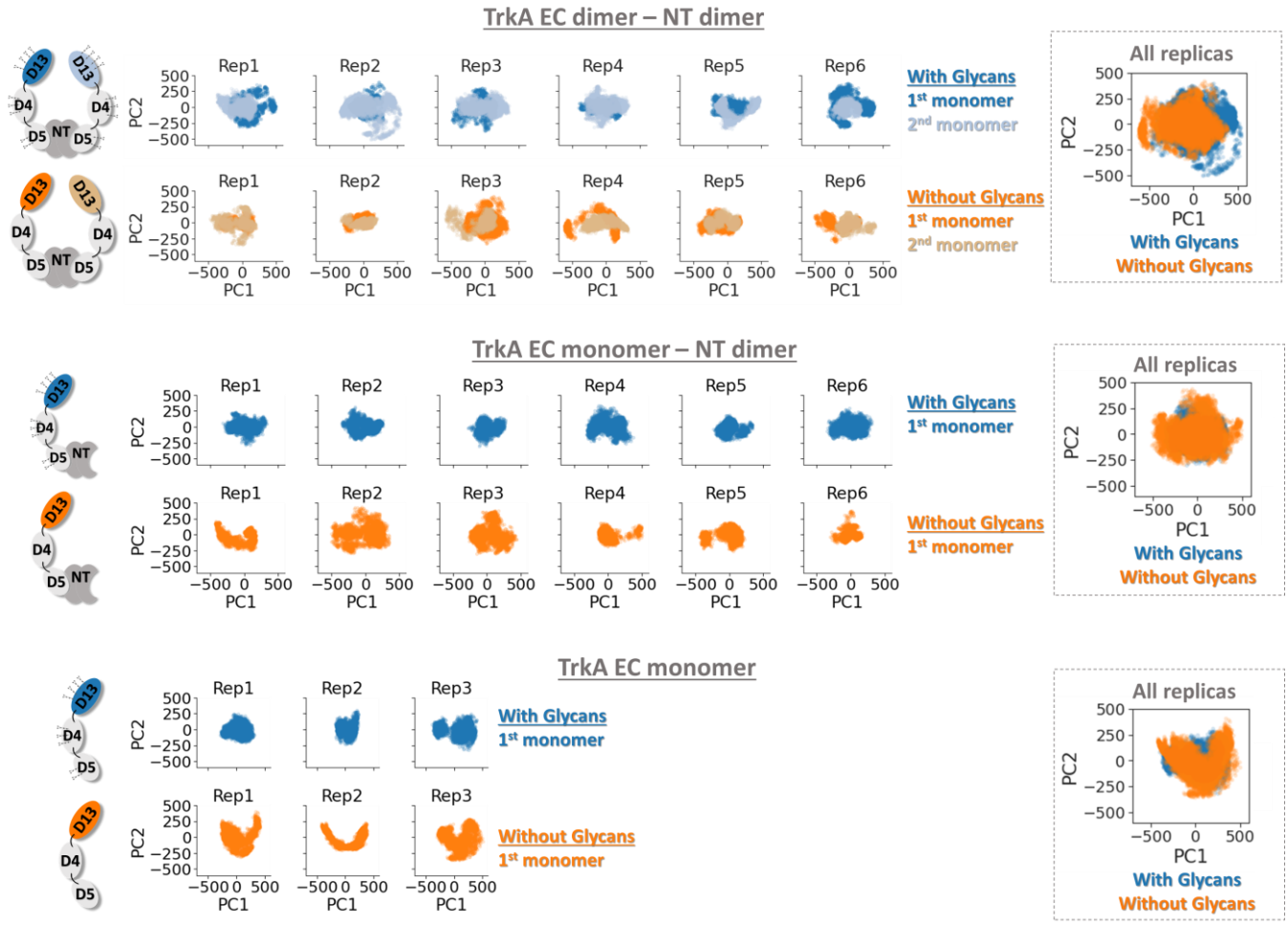


Figure 3.18: First and second principal components of the D13 domain during the simulations of the glycosylated and not TrkA systems. PCA analysis was performed to the D13 domains after alignment of

each EC segment to the D4 and D5 domains. The color coding of each domain is shown in the schematic representations of the systems at the left of the plots.

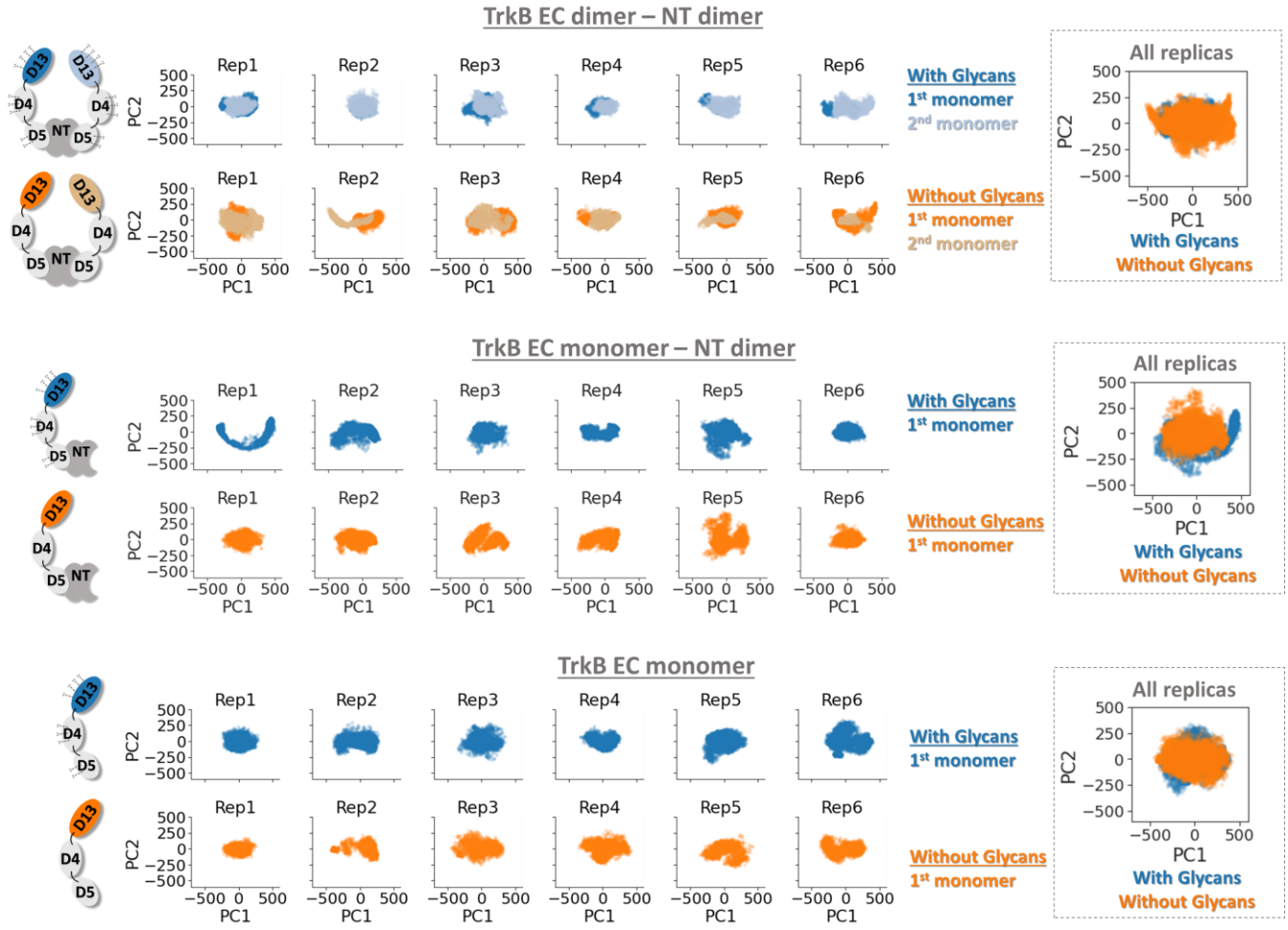


Figure 3.19: First and second principal components of the D13 domain during the simulations of the glycosylated and not TrkB systems. PCA analysis was performed to the D13 domains after alignment of each EC segment to the D4 and D5 domains. The color coding of each domain is shown in the schematic representations of the systems at the left of the plots.

Since most of the D13 domain conformations originate from hinge motions in the linker between the D13 and D4 domains, I performed PCA on the dihedral space of this linker in all systems. The results for TrkA and TrkB are presented in Figure 3.20 and Figure 3.21. As in the PCA in the coordinate space of D13 domain, the PCA in the dihedral space of the linker does not reveal any important differences between the glycosylated and non-glycosylated states.

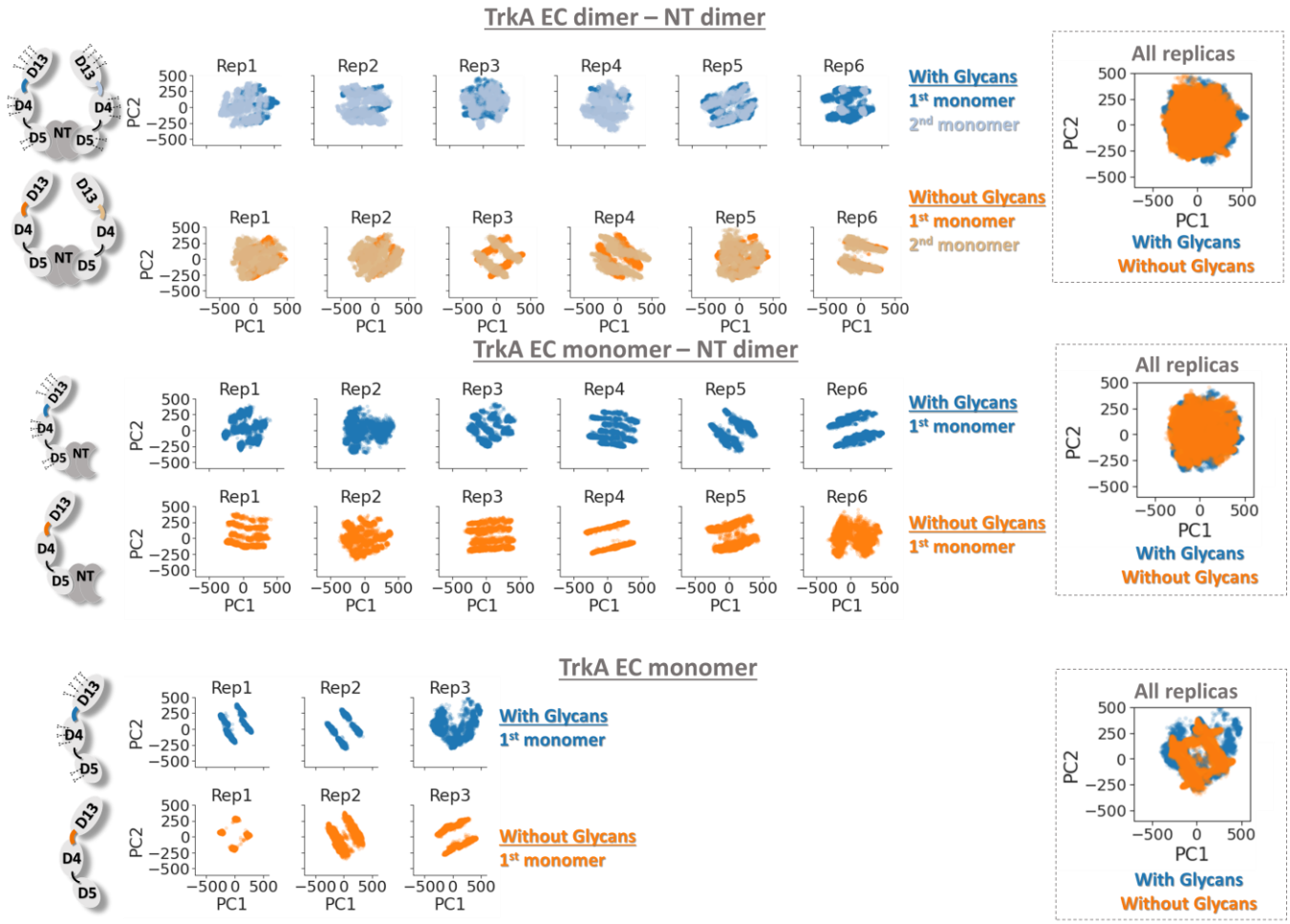


Figure 3.20: First and second principal components of the linker between the D13-D4 domains phi and psi dihedrals during the simulations of the glycosylated and not TrkA systems. Dihedral PCA analysis was performed to the linker after alignment of each EC segment to the D4 and D5 domains. The color coding of each domain is shown in the schematic representations of the systems at the left of the plots.

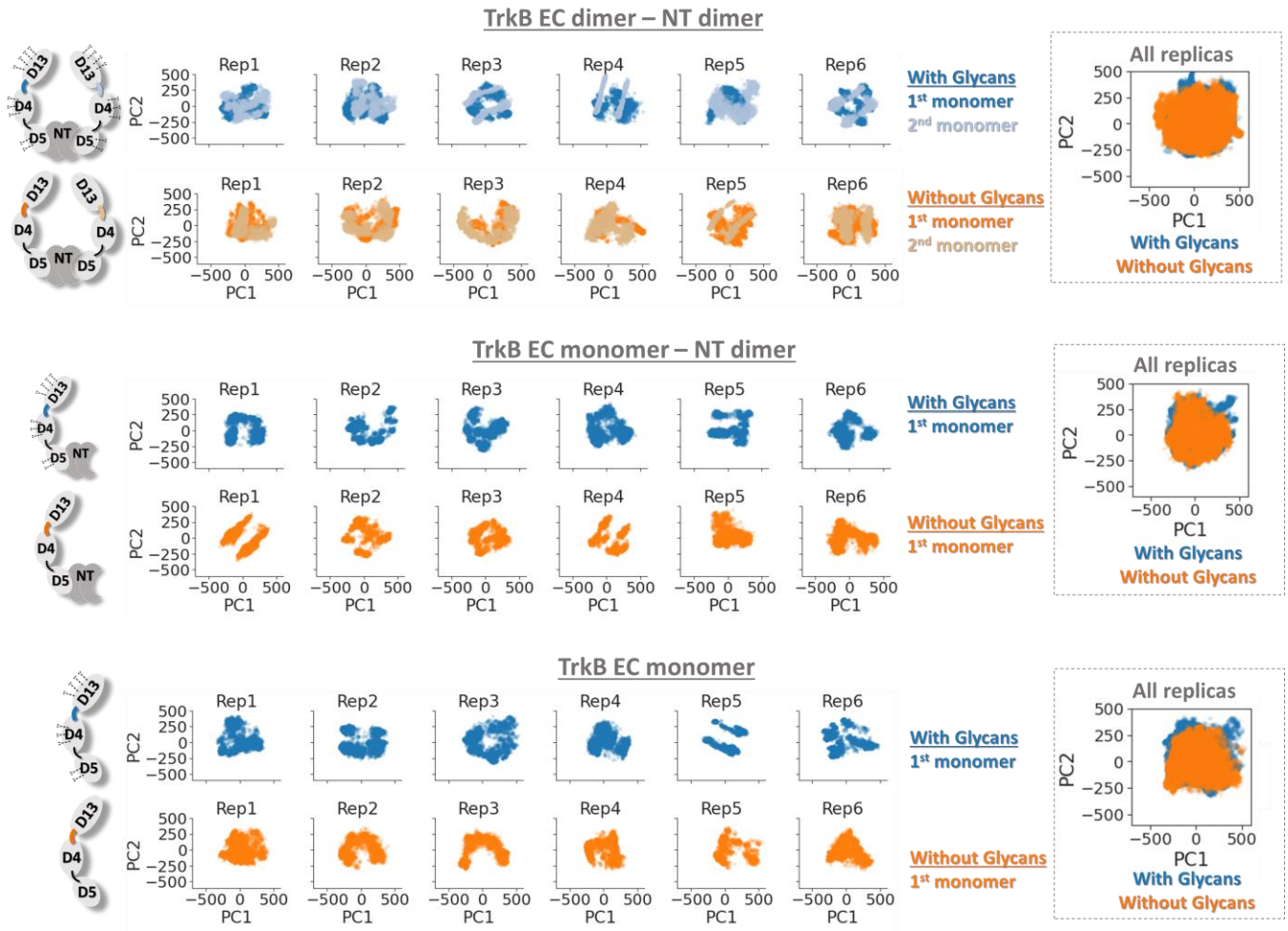


Figure 3.21: First and second principal components of the linker between the D13-D4 domains phi and psi dihedrals during the simulations of the glycosylated and not TrkA systems. Dihedral PCA analysis was performed to the linker after alignment of each EC segment to the D4 and D5 domains. The color coding of each domain is shown in the schematic representations of the systems at the left of the plots.

The time evolution of the principal components was also visualized to monitor the exploration of the PC1 and PC2 space over time. This was done for both the PCA in coordinate space of the D13 domain (Figure A.6 and Figure A.7) and in dihedral space of the linker between D13-D4 domains (Figure A.8 and Figure A.9). The systems seem to explore specific regions of the PC1 and PC2 space. In cases where some conformational change takes place during the simulations, and thus the system moves from one region to another in the PC1/PC2 space, no revisiting of the initial region takes place. This means that the systems do not oscillate between the different states, which is indicative of incomplete sampling. This is

actually expected considering the big size of the systems, the short simulation times (500 ns) and the fact that the observed conformational changes happen in time scales of hundreds of ns or higher.

A final analysis that I performed in the glycosylated systems was the calculation of contacts between the glycans attached in the several positions and the protein residues. The results are presented in Figure 3.22 for both TrkA and TrkB. A contact is defined when any heavy atom (no hydrogens) from the glycans is within 4.5 Å distance from any heavy atom of a protein residue. Only contacts of higher than 10% occupancy are shown. A closer look into the interactions shows that many glycans interact with a residue two positions before the Asparagine in which they are attached. To examine the possibility of these Asn residues adopting specific conformations that allow the glycans to interact with two residues before, I calculated the χ1 dihedral angles of these Asparagines (Figure A.10 and Figure A.11). However, no particular behavior was observed for these residues, except for the more restricted χ1 space when the Asn residues are glycosylated.

For the TrkA systems, the interaction between the glycan in Asn202 and Val200 EC occurs with more than 10% occupancy in all systems and monomers. Then, the glycan at Asn281 interacts with Gln279 in all TrkA systems. In fact, this interaction has ~80% occupancy in the second monomer of the TrkA dimer system. Apart from these interactions, the glycans seem to form contacts with several Glutamic acid, Arginine and Glutamine residues in all TrkA systems.

For the TrkB systems, mostly contacts of lower occupancy occurred compared to TrkA. Among the most stable contacts is the one between the glycan in Asn205 and Ala203 of TrkB for all systems and monomers. Also, the interaction between the glycan in Asn241 and Trp230 occurs quite often and in more than 20% occupancy. Many glycans interact with Glutamic acid and Aspartic acid residues on TrkB.

Overall, the contact analysis between the glycans and the protein residues did not reveal any specific contacts and the glycans mostly interact with their environment rather than any distant residues. This might be because of the conservative modeling of the glycans, which includes only the first five oligosaccharides (glycan core). Perhaps the long antennae that are normally present in physiological conditions would be able to form more specific interactions. On the other hand, since there is usually a lot of variability in the glycan type based on the cell type, different structures of glycans can be found in a single position. According to this, one would not expect specific interactions of the receptor with a particular type of glycan structure, as this might change. Therefore, the presence of the glycans might affect receptor conformations mostly through steric contacts rather than specific interactions, as seen here.

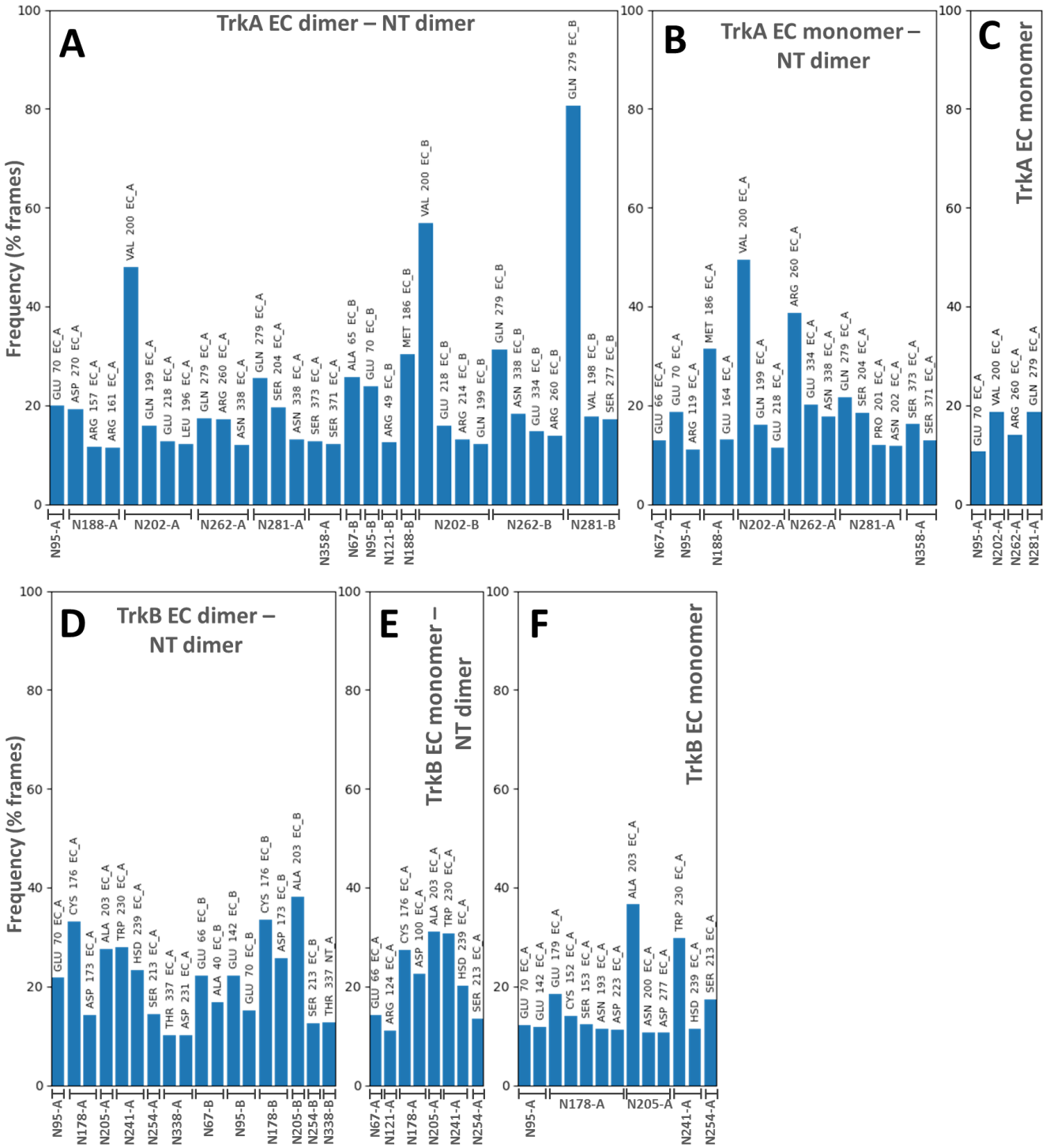


Figure 3.22: Contact analysis between the glycans attached in different Asparagine residues and the protein residues from all glycosylated TrkA and TrkB systems. A contact is defined when any heavy atom (no hydrogens) from the glycans is within 4.5 Å distance from any heavy atom of a protein residue. The glycans are indicated with the Asparagine residues from each receptor monomer (with A and B) in the x-axis, while the protein residues that interact with the glycans are indicated at the top of each bar. The

height of the bars shows the percentage of the total simulation time in which the corresponding contacts are present. Only contacts of higher than 10% occupancy are shown. Each subplot presents the contacts for each TrkA and TrkB glycosylated system.

3.4 Conclusions

In the current chapter of the thesis, I investigated the effects of glycosylation in the EC domains of TrkA and TrkB receptors with MD simulations. To model the glycosylation, I used experimental information about the positions of the glycans in the receptors, while for the chemical structure of the glycans, I chose to model the glycan core, i.e. the first five oligosaccharides, as no experimental information about the types of glycans is available. I modeled the glycosylated and non-glycosylated forms of the TrkA and TrkB receptors as (i) dimers bound to NGF and NT-4/5 neurotrophins, respectively, (ii) as monomers bound to the NTs, and (iii) as monomers alone. Then, I simulated all the systems with unbiased MD simulations for 500 ns using several replicas starting from different velocities.

Visualization of all glycan conformations that were sampled during the simulations showed that TrkA has glycans in all the length of its EC domain, while the glycans in TrkB are mostly accumulated in its N-terminal region. Another important difference between the two receptors is that TrkB does not have a glycan shielding its D5 domain, as in TrkA where the glycan at residue Asn 358 shields the solvent exposed area of the D5 domain. Despite this difference, the effect of glycan shielding was the same in both receptors, with up to 29% of the receptor accessible surface area being covered by glycans when a large probe radius was used for the calculation. This means that the shielding of the receptors from the glycans would probably disturb mostly the binding of larger proteins rather than small molecules. This may be relevant for protection of the receptors from proteolytic cleavage by e.g. secretases, which is something that happens physiologically for the p75 receptor.[280]

I also examined the contact area between receptors and NTs, and the glycans at the D5 domains of the receptors were shown to be part of this interface. This was even more apparent for the TrkB receptor, which has an extra glycan attached in Asn 338, which is in close proximity to the NT and which is present only in TrkB and not TrkA. This glycan seems to interact with the NT during the simulations and to increase the contact area of the D5 with the NT.

The conformations of the receptors were also affected by the presence of the glycans. In general, large RMSD values were sampled during the simulations for both glycosylated and non-glycosylated states, due to hinge motions between the domains. However, the glycosylated forms of the receptors displayed higher values of radius of gyration compared to the non-glycosylated ones, which means that glycosylation promoted the extended conformations of the receptors. Visualization of the trajectories showed that when the glycans are not present, the EC domains bend and interact with each other or with the NT, while for the case of the EC monomers, the D13 domain interacts with the D5. Similar closed conformations of the EC domains have been reported before from enhanced sampling simulations.[225]

The closed conformations of the non-glycosylated receptors were further analyzed with calculations of the distances between the different domains. Comparison of the glycosylated and non-glycosylated states showed that smaller D13-D13 (from the two monomers) distances were sampled in the non-glycosylated systems of both TrkA and TrkB. Also, smaller D13-D5 distances occurred for the non-glycosylated systems, which is direct evidence of the more closed conformations of the non-glycosylated receptors. This was more apparent for the TrkA systems, but also the TrkB non-glycosylated systems seem to explore more conformations of smaller distance compared to when glycans are present. Principal component analysis showed that, especially for the TrkA systems, the first and second principal components of the glycosylated systems had different direction than those of the non-glycosylated systems, with the dot products of the vectors being close to zero.

Contact analysis between the glycans and the protein residues did not reveal any important specific interactions, suggesting that the reason for the glycans promoting the extended conformations of the receptors is mostly due to steric hindrance of the closed conformations when the glycans are present. This is in agreement with the fact that many of the glycans are attached at the linkers between the subdomains or close to them. This way, the glycans can limit the hinge motions of the receptors and keep them in more extended conformations, which might facilitate neurotrophin binding. In the systems where the NT and glycans were not present, the D13 domain was able to bend and interact with the D5 domain, which is the binding site of the NTs. This could hinder NT binding under physiological conditions, or reduce the rate of binding and thus receptor activation by the neurotrophin.

4 Transmembrane domain dimerization mechanism

4.1 Introduction to transmembrane helix function

Neurotrophins bind in the extracellular (EC) segments of receptor dimers and this leads to the activation of the intracellular (IC) kinase domain. The activation signal is transmitted to the IC domains, through the transmembrane (TM) helical domain, resulting in initiation of signaling cascades within the cell (Figure 1.2). The important role of the TM domain in receptor activation has led to more in-depth studies of the molecular and structural characteristics of this domain.

The predominant model for receptor function requires its dimerization, which in the case of the Trks leads to trans-autophosphorylation of tyrosines in the kinase domains, even though it may also be possible for the receptors to act as monomers.[281,282] Biomolecular experiments have shown that TrkA and TrkB can exist in the cell membrane as preformed, yet inactive, homodimers prior to NT binding,[61–63] or even higher oligomerization states.[68] Truncation constructs have suggested that TrkA and TrkB dimerization in the absence of NT is mediated by the TM and IC domains,[74] while p75 forms disulfide-linked homodimers, through C257 in the TM domain, independently of NT binding.[283,284] This indicates that receptor dimers, and thus their respective TM domains, can exist in at least two states; the active and inactive one.

Franco *et al.* recently solved the structure of the TM homodimer of TrkA in dodecylphosphocholine (DPC) lipid micelles with Nuclear Magnetic Resonance (NMR), revealing an arrangement in which the TM helices interact through the $L_{424}xxF_{427}A_{428}xxF_{431}$ motif (yellow residues in Figure 4.1A,B).[60] Mutagenesis studies revealed that when TrkA is active, the TM monomers interact through another interface, with residue V418 playing a central role, suggesting that the NMR interface corresponds to the inactive state of the receptor.[60] Both active and inactive interfaces involve residues in the N-terminal half of the TM helix, with the authors concluding that activation takes place through a rotation of the TM helices around their long axis.[60] Interestingly, the N-terminal TM region of TrkA encompasses three GxxxG-like putative dimerization motifs, $G_{417}xxxA_{421}$, $S_{419}xxxG_{423}$, and $A_{421}xxxA_{425}$, which might facilitate this rotation (magenta residues in Figure 4.1A). One dimerization motif exists also in the TM sequence of TrkB at

residues $A_{440}xxxG_{444}$ (magenta residues in Figure 4.1A,C). Particularly for TrkB, Casarotto *et al.* recently showed with molecular dynamics (MD) simulations that it can form stable TM dimers interacting in the $A_{440}xxxG_{444}$ motif ($A_{439}xxxG_{443}$ in the paper where they used the rat TrkB).[91] In this arrangement, the TM helices of TrkB can form a binding site for the antidepressant drug fluoxetine, which can act as a TrkB agonist.[91]

The aforementioned experimental and computational findings signify the importance of the TM domain in receptor activation and suggest the existence of more than one state of the TM domain. In this chapter of the thesis, I have explored the accessible states of the TM domains of TrkA and TrkB receptors by utilizing atomistic and coarse-grained, unbiased and enhanced sampling molecular dynamics simulations. The goal is to identify all possible states of the TM domains of TrkA and TrkB receptors and possibly understand their biological significance.

4.2 Methodology

Protein models

For the simulations of the transmembrane (TM) domains, I used the Nuclear Magnetic Resonance (NMR) structure of the TrkA-TM dimer with PDB ID 2N90 (Figure 4.1B).[60] For the TrkB-TM dimer, for which there is no experimental structure available, I used bioinformatics web servers and databases to predict the TM and helical parts of the sequence. Specifically, I used UNIPROT,[285] ELM,[286] MEMSAT (PHYRE^2),[287] MEMSAT-SVM,[288] TMpred,[289] TMHMM[290] and PredictProtein[291] for TM sequence prediction, and Jpred4,[292] PSIPRED,[293] NPSA-PRABI[294] and CFSSP[295] for secondary structure prediction (Table 4.1). Based on the predictions from the webservers, I chose to model the V433-R458 sequence as the TM helix of TrkB. Then, I built a homology model of the TrkB-TM homodimer with the AutoModel class of MODELLER v.9.23[246] using the TrkA NMR structure 2N90 as a template (Figure 4.1C). The TM sequences of TrkA and TrkB that I simulated are presented in Figure 4.1A. I also simulated the TM helical dimers of the Ephrin type-A receptor 1 (EphA1) and the Glycophorin A (GpA), which are model systems of TM helix dimerization due to their well-characterized TM interfaces. My goal was to test the protocols that I applied to the TrkA and TrkB receptors. For EphA1, I used the NMR structure with PDB ID 2K1L,[296] while for GpA, I used the NMR structure with PDB ID 1AFO.[297]

Table 4.1: Bioinformatics webservers and databases predicting the TM part (upper table) or secondary structure type (lower table) of the TrkB sequence. The TM sequence predicted by each server is highlighted with green in the upper table, while the helical or strand predicted regions are highlighted with green and orange, respectively, in the lower table. The dashed lines define the region that I modeled as helical in the final model.

Prediction Source	TM predicted sequence
	427 433 458 465
UniProt	GREHLSVYAVVVIASVVGFCLLVMLFLLKLARHSKFGMK
ELM	GREHLSVYAVVVIASVVGFCLLVMLFLLKLARHSKFGMK
MEMSAT (PHYRE[2])	GREHLSVYAVVVIASVVGFCLLVMLFLLKLARHSKFGMK
MEMSAT-SVM	GREHLSVYAVVVIASVVGFCLLVMLFLLKLARHSKFGMK
TMpred (in to out helix)	GREHLSVYAVVVIASVVGFCLLVMLFLLKLARHSKFGMK
TMpred (out to in helix)	GREHLSVYAVVVIASVVGFCLLVMLFLLKLARHSKFGMK
TMHMM	GREHLSVYAVVVIASVVGFCLLVMLFLLKLARHSKFGMK
PredictProtein	GREHLSVYAVVVIASVVGFCLLVMLFLLKLARHSKFGMK

Prediction Source	Strand-helix predicted sequence
Jpred4	GREHLSVYAVVVIASVVGFCLLVMLFLLKLARHSKFGMK
PSIPRED	GREHLSVYAVVVIASVVGFCLLVMLFLLKLARHSKFGMK
NPSA-PRABI	GREHLSVYAVVVIASVVGFCLLVMLFLLKLARHSKFGMK
CFSSP	GREHLSVYAVVVIASVVGFCLLVMLFLLKLARHSKFGMK

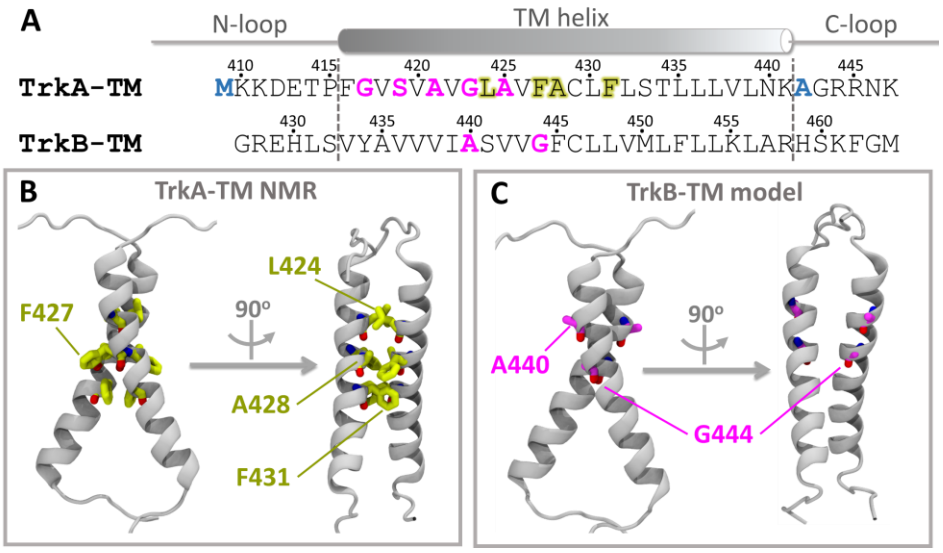


Figure 4.1: (A) Sequences used for the simulations of the transmembrane (TM) domains of TrkA and TrkB receptors. The TM helical part from the TrkA NMR structure (PDB ID: 2N90)[60] is indicated with the dashed lines, while the N- and C-terminal loop sequences in the solved NMR structure are also shown. Mutated

residues in the NMR sequence compared to the human TrkA are colored blue. Residues participating in the GxxxG-like motifs are colored magenta, while the TrkA NMR interface LxxFAxxF motif is highlighted in yellow. (B) Cartoon representation of the TrkA NMR structure (PDB ID: 2N90), with the interface LxxFAxxF motif shown in yellow licorice representation. (C) Cartoon representation of the TrkB-TM dimer homology model, with the AxxxG motif shown in magenta licorice representation.

Coarse-grained (CG) simulations

I performed coarse-grained (CG) molecular dynamics (MD) simulations of the TrkA, TrkB TM dimers in self-assembled dodecyl-phosphocholine (DPC) micelles, which is the environment in which the NMR structure of TrkA was solved,[60] using the Martini 2.2[232,233] force field (Systems 1-2, Table 4.2). I converted the initial atomistic models of the proteins to Martini CG level with the martinize.py script and I used an elastic network model to preserve the secondary structure of the TM helices while retaining the intrinsic flexibility of the N-loop and C-loop regions during the simulations. The elastic network applies additional harmonic restraints with an elastic force constant of 500 kJ/(mol·nm^2) and a distance cutoff range of 5-9 Å. Then, I inserted the CG protein models in a simulation box together with 100 DPC molecules at random positions, in order to retain a detergent-to-protein molar ratio (DPR) of 50:1, as used for the determination of the NMR structure of TrkA.[60] I solvated the systems with the standard non-polarizable water model (NPW) with 10% of the anti-freeze water type (WF) and I neutralized the charge of the box by adding Na^+ and Cl^- ions at 150 mM concentration. I also performed extra test simulations (Systems 3-4, Table 4.2) with ionic strength 24 mM and detergent-to-protein molar ratio (DPR) 40:2, as presented in the experimental section of the 2N90 NMR structure in the PDB website, to examine the influence of these parameters in the sampled configurations of the TM homodimers.

All simulations were run with the GROMACS v.2020 MD engine.[262] I started each simulation with a steepest-descent energy minimization, which was followed by a 2 μs equilibration in the NPT ensemble with the protein helical backbone being restrained with a force constant of 4000 kJ/(mol·nm^2), in order to retain the initial dimer arrangement, while the DPC micelles self-assembled around the protein TM domain. I assessed the self-assembly of the micelles by monitoring the radius of gyration of the DPC lipids, with constant low values indicating micelle formation (Figure A.12). I ran additional test simulations (Systems 5-6, Table 4.2) to examine the importance of the equilibrated micelle in the TM arrangements. I ran equilibration simulations at a constant temperature of 310 K, maintained using the

velocity rescale thermostat[298] with coupling constant 1 ps, and at a constant pressure of 1 bar, maintained with the Berendsen barostat.[299] I also used a coupling constant of 4 ps to maintain isotropic pressure coupling with a compressibility of $4.5 \cdot 10^{-5}$. A time step of 20 fs was applied. I treated the non-bonded interactions with a reaction field for Coulomb interactions, and the cutoff distance that I used for these and for van der Waals interactions was 1.1 nm. I ran 20 replica equilibration simulations for both TrkA and TrkB systems starting from different initial protein and lipid positions, as well as different initial velocities. Ten replicas in which equilibrated micelles (Figure 4.2A) had been successfully formed by the end of the 2 μs equilibration were progressed to production simulations (Figure A.12). I ran the production simulations with fully unrestrained systems for a duration of 20 μs in the NPT ensemble at a constant temperature of 310 K, maintained using the velocity rescale thermostat with coupling constant 1 ps, and at a constant pressure of 1 bar, maintained with the Parrinello-Rahman barostat.[266,267] I used a coupling constant of 12 ps to maintain isotropic pressure coupling with a compressibility of $3 \cdot 10^{-4}$. I applied a time step of 20 fs. I treated the non-bonded interactions with a reaction field for Coulomb interactions, and the cutoff distance for the Coulomb and van der Waals interactions was 1.2 nm.

Finally, I performed some additional test CG simulations with the Ephrin type-A receptor 1 (EphA1) and the Glycophorin A (GpA) TM homodimers (Systems 7-10, Table 4.2). My goal was to test the efficiency of the protocol that I applied to the TrkA and TrkB receptors, in reproducing the well-characterized TM interfaces of those model systems. I tested the protocol by embedding the TM dimers in a 80x80 Å 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) bilayer grid using the insane script.[300] Then, I simulated the systems in the same way as presented above, with the Martini 2.2 force field, and also with the newer Martini 3,[240] to also examine the influence of the force field in the sampled TM arrangements. For the simulations in the POPC bilayer, I used semi-isotropic pressure coupling, to ensure that the x and y dimensions of the bilayer are adjusted concomitantly.

Atomistic simulations

I performed all-atom (AA) MD simulations of the TrkA and TrkB TM dimers in DPC micelles using the CHARMM36m force field[263] (Systems 11-12, Table 4.2). I initiated the simulations in DPC micelles from the ten replica CG systems containing formed, equilibrated micelles, after back-mapping to atomistic detail. For this purpose, I built a map for DPC lipids to define the CG beads that correspond to specific atoms (Figure A.13). For the back-mapping, I used the backward.py script developed by the Martini

team.[301] I ran all simulations with the GROMACS v.2020 MD engine.[262] I started each simulation with a steepest-descent energy minimization, which was followed by one NVT and three NPT equilibration steps in which gradual removal of restraints on protein atoms took place. Specifically, I ran the first NVT equilibration at 310 K for 200 ps with the protein helical backbone being restrained with a force constant of 4000 kJ/(mol·nm^2) and the side chains with a force constant of 2000 kJ/(mol·nm^2). Then, I ran the subsequent three (A, B and C) NPT equilibrations at 310 K and 1 bar pressure for 1 ns each with diminishing backbone (BB) and side chain (SC) restraints with force constants of A: BB:4000 kJ/(mol·nm^2), SC:2000 kJ/(mol·nm^2), B: BB:2000 kJ/(mol·nm^2), SC:1000 kJ/(mol·nm^2), and C: BB:1000 kJ/(mol·nm^2), SC:500 kJ/(mol·nm^2). For the first three equilibration steps I used the Berendsen[299] thermostat and barostat and for the last one I used the Nosé-Hoover thermostat[264,265] and the Parrinello-Rahman barostat.[266,267] I used a coupling constant of 5 ps to maintain isotropic pressure coupling with a compressibility of $4.5 \cdot 10^{-5}$. I applied a time step of 1 fs in the first equilibration step and then I increased the time step to 2 fs for the next equilibration steps. I treated the non-bonded electrostatic interactions with PME[268,269] with a cutoff distance at 1.2 nm and I chose the non-bonded van der Waals interactions to be calculated with a cutoff of 1.2 nm and a switching distance of 1.0 nm. I also employed the LINCS algorithm[270] to keep fixed the length of all hydrogen-containing bonds. Subsequently, I ran production runs for 1 μs for each of the 10 replicas with the protein completely unrestrained. For the production runs, I used the same simulation parameters as for the last equilibration step.

Coarse-grained metadynamics (CG-MetaD) simulations

I performed well-tempered metadynamics[302] simulations of the TrkA and TrkB TM dimers in 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) bilayer in the CG level with the Martini 2.2 force field (Systems 9-11, Table 4.2) to calculate the free energy surfaces of dimerization. I embedded the protein models in an 80x80 Å POPC bilayer grid using the insane script.[300] I performed energy minimization with the steepest-descent algorithm and then equilibration with the protein restrained as described above in the "Coarse-grained (CG) simulations" section. After removing the restraints, I ran the production CG-MetaD simulations.

I ran the CG-MetaD simulations with the GROMACS v.2020 patched with Plumed v.2.7.[303] As a biasing collective variable (CV), I chose the distance between the centers of geometry of the backbone (BB) beads of the two helix monomers (Figure 4.2B). A Gaussian potential was deposited on the CV space

every 5000 steps with decreasing height $W = W_0 e^{-V(s,t)/(f-1)T}$, where W_0 = 0.05 kJ/mol is the initial Gaussian height, T = 310 K is the simulation temperature, f = 10 is the bias factor and V(s,t) is the bias potential as a function of the CV (s) and the time. I set the Gaussian width to 0.05 nm, which was approximately the ½ of the CV fluctuations in the unbiased simulations. These metadynamics parameters have been previously used successfully for the reconstruction of the free energy surface of the EGFR TM domain dimerization.[238] The rest of the simulation parameters that I used were the same as for the unbiased CG simulations discussed before, with the exception that here I used semi-isotropic pressure coupling because of the membrane system. I ran the CG-MetaD simulations for more than 100 μs per system, with a total of 388 μs for all CG-MetaD simulations.

Table 4.2: Summary of the simulations that I performed. Test simulations are denoted with *. I performed simulations for the TrkA, TrkB TM dimers. In systems 3 and 4, "I" is the ionic strength and "DPR" the detergent-to-protein ratio. In systems 7-10, I specify the force field with Martini 2.2 (M2) and Martini 3 (M3). The simulation types that I performed are the all-atom (AA), coarse-grained (CG) with Martini 2.2 and coarse-grained metadynamics (CG-MetaD).

System No	Protein	Lipid environment	Simulation type	Replicas	Time/Replica (μs)
1	TrkA TM	DPC micelle	CG	10	20
2	TrkB TM	DPC micelle	CG	10	20
3*	TrkA TM	DPC micelle, I = 24 mM	CG	10	20
4*	TrkA TM	DPC micelle, I = 24 mM, DPR = 40:2	CG	10	20
5*	TrkA TM	DPC non-equil. micelle	CG	10	5
6*	TrkB TM	DPC non-equil. micelle	CG	10	5
7*	EphA1	POPC bilayer	CG-M2	10	20
8*	EphA1	POPC bilayer	CG-M3	10	20
9*	GpA	POPC bilayer	CG-M2	10	20
10*	GpA	POPC bilayer	CG-M3	10	20
11	TrkA TM	DPC micelle	AA	10	1
12	TrkB TM	DPC micelle	AA	10	1
13	TrkA TM	POPC bilayer	CG-MetaD	1	153
14	TrkB TM	POPC bilayer	CG-MetaD	1	120

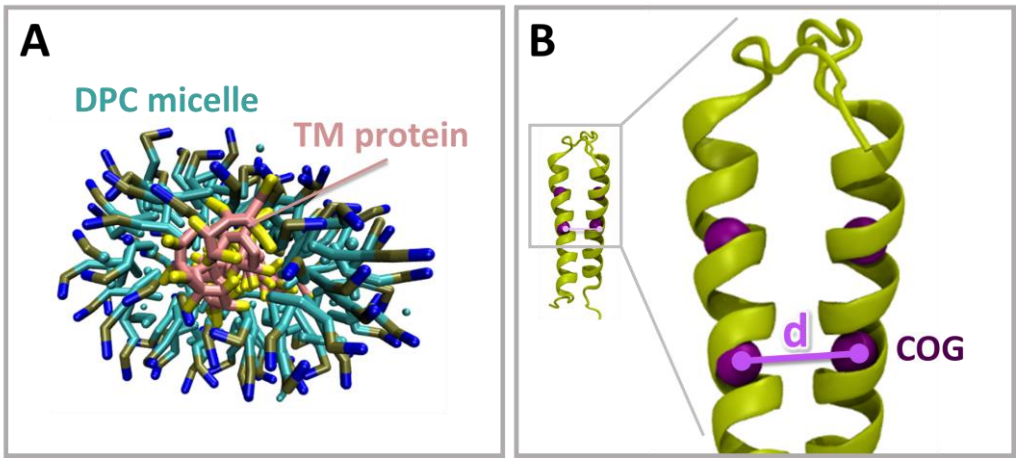


Figure 4.2: (A) Equilibrated micelle formed by DPC lipids (cyan) with the TrkA TM helical protein (pink) in the middle in CG representation. (B) Schematic representation of the collective variable (CV) used in the CG-MetaD simulations. The distance between the centers of geometry (COG) of the backbone beads of the two helices was used as the biasing CV. In the figure, also the COGs of the N-terminal half of the helices are also shown, as they were later used for subsequent analysis.

Analysis of simulations

I post-processed all frames of the simulations with the *gmx trjconv* tool of GROMACS to make molecules whole, remove periodic boundary condition (pbc) jumps and center the proteins in the simulation box. I saved frames every 1 ns. I also aligned the systems to the backbone of the first helix throughout the trajectories. I performed the graphical rendering of protein structures from the trajectories with VMD 1.9.3.[271]

Order parameters

I calculated the order parameters that describe the various arrangements of the TM helices that were sampled during the simulations. These order parameters were (1) the distance between the centers of geometry (COG) of the two helices, (2) the crossing angle between the two helices defined as the dihedral angle formed by the COGs of the two helices and the COGs of the N-terminal halves of the two helices, (3) the phase and position which correspond to the rotation around each helix, defined as the dihedral formed by the upper COG of one helix, the COG of the same helix, the COG of the second helix and one residue, which was Ala 428 for TrkA and Gly 444 for TrkB, (4) the distances of the N- and C-

termini of the two helices, defined by their first and last residues. For the calculation of these order parameters, I used MDAnalysis[272,273] and the numpy.linalg.norm function of the Numpy library.[277] I calculated these parameters for all frames and all replicas.

Kernel Density Estimation (KDE)

I have visualized the distribution densities of the aforementioned order parameters with Kernel Density Estimation (KDE), which smooths the density by summing kernels on the various sampled data points. I calculated and visualized the 2D KDE of two order parameters with the seaborn.kdeplot function of Seaborn.[275]

Contact analysis

I calculated the residues from each helix that form contacts with the residues of the other helix with MDAnalysis. I defined a contact as formed when any CG bead of a residue is within a distance of less than 6 Å of another bead. I calculated the contacts for all frames of all replicas. Then, I post-processed the data with the Pandas library[278,279] and I visualized the results with Matplotlib.[274]

CG-MetaD analysis

In order to assess the convergence of the CG-MetaD simulations, I examined the diffusive behavior of the biasing CV over time. For this purpose, I calculated the time evolution of the distance between the COGs of the two helices with Plumed. I also examined whether the height of the deposited Gaussian potentials decreases over time, which would be an indication of convergence for well-tempered metadynamics. Another convergence test that I performed was to test the time evolution of the free energy in the biasing CV space. I also calculated how the free energy difference between the bound and unbound states changes over time. Finally, I applied the block-analysis technique to calculate the average free energy across the blocks and the error as a function of block size. Specifically, I calculated the error every 10 blocks from one to 1000 blocks.

For the 2D energy landscapes, I used a reweighting protocol to compute the Boltzmann distribution of CVs different from those biased in the original simulation. In this manner I reconstructed the free energy surface (FES) as a function of both the original biasing CV (the interhelix distance d) and of a CV corresponding to the interhelix crossing angle (Ω). Additionally, I reconstructed the FES as a function of other commonly used descriptors for helix dimer arrangements. These descriptors were the rotation around each helix and the N- and C-termini distances that I have already described in the “Order

parameters" section. For the FES, I calculated these order parameters with Plumed, as well as their histograms and the conversion to free energy.

4.3 Results

The transmembrane domain of Receptor Tyrosine Kinases (RTKs) is responsible to transmitting the activation signal from the extracellular (EC) to the intracellular (IC) compartment. For the TrkA receptor, the arrangement of the TM helices in the homodimeric state of the receptor has been solved with NMR.[60] Mutagenesis studies suggest that there are two functional interfaces of the TM helices that correspond to the active and inactive states of the receptor, with a swapping between the two taking place with a rotation of the helices.[60] For the TrkB receptor, the TM domain has been identified as a drug target.[91] All this evidence showcases the importance of the TM domain, and thus I examined with Molecular Dynamics (MD) simulations the most favorable TM helix arrangements.

In order to decide which level of molecular representation is more suitable to efficiently sample all the relevant TM arrangements, I compared the sampling ability of all-atom (or atomistic) and coarse-grained (CG) MD simulations. In all-atom simulations, as indicated by the name, all the atoms of the molecules are explicitly represented by the force field, whereas in CG simulations several atoms are grouped together and represented as one bead. I used the Martini force field, which models four heavy atoms as one bead, because this force field is extensively used in biomolecular simulations and has been successfully used before to correctly capture the arrangements of several TM helical dimers.[235,237–239,304] Both levels of molecular representation have their own strengths and weaknesses. Atomistic force fields capture more accurately the physicochemical and steric characteristics of the molecules, but since they model all the atoms explicitly, they are computationally expensive and one might need to simulate long times to achieve adequate sampling of all important configurations. On the other hand, CG force fields miss some chemical features, such as the specificity of hydrogen bonds, but they make long simulations computationally tractable because of the small number of modeled atoms.

In this respect, I simulated the TM helical homodimers of TrkA and TrkB receptors in self-assembled DPC micelles, which is the lipid environment used for the TrkA NMR structure, with atomistic and CG simulations, using the CHARMM36m and Martini 2.2 force fields, respectively. I ran 10 replica simulations for each case, with 1 μs per replica for the atomistic simulations and 20 μs per replica for the CG simulations. In order to evaluate the sampling of different configurations, I plotted the Kernel

Density Estimates (KDE) of the distributions of six order parameters that describe the various TM arrangements. These parameters were (1) the distance between the centers of geometry (COG) of the two helices, (2) the crossing angle between the two helices, (3) the phase and position which correspond to the rotation around each helix, and (4) the distances of the N- and C-termini of the two helices. The results are presented in Figure 4.3 and Figure 4.4 for TrkA and TrkB, respectively.

Comparison of the maxima that occur in the atomistic and CG simulations shows that the order parameter space that is sampled in the atomistic simulations is more restricted than the one in the CG simulations. This is evident by the sampling of only left-handed arrangements for the TM dimer, which are indicated with positive crossing angles, in the atomistic simulations, in contrast to both left- and right-handed arrangements in the CG simulations. This reduced sampling in the all-atom level leads to the system visiting mostly configurations close to the initial NMR structure, as indicated by the maxima near the NMR spot in Figure 4.3 and Figure 4.4.

Additionally to the restricted sampling, the atomistic simulations seem to also lead to arrangements with high crossing angles, i.e. more than 60^{o}, which are probably not biologically relevant as they could not occur in the membrane environment. In contrast to this, CG simulations sampled arrangement that could also occur in membranes, with crossing angles between -60^{o} and 60^{o}. This indicates that the CG simulations are more adequate for sampling the relevant TM helix arrangements than the atomistic ones.

In the coarse-grained level, various different arrangements occurred for TrkA, which differ from the NMR structure. One dominant arrangement was more distinct in which the two helices interacted through a different interface than the NMR one (purple versus orange helices in Figure 4.3). In fact, this dominant arrangement is right-handed compared to the left-handed NMR arrangement. Also, the smaller crossing angle of the dominant dimer at -30^{o} compared to 40^{o} for the NMR, and the smaller N- and C- termini distances suggest more parallel TM arrangement of the helices.

For the case of the TrkB receptor, three dominant configurations were sampled (Figure 4.4). One of them corresponds to a left-handed dimer, which is quite close in the order parameter space to the initial model that I built, for which I used the TrkA NMR structure as a template. The other two are right-handed dimers with a difference in the interhelical distance of ~2 Å.

The comparison between all-atom and CG simulations revealed that CG simulations are more efficient in sampling the biologically relevant configurations of the TM helices. However, the

distributions of the TrkA TM homodimer sampled during the CG simulations indicate that the NMR structure is not very stable, as the point corresponding to the NMR structure is not localized in a density maximum in Figure 4.3. In the atomistic simulations, the NMR point is closer to density maxima but not exactly on them. Visualization of the dimer arrangements that were sampled during the atomistic simulations revealed various different configurations (Figure A.14), and thus the NMR structure does not seem to be stable in atomistic simulations either. This led to the need to examine the influence of the initial conditions of the simulations in the stability of the NMR structure and the sampled arrangements.

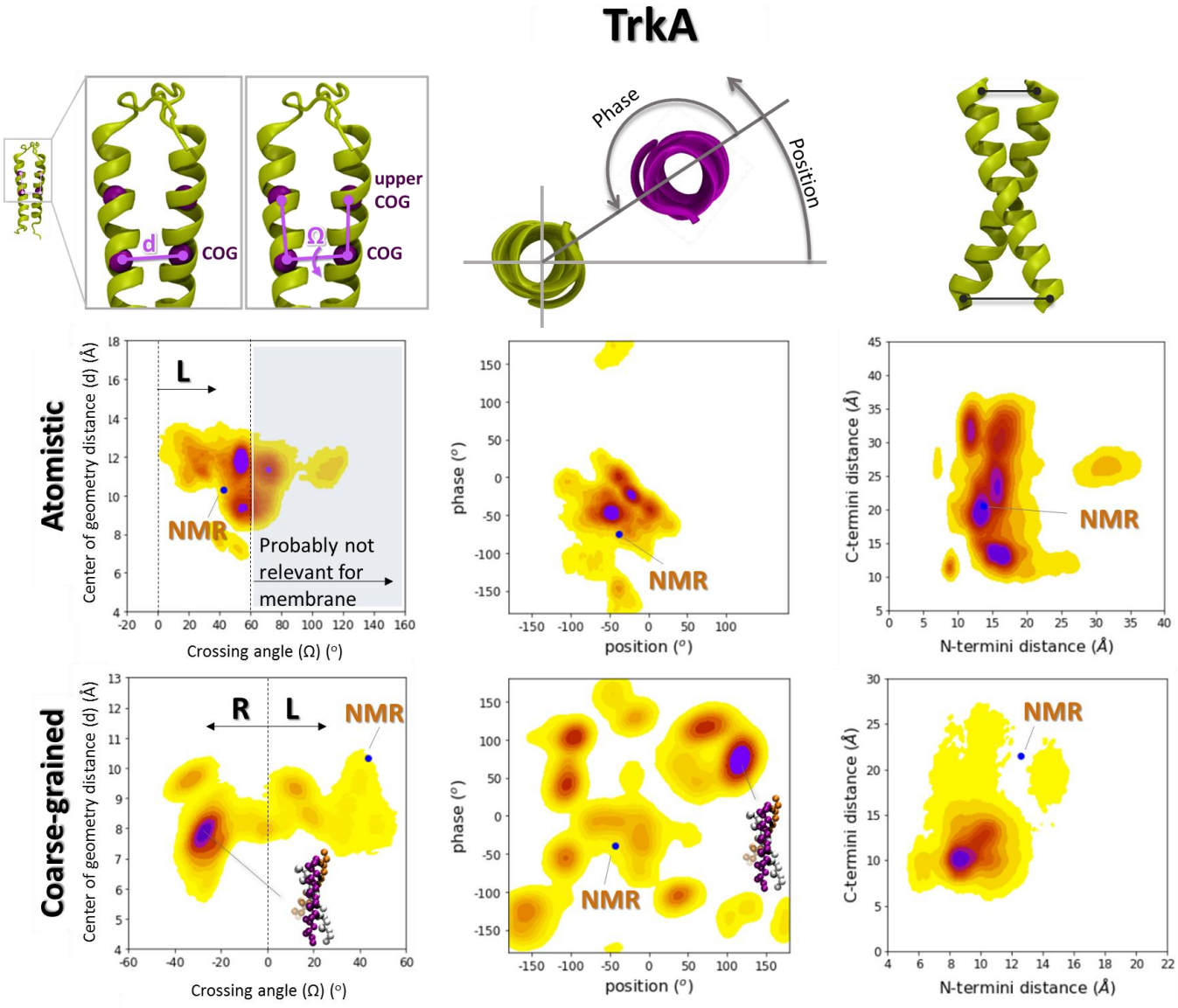


Figure 4.3: Comparison of sampling in atomistic and coarse-grained simulations for the TrkA TM homodimer. I calculated the KDE plots for six different order parameters: interhelical distance (d), interhelical crossing angle (Ω), phase (rotation around 1st helix), position (rotation around 2nd helix), N-

and C- termini distances. The parameters are also visualized schematically above the plots. Right- and left-handed arrangements are indicated with (R and L). The NMR state is shown with a blue dot. In the coarse-grained simulations, I have visualized the dominant sampled state.

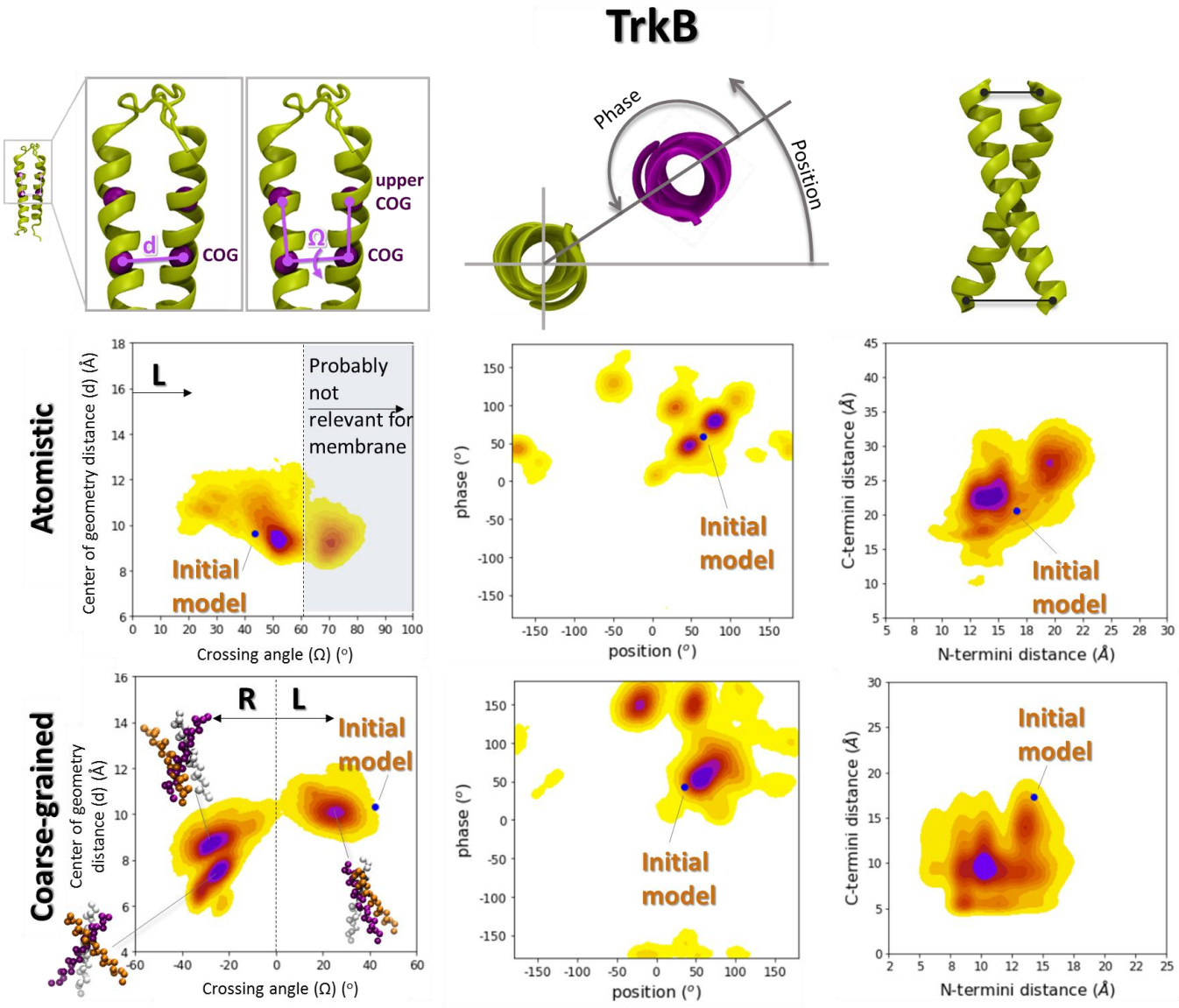


Figure 4.4: Comparison of sampling in atomistic and coarse-grained simulations for the TrkB TM homodimer. I calculated the KDE plots for six different order parameters: interhelical distance (d), interhelical crossing angle (Ω), phase (rotation around 1st helix), position (rotation around 2nd helix), N- and C- termini distances. The parameters are also visualized schematically above the plots. Right- and

left-handed arrangements are indicated with (R and L). The NMR state is shown with a blue dot. In the coarse-grained simulations, I have visualized the dominant sampled state.

I performed test simulations with a different ionic strength than 150 mM, which corresponds to physiological conditions, to 24 mM which, as indicated in the PDB entry of the TrkA NMR structure (PDB ID: 2N90),[60] was used in the experimental conditions. The results are presented in Figure A.15, System 3. I also examined a different detergent-to-protein molar ratio (DPR) than the 50:1 that is stated in the NMR publication.[60] Specifically, I modeled a DPR of 40:2 in agreement with the PDB entry (Figure A.15, System 4). Finally, I tested the importance of starting from an equilibrated micelle. Hence, I simulated the self-assembly of DPC lipids, while the proteins were allowed to freely move (Figure A.15, System 5). Comparison of the configurations that were sampled in these tests (Figure A.15) and in the initial CG simulations in micelles (Figure 4.3) reveals very similar order parameter distributions. This means that the arrangements that are sampled do not depend so much on the simulation conditions, which enhances the confidence in the results of the initial protocol (Figure 4.3). What is more, the NMR structure is not stable in any of the tests and this raises questions about the conditions under which it was solved.

As a final evaluation of the CG simulation protocol and force field, I simulated two model systems for TM helix dimerization: those of the Ephrin type-A receptor 1 (EphA1) and the Glycophorin A (GpA) TM homodimers. The results shown in Figure A.16 reveal that NMR structures of the two proteins are quite stable during the simulations and they correspond to density maxima. The sampled arrangements are also similar between Martini 2.2 and the newer Martini 3 force field, which suggests the robustness of the simulation protocol.

As mentioned before, mutagenesis studies have revealed two interfaces; an active and an inactive, the latter of which corresponds to the NMR structure.[60] A visualization of the residues that form the two interfaces is shown in Figure 4.5. Interestingly, both interfaces are closer to the N-terminus, where there are three GxxxG-like motifs, which might facilitate the rotation of the helices.

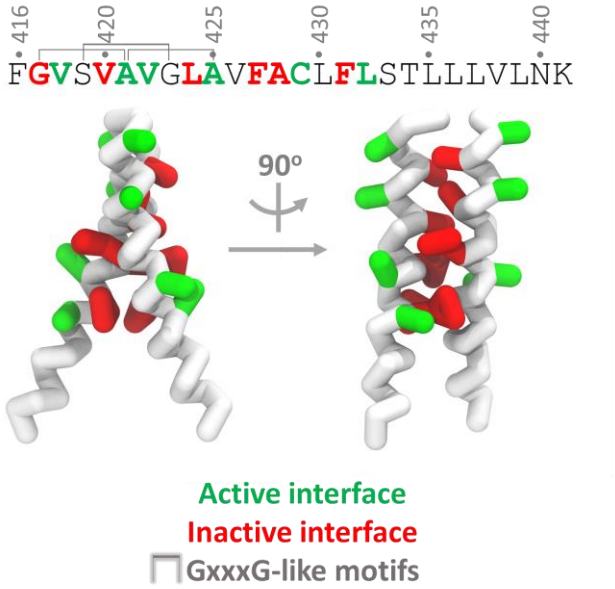
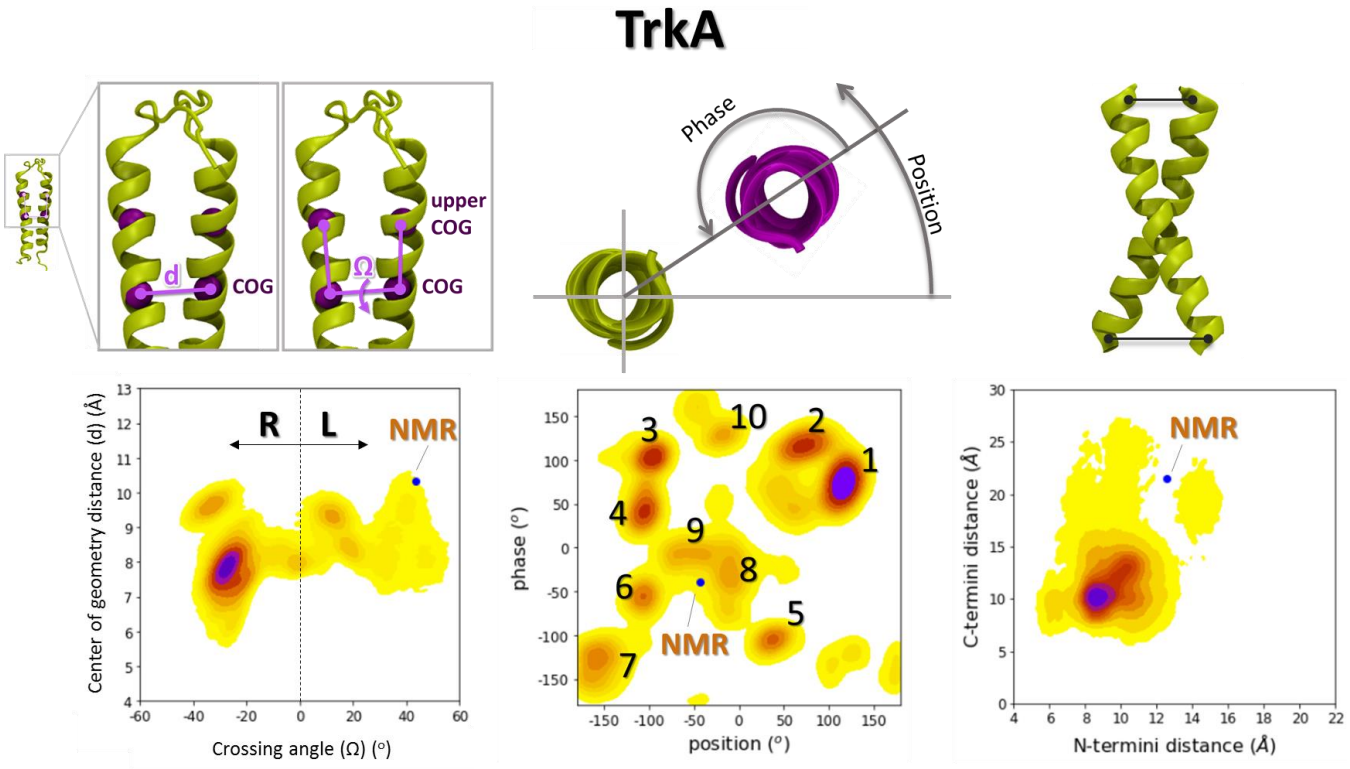


Figure 4.5: TrkA TM NMR structure (PDB ID: 2N90) in CG representation level. The residues of the active and inactive interfaces are shown in green and red, respectively. The three GxxxG-like motifs are also indicated in the sequence.

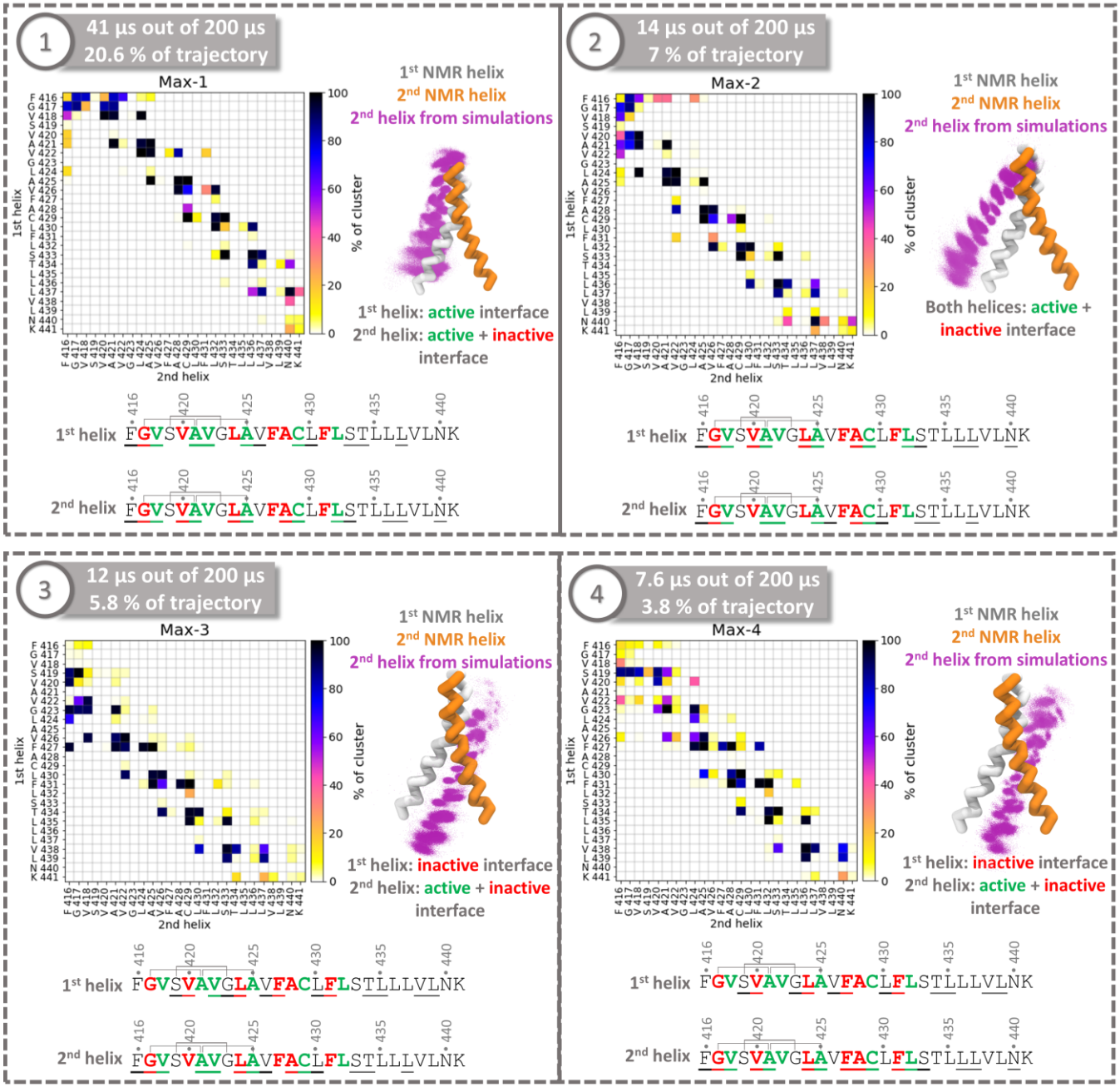
As a next step, I analyzed the interfaces of the TrkA TM dimer from the CG simulations. Root mean squared deviation (RMSD) and crossing angle calculations revealed differences from the initial structures for both TrkA and TrkB (Figure A.17 and Figure A.18). The various arrangements are clearly indicated by the many density maxima that occurred in the phase and position 2D heatmaps (Figure 4.3, CG and Figure 4.6, upper plots). This is because the phase and position are equivalent to the rotation around the first and second helix and thus the various maxima correspond to different dimer interfaces.

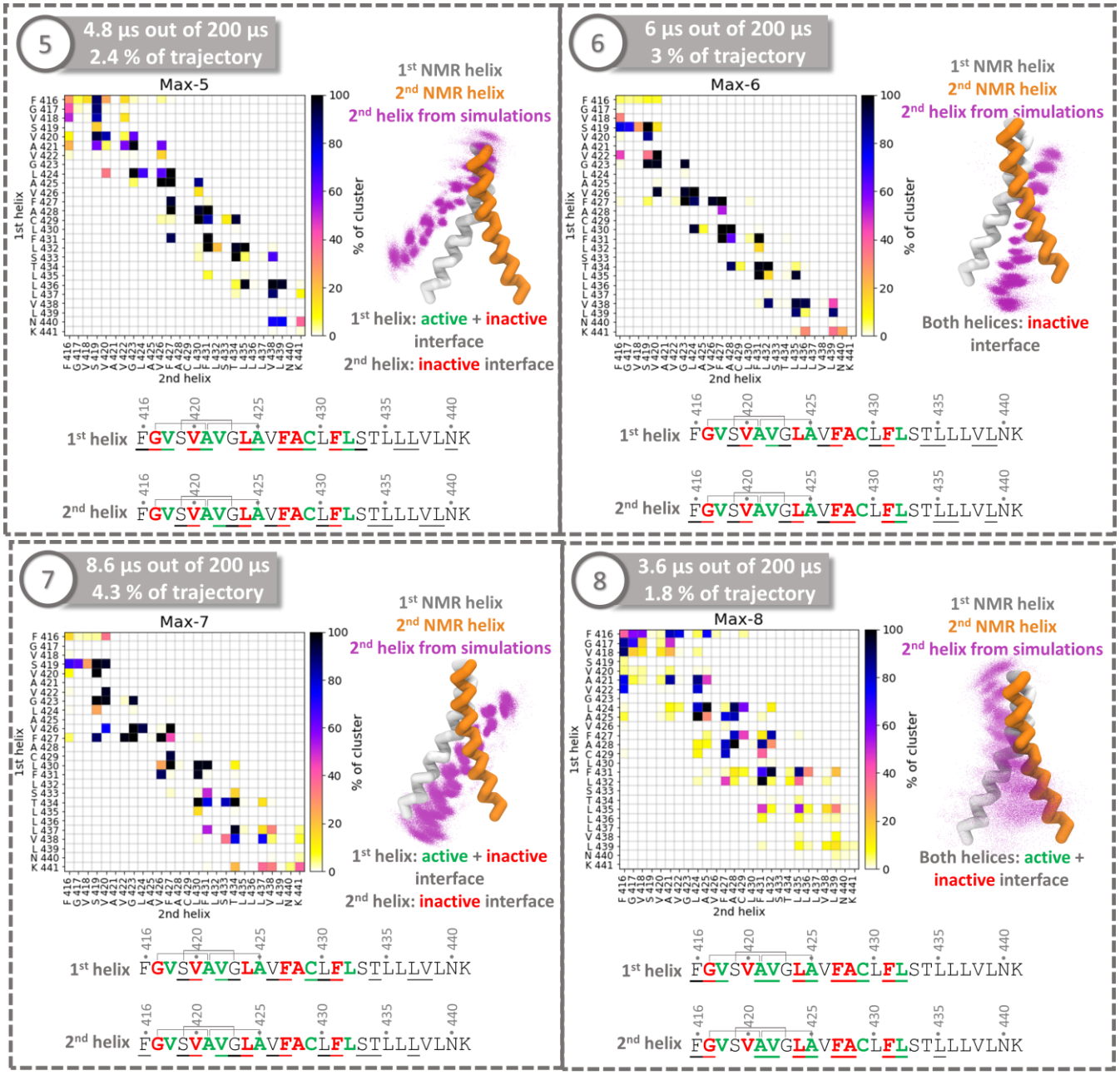
To investigate the interfaces that occurred and the protein residues that participate in them, I extracted the frames from each of the maxima and performed a contact analysis between the helical monomers. The results are presented in Figure 4.6. This analysis shows that the biggest cluster (1st maximum) represents about 20% of all the replica trajectories. I calculated the contacts between the two helices, by defining a contact when two residues are within 6 Å distance. Then I examined the contacts and I underlined the residues that interact the most in the sequences shown in Figure 4.6. For

the 1st and most populated cluster, the 1st helix interacts with the 2nd one mostly through residues that have been identified to be part of the active interface of the dimer, while the 2nd helix interacts through both active and inactive interface residues. However, since the active and inactive interface residues are very close to each other, it seems that some of the inactive residues could even participate in the active interface. Thus, the helix arrangements from the 1st maximum could correspond to the active state of the TrkA receptor. Interestingly, this is a right-handed dimer in contrast to the left-handed NMR structure.



4 Transmembrane domain dimerization mechanism




5
4.8 µs out of 200 µs
2.4 % of trajectory
Max-5
1st NMR helix
2nd NMR helix
2nd helix from simulations
1st helix: active + inactive interface
2nd helix: inactive interface
% of cluster
1st helix
2nd helix
1st helix FGVSVAVGLAVFACLFLSTLLLVLNK
2nd helix FGVSVAVGLAVFACLFLSTLLLVLNK
6
6 µs out of 200 µs
3 % of trajectory
Max-6
1st NMR helix
2nd NMR helix
2nd helix from simulations
Both helices: inactive interface
1st helix FGVSVAVGLAVFACLFLSTLLLVLNK
2nd helix FGVSVAVGLAVFACLFLSTLLLVLNK
7
8.6 µs out of 200 µs
4.3 % of trajectory
Max-7
1st NMR helix
2nd NMR helix
2nd helix from simulations
1st helix: active + inactive interface
2nd helix: inactive interface
1st helix FGVSVAVGLAVFACLFLSTLLLVLNK
2nd helix FGVSVAVGLAVFACLFLSTLLLVLNK
8
3.6 µs out of 200 µs
1.8 % of trajectory
Max-8
1st NMR helix
2nd NMR helix
2nd helix from simulations
Both helices: active + inactive interface
1st helix FGVSVAVGLAVFACLFLSTLLLVLNK
2nd helix FGVSVAVGLAVFACLFLSTLLLVLNK

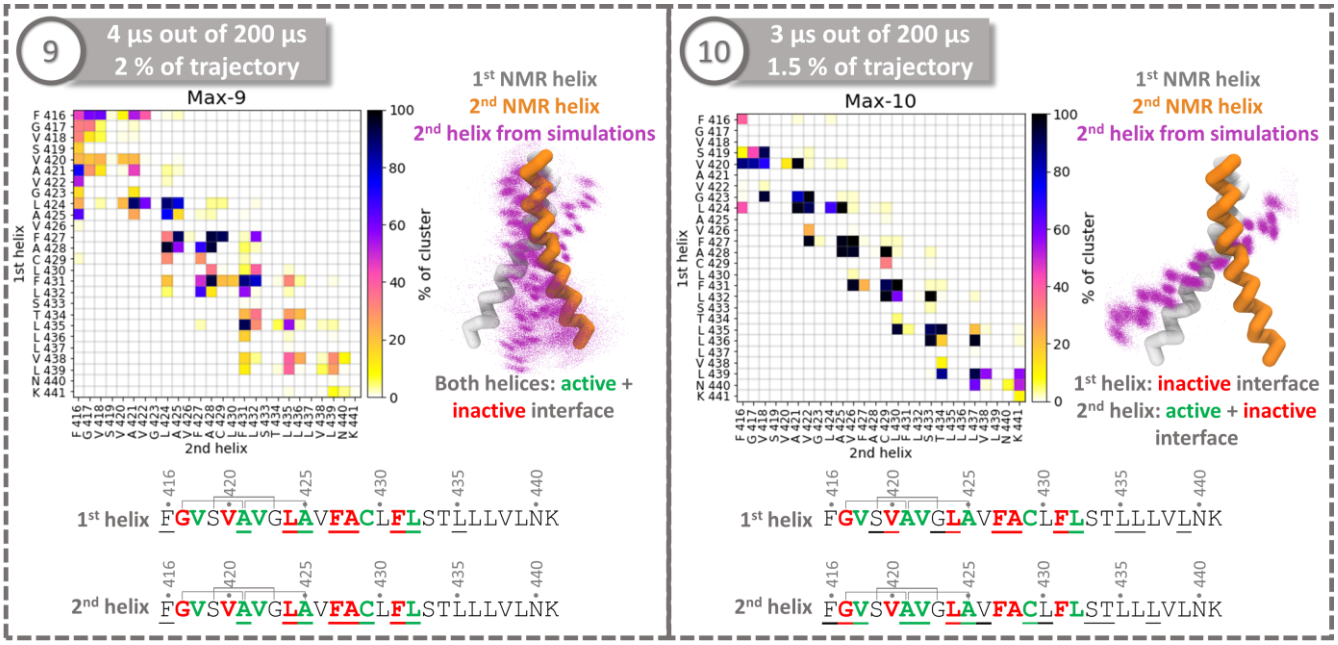
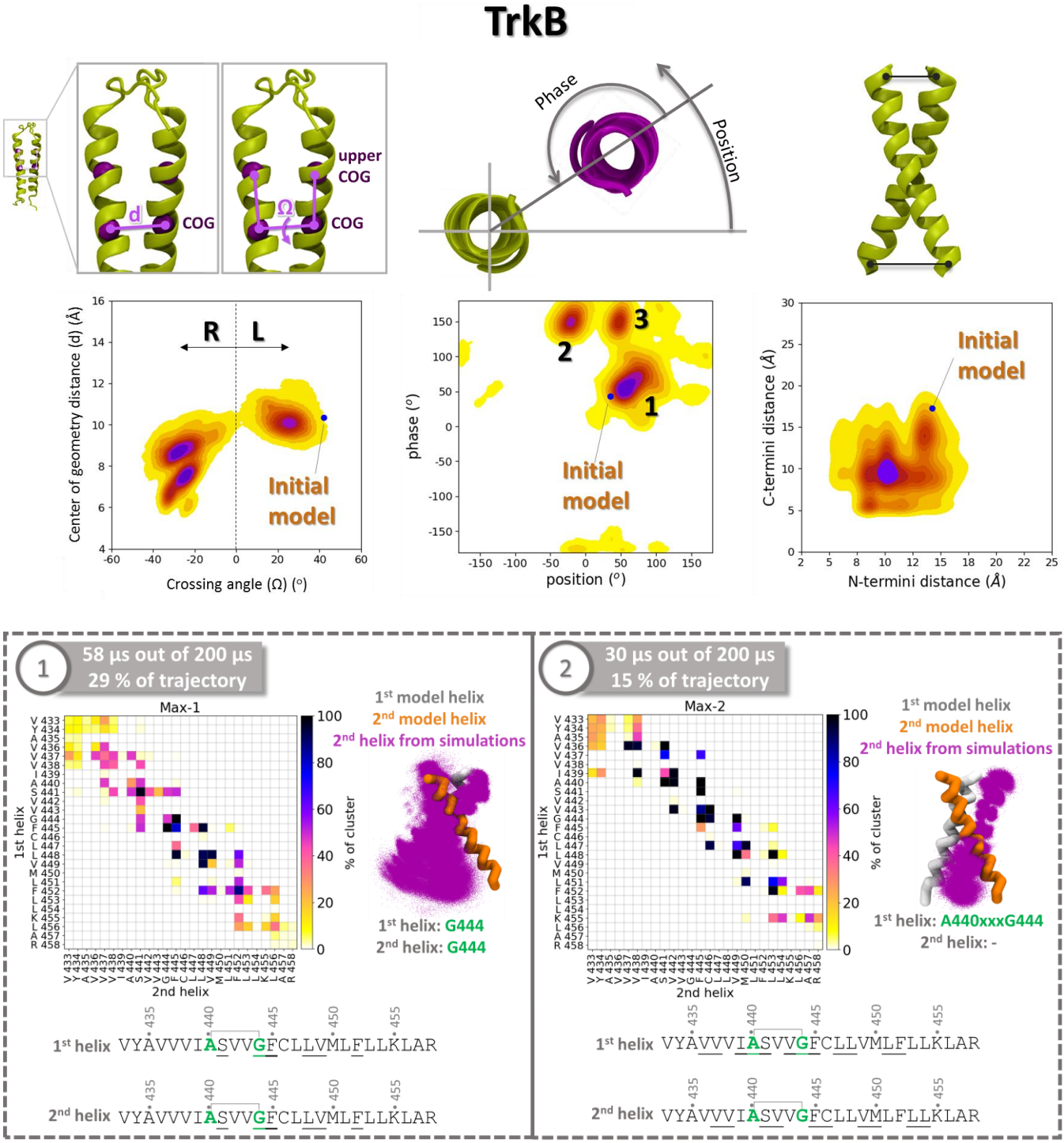


Figure 4.6: Analysis of ten TrkA arrangements from the CG simulations in micelles. The maxima are numbered in the phase-position 2D heatmap and then analyzed in each subsequent box. In each box, I have calculated the percentage of the trajectory that was spent in this maximum, the contact map between the two helices and the residues that interact the most from each helix are underlined in the two sequences on the bottom. I have also visualized the arrangements by aligning all the frames to the 1st helix and then displaying the backbone of the 1st helix in white, the 2nd helix from the NMR structure in orange, and the 2nd helix from the simulation frames in purple dots.

The other arrangements sampled in the remaining density maxima correspond to both active and inactive interfaces, with almost all arrangements being a combination of the two. Only, the 6th arrangement involves only inactive interface residues. Moreover, all TM dimer configurations are right-handed dimers, except for the 8th and 9th maxima configurations. This indicates that the right-handed arrangements are favored over the left-handed ones, and this explains all the high maximum at ~-30° crossing angle in the d-Ω 2D heatmap.


TrkB
d
COG
upper
COG
Ω
COG
Phase
Position
Center of geometry distance (d) (Å)
R
L
Initial
model
Crossing angle (Ω) (°)
phase (°)
3
2
1
Initial
model
position (°)
C-termini distance (Å)
Initial
model
N-termini distance (Å)
1
58 µs out of 200 µs
29 % of trajectory
Max-1
1st model helix
2nd model helix
2nd helix from simulations
1st helix
2nd helix
% of cluster
1st helix: G444
2nd helix: G444
1st helix VYAVVVIASVVGFCLLVMLFLLKLAR
2nd helix VYAVVVIASVVGFCLLVMLFLLKLAR
2
30 µs out of 200 µs
15 % of trajectory
Max-2
1st model helix
2nd model helix
2nd helix from simulations
1st helix
2nd helix
% of cluster
1st helix: A440xxxG444
2nd helix: -
1st helix VYAVVVIASVVGFCLLVMLFLLKLAR
2nd helix VYAVVVIASVVGFCLLVMLFLLKLAR

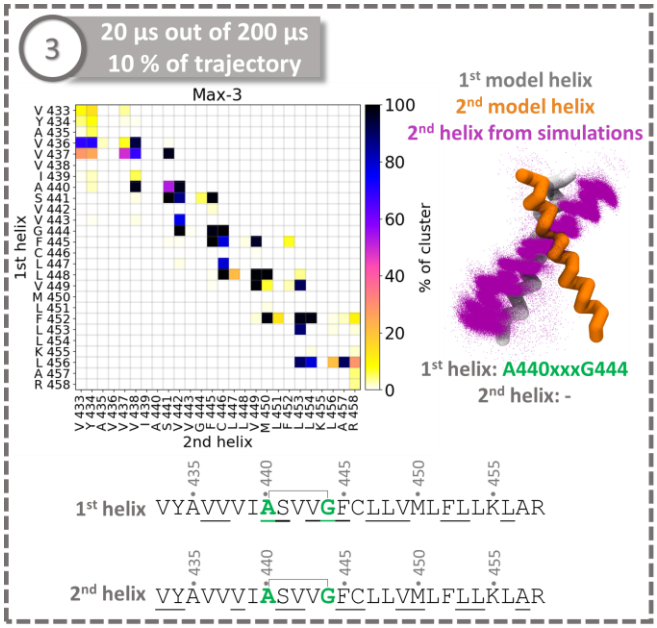


Figure 4.7: Analysis of three TrkB arrangements from the CG simulations in micelles. The maxima are numbered in the phase-position 2D heatmap and then analyzed in each subsequent box. In each box, I have calculated the percentage of the trajectory that was spent in this maximum, the contact map between the two helices and the residues that interact the most from each helix are underlined in the two sequences on the bottom. I have also visualized the arrangements by aligning all the frames to the 1st helix and then displaying the backbone of the 1st helix in white, the 2nd helix from the NMR structure in orange, and the 2nd helix from the simulation frames in purple dots.

I also analyzed the arrangements of TrkB dimer from the CG simulations. As for TrkA, I extracted the frames that correspond to each of the maxima in the phase-position 2D heatmap, because they indicate different rotation angles, and thus different interfaces. Based on this, I extracted the configurations from the three maxima in Figure 4.7 phase-position diagram. The 1st and most populated maximum accounts for 29% of the trajectory of all replicas together. In contrast though to TrkA, for which the phase-position density plot distinguishes well the various arrangements, two different helix arrangements of TrkB are localized in the 1st maximum, as shown from the visualization of the cloud of extracted frames. In these two arrangements, the helices interact through the G444 of the A440xxxG444 motif, and only around 50% of the cluster interacts through the whole motif. The 2nd and 3rd maxima correspond to specific arrangements each and the 1st helix in both of them interacts with the 2nd one through the A440xxxG444 motif. Because the phase-position 2D heatmap could not distinguish the two arrangements in the 1st

maximum, I also extracted the frames from the maxima in the interhelical distance-crossing angle 2D heatmap to examine whether these descriptors distinguish better the various arrangements (Figure A.19). However, the 1st maximum corresponds to even more arrangements. The first three maxima configurations interact through the AxxxG motif only in one helix or just through the G444 residue. Interestingly, in the 4th maximum both helices interact through the whole AxxxG motif. This might correspond to the active state of the receptor, with the rest of the arrangements being intermediate states between the active and inactive ones.

The various TM arrangements that were sampled in the CG simulations of TrkA and TrkB raise the question of whether the sampling is complete. To examine this, I performed enhanced sampling CG well-tempered metadynamics (CG-MetaD) simulations. In these type of simulations, Gaussian potentials are deposited very several steps on the free energy of a collective variable (CV), allowing this way the CV to explore all possible values. For these TM systems, I used the interhelical distance, d, as a biasing CV to explore both bound states of the TM dimer (low d values) and unbound ones (high d values). Metadynamics allows also to estimate free energy differences between two states along the CV space. In order to investigate the convergence of the CG-MetaD simulations, I performed a series of tests. These were the time evolution of (1) the diffusion of the biasing CV (d), (2) the decrease of Gaussian hills, (3) the free energy profile along the CV and (4) the free energy difference between bound and unbound states. I also examined the free energy error as a function of simulation blocks of increasing size.

The convergence analysis for TrkA is presented in Figure 4.8. Examination of the evolution of the biasing CV (d) over time shows a diffusive behavior, with many bound (d < 1.5 nm) and unbound (d > 1.5 nm) states being sampled, which is necessary for convergence (Figure 4.8A). In well-tempered metadynamics, the heights of the Gaussian hills decrease over time, which is an indication of convergence. In the case of TrkA, the Gaussian height does decrease over time, but it does not become zero for a long time. This indicates that the system is probably close to convergence, but it has not fully converged yet (Figure 4.8B). This is also shown in the evolution of the free energy in the CV space and the free energy difference between bound and unbound states over time, which seem to be very close to stabilization (Figure 4.8C,D). Similarly, the free energy error gets close to reaching a plateau (Figure 4.8E). With this sampling of the CG-MetaD simulations, the free energy difference between the bound and unbound states of the TrkA TM homodimer is predicted to be ΔG_{bind} = -54 ± 0.33 kJ/mol (-12.9 ± 0.08 kcal/mol). The experimentally measured binding free energy is -1.9 ± 0.2 kcal/mol,[60] which suggests that the binding is over-estimated, probably because of the sticky nature of the Martini 2.2 force field.[305,306]

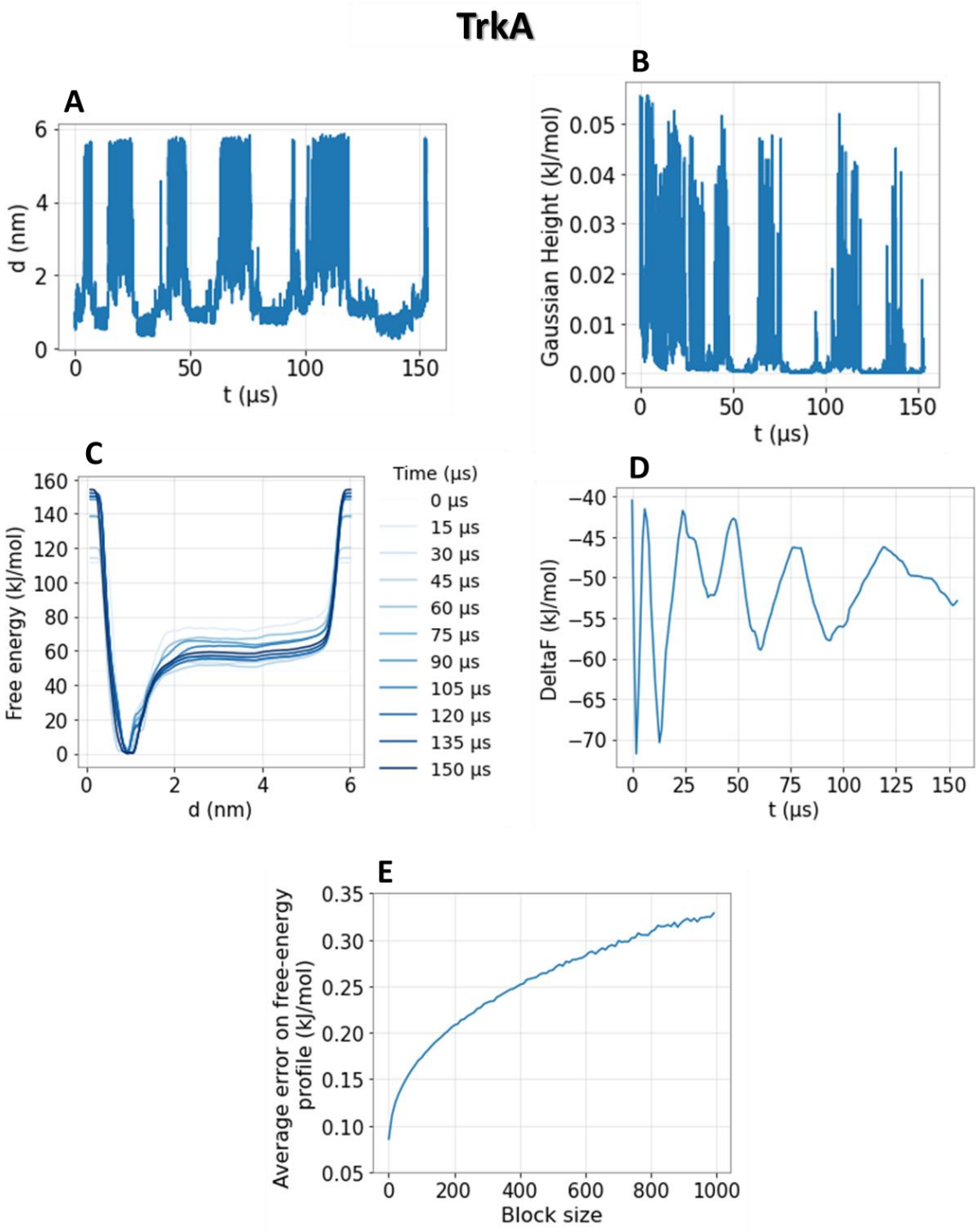


Figure 4.8: Convergence evaluation of CG-MetaD simulations on TrkA-TM dimer. (A) Time evolution of biasing CV (interhelical distance, d). (B) Time evolution of Gaussian height. (C) Time evolution of free energy along CV space. (D) Time evolution of free energy difference (ΔF) between bound state (d < 1.5 nm) and unbound state (d > 1.5 nm). (E) Free energy error as a function of simulation block sizes.

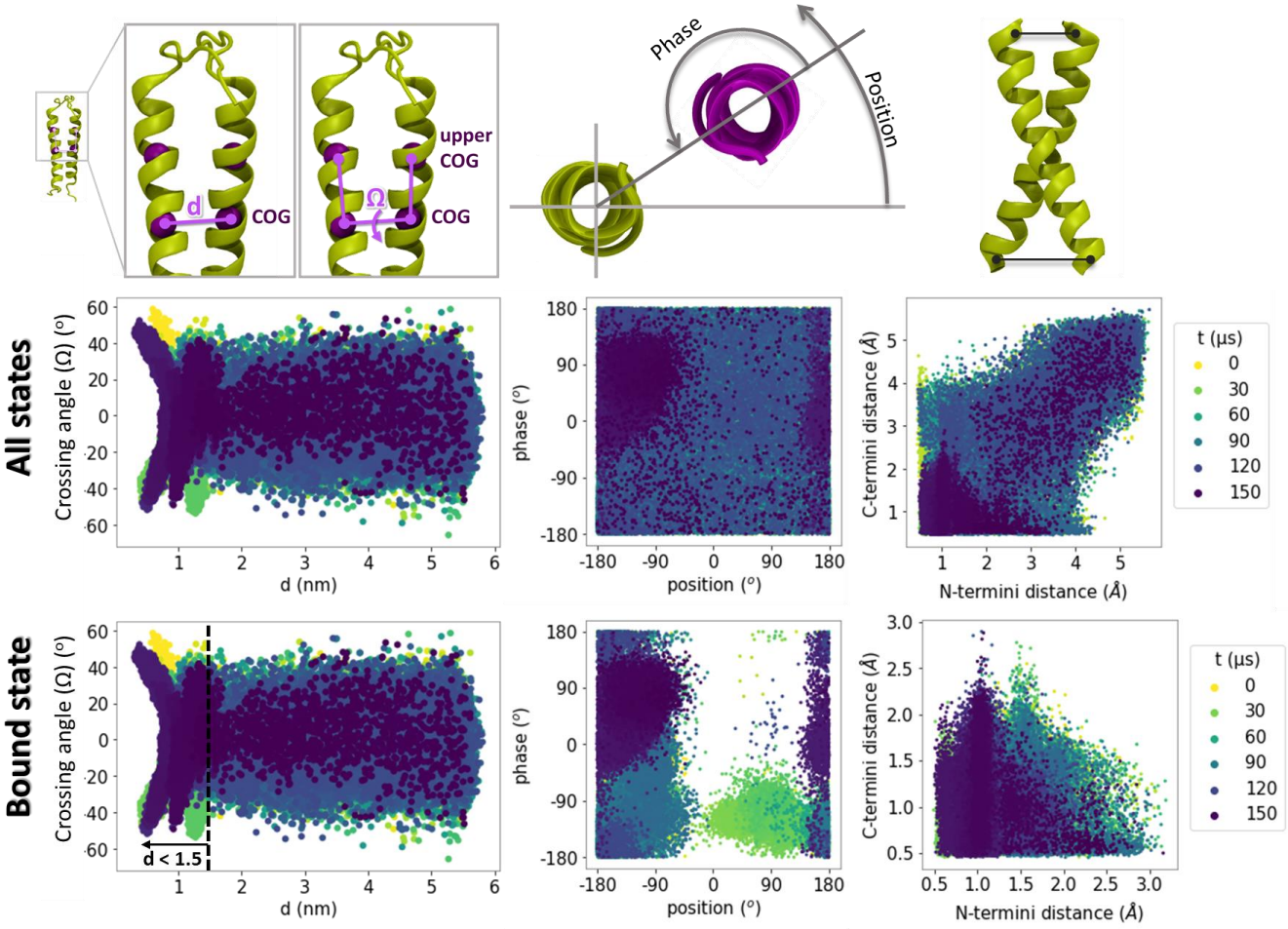


Figure 4.9: Time evolution of the six order parameters that describe the arrangements of the TrkA-TM helices during the CG-MetaD simulations. The values that correspond to the bound state (d < 1.5 nm) of the helices and those from all states (both bound and unbound) are plotted in the two rows of the figure.

Apart from the biasing CV, which is the interhelical distance (d), as mentioned before, there are some additional order parameters that are needed to describe the helix arrangements: the interhelical crossing angle, the phase and position (i.e. the rotation around each helix) and the N- and C-termini distance (Figure 4.9). The plots show good exploration of all order parameter space during the CG-MetaD simulation. When considering only the bound state, specific regions of the phase-position space seem to be preferred, which probably corresponds to specific arrangements in the bound state.

The above analysis is the basis for examining the free energy surface (FES) along those order parameters that were not biased during the simulation. In order to retrieve the FES from non-biasing CVs, I performed reweighting of the simulation to remove the bias from the simulation. This way, a

weight is assigned to each frame based on the bias that was deposited in that frame. Thus, the arrangements that were sampled in the various frames contribute differently in the FES. In order to investigate if the reweighting was correct, I compared the free energy over the distance (d) from the deposited Gaussian hills (MetaD bias) and from the reweighting protocol (Figure 4.10A). The two profiles are the same, meaning that the reweighting was performed successfully.

With the simulation being reweighted, I could then explore the FES along different CVs from the biased one. First, I calculated the FES in the 2D space of the interhelical crossing angle (Ω) and the distance between the two helices (d) (Figure 4.10B). The FES shows that the free energy difference between bound (d < 1.5 nm) and unbound states (d > 1. nm) is around -54 kJ/mol, as already shown from the 1D FES plot. The extra dimension of the crossing angle shows that there is a region from -25° to 25° of very similar free energy (within 3 kJ/mol). This free energy minimum coincides with an interhelical distance of 1 nm.

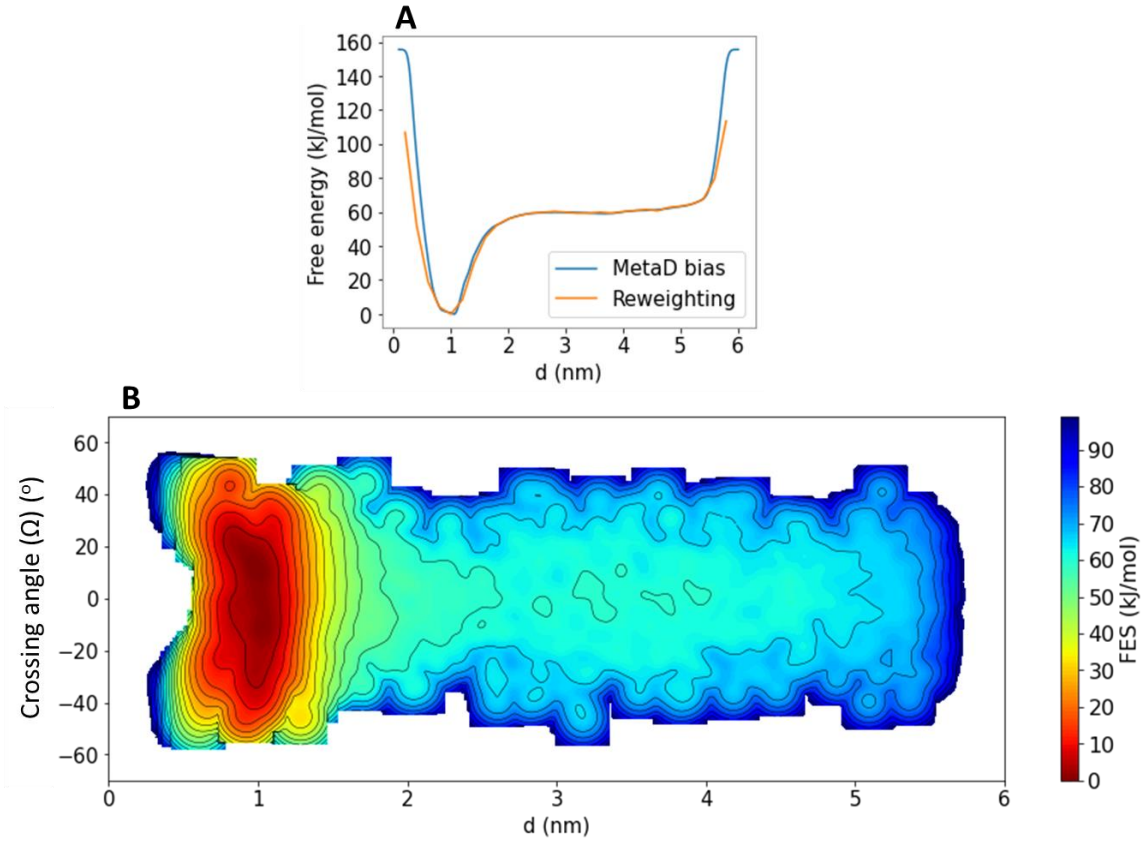


Figure 4.10: Reweighting on TrkA CG-MetaD simulations. (A) Comparison of free energy profile on biasing CV, interhelical distance, calculated from the Gaussian metadynamics bias (MetaD bias) and from

the reweighting procedure. (B) Free energy surface (FES) on crossing angle (Ω) – interhelical distance (d) space after reweighting. Isosurfaces are plotted every 3 kJ/mol.

Next, I analyzed the FES on the six order parameters, this time only for the frames that correspond to the bound state (Figure 4.11). The Ω-d FES, with the isosurfaces plotted for each kJ/mol, shows that there are two minima at ~ 10° and -10° crossing angle of the same energy. The phase-position diagram on the other hand reveals three regions of low energy, probably corresponding to three different stable arrangements.

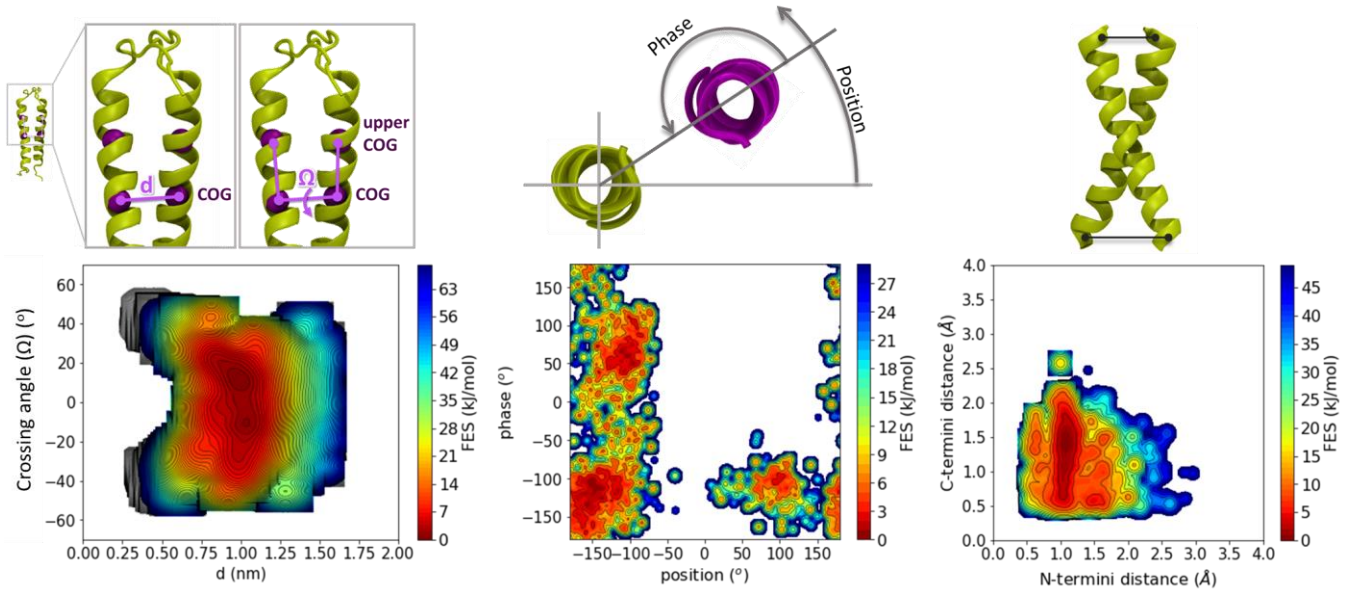


Figure 4.11: Free energy surfaces (FES) on crossing angle (Ω) – interhelical distance (d), phase-position and N- and C-termini distances after reweighting. Isosurfaces are plotted every 1 kJ/mol.

The analysis of the TrkB CG-MetaD simulations was very similar to that of TrkA (Figure 4.12). The convergence analysis showed that the TrkB system is also getting close to convergence, but the non-zero Gaussian heights show that it has not reached it yet (Figure 4.12B). However, the sampling already reveals some differences from TrkA. The TrkB TM dimer has lower binding free energy with ΔG_{bind} = -44 ± 0.30 kJ/mol (10.5 ± 0.07 kcal/mol) (Figure 4.12C,D,E), in contrast to -54 ± 0.33 kJ/mol for TrkA.

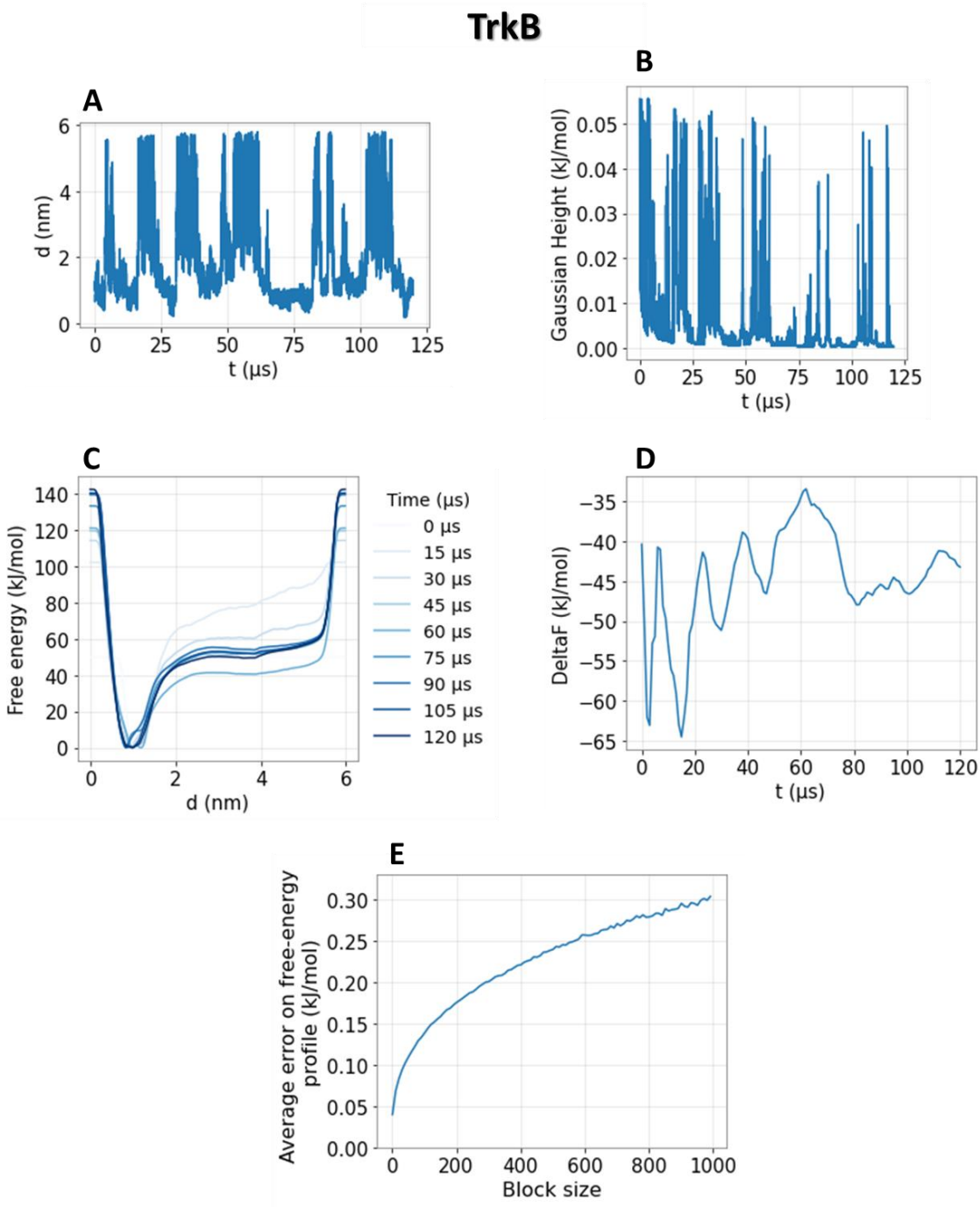


Figure 4.12: Convergence evaluation of CG-MetaD simulations on TrkB-TM dimer. (A) Time evolution of biasing CV (interhelical distance, d). (B) Time evolution of Gaussian height. (C) Time evolution of free energy along CV space. (D) Time evolution of free energy difference (ΔF) between bound state (d < 1.5 nm) and unbound state (d > 1.5 nm). (E) Free energy error as a function of simulation block sizes.

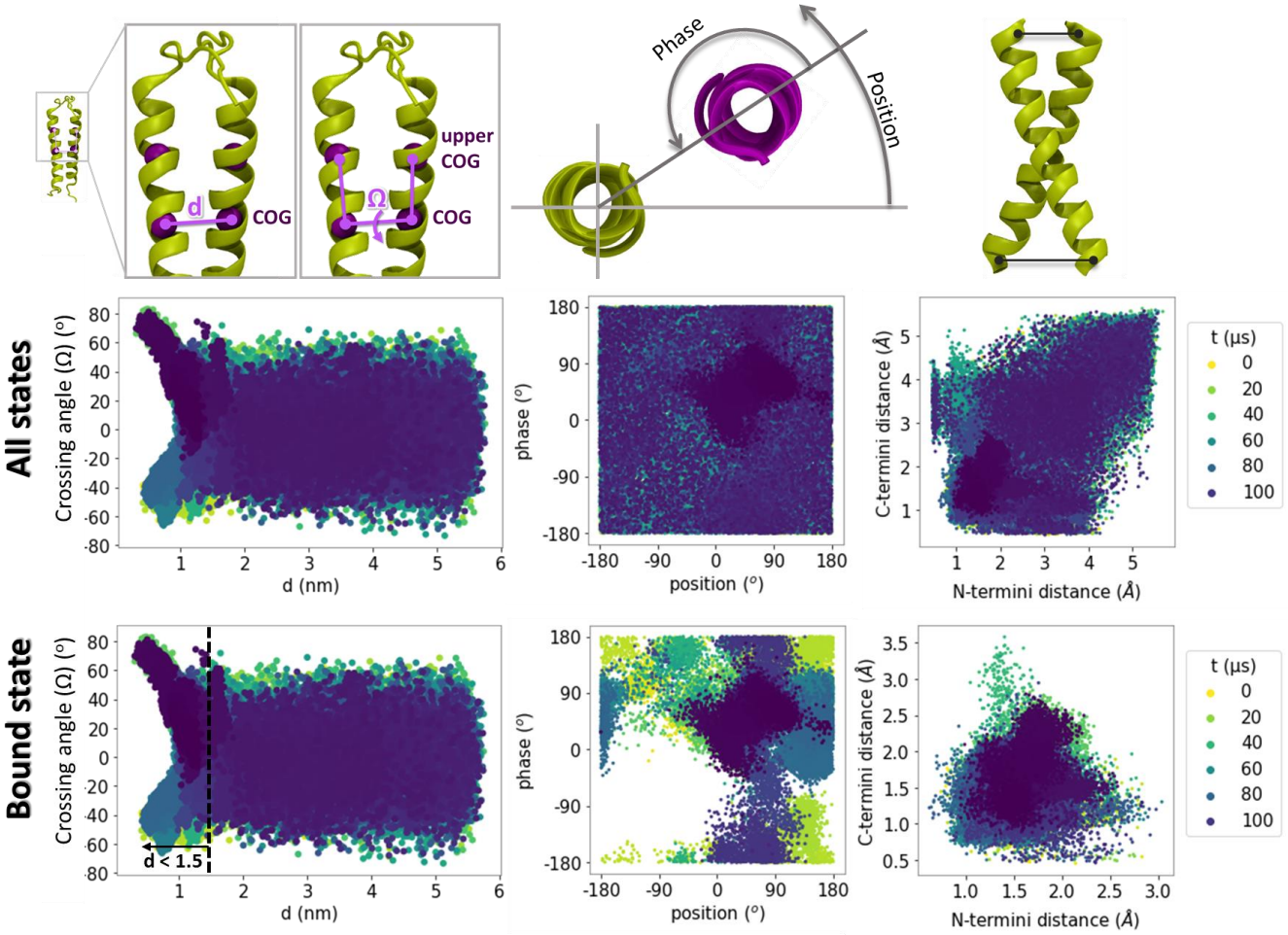


Figure 4.13: Time evolution of the six order parameters that describe the arrangements of the TrkB-TM helices during the CG-MetaD simulations. The values that correspond to the bound state (d < 1.5 nm) of the helices and those from all states (both bound and unbound) are plotted in the two rows of the figure.

As for TrkA, I also examined the exploration of the six order parameter space for TrkB. When considering only the bound state, specific regions of the phase-position space seem to be preferred, which probably corresponds to specific arrangements in the bound state (Figure 4.13). To retrieve the FES along the order parameters, I performed reweighting as for TrkA. The free energy profiles along the biasing CV were the same when calculated from the metadynamics bias and from the reweighting protocol, meaning that the reweighting was performed successfully (Figure 4.14A).

Then I calculated the FES along the order parameters. First, I calculated the FES in the 2D space of the interhelical crossing angle (Ω) and the distance between the two helices (d) (Figure 4.14B). The FES shows that the free energy difference between bound (d < 1.5 nm) and unbound states (d > 1. nm) is

around -44 kJ/mol, as already shown from the 1D FES plot. The extra dimension of the crossing angle shows that there are two main minima at crossing angles ~ 30º and -40º, with arrangements at 0º crossing angle being less favorable. This means that higher absolute values of crossing angles were sampled for TrkB than for TrkA, which might be due to the existence of only one GxxxG-like motif in the sequence of TrkB, compared to three for TrkA. The free energy minima coincide with an interhelical distance of 1 nm.

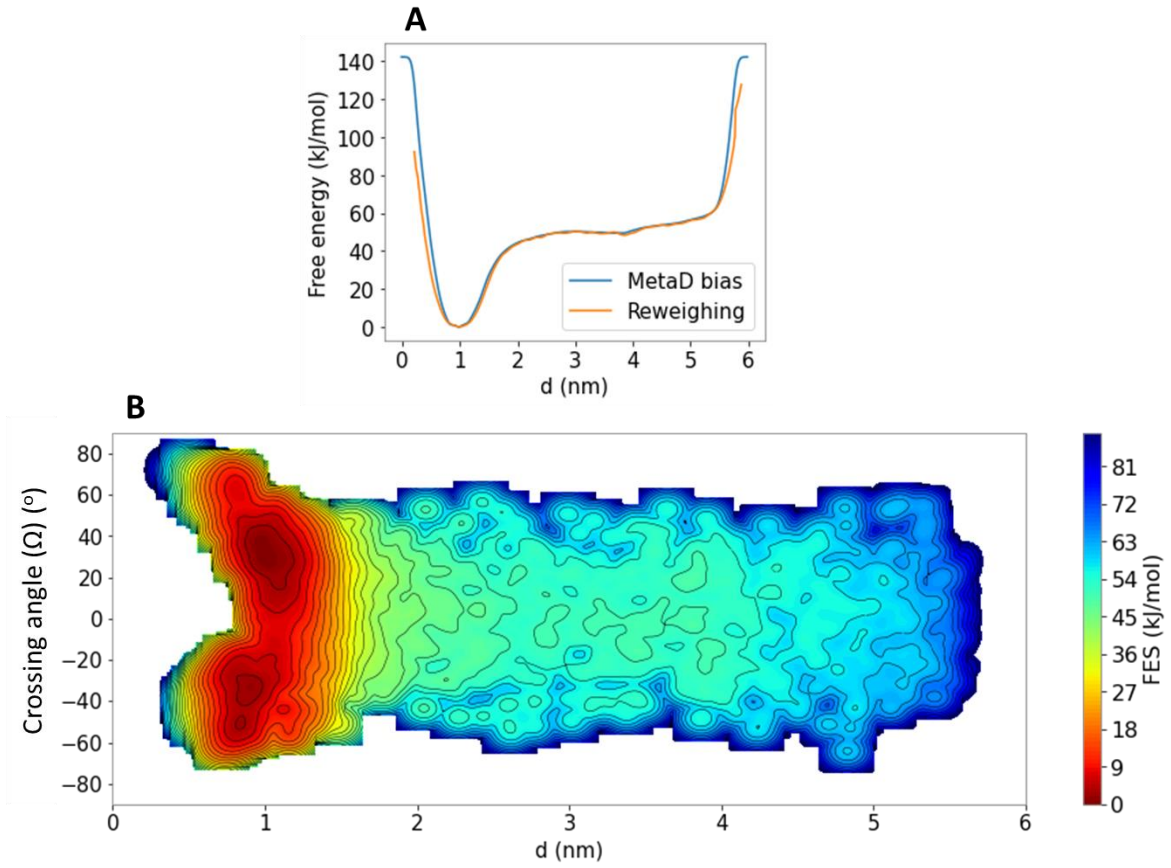


Figure 4.14: Reweighting on TrkB CG-MetaD simulations. (A) Comparison of free energy profile on biasing CV, interhelical distance, calculated from the Gaussian metadynamics bias (MetaD bias) and from the reweighting procedure. (B) Free energy surface (FES) on crossing angle (Ω) – interhelical distance (d) space after reweighting. Isosurfaces are plotted every 3 kJ/mol.

Next, I analyzed the FES on the six order parameters, this time only for the frames that correspond to the bound state (Figure 4.15). The Ω-d FES, with the isosurfaces plotted for each kJ/mol, shows that

there are two minima at ~ 30° and -40° crossing angle of the same energy. The phase-position diagram on the other hand reveals one main region of low energy, which is though spread at angles 50°-100°.

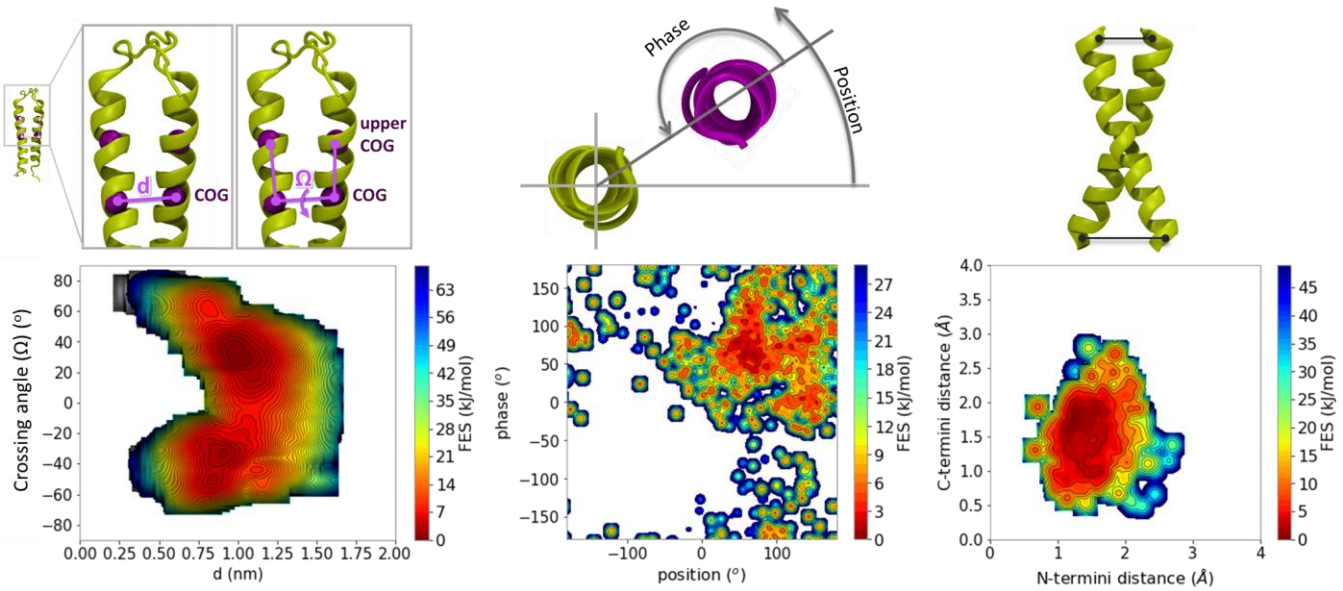


Figure 4.15: Free energy surfaces (FES) on crossing angle (Ω) – interhelical distance (d), phase-position and N- and C-termini distances after reweighting. Isosurfaces are plotted every 1 kJ/mol.

4.4 Conclusions

In this chapter of the thesis, I investigated the arrangements of the TM helical dimer of TrkA and TrkB neurotrophin receptors. For the TrkA receptor, the arrangement of the TM helices in the homodimeric state of the receptor has been solved with NMR.[60] Mutagenesis studies suggest that there are two functional interfaces of the TM helices that correspond to the active and inactive states of the receptor, with a swapping between the two taking place with a rotation of the helices.[60] For the TrkB receptor, the TM domain has been identified as a drug target.[91]

I performed atomistic and coarse-grained (CG) simulations for both TrkA and TrkB systems. Analysis of the results revealed that the sampling in the atomistic simulations is more restricted that in the CG ones, with arrangements of high crossing angles occurring in the all atom simulations, which would not be probable in the native membrane environment of the receptors. On the other hand, reasonable arrangements occurred in the CG simulations, together with more extended sampling and thus I continued with the analysis of the CG simulations. For TrkA, various different arrangements occurred

which differ from the NMR structure. Test simulations with different initial conditions showed that these arrangements are really favored irrespectively of the conditions of the simulations, thus raising questions about the conditions under which the NMR structure was retrieved.

From the sampled arrangements of TrkA, the most dominant one is right-handed helical dimer, in contrast to the left-handed NMR arrangement. The helices in this arrangement interact through residues that have been identified to be part of the active interface of the dimer,[60] which suggests that this configuration might correspond to the active state of the receptor. Understanding how the active dimer looks like can be helpful for the design of small molecules that can stabilize this arrangement and thus act as agonists of the TrkA receptor, similarly to the recent TrkB agonists.[91]

For the TrkB receptor, three maxima were sampled in the order parameter space. The most populated maximum of the phase-position heatmap contained two arrangements that interacted partially through the AxxxG motif and thus it could also correspond to the active arrangement of the TrkB receptor.

To achieve better sampling and estimate the binding free energy of the TrkA and TrkB homodimers, I performed CG metadynamics simulations. Even though the convergence of these simulations was not perfect, it demonstrated a difference in the binding free energy of the two receptors, with TrkA having a tighter binding by 10 kJ/mol. Also, the heatmaps from the metadynamics simulations have a similar profile with the unbiased CG simulations in the bound state of the receptors. This provides some confidence about the sampling achieved in the unbiased simulations.

Overall, the results presented in this chapter can provide the basis for the understanding of receptor activation in the TM segment and can be used for future drug design efforts for the development of receptor activators.

5 Dynamic behavior of full-length receptors in the neuronal membrane

5.1 Introduction to the neuronal membrane composition and its importance for the receptors

The neurotrophin receptors are physiologically found in neuronal membranes, which are known to have high concentration of cholesterol and sphingolipids, including ceramides, cerebrosides and sphingomyelin.[307–309] Experimental studies have shown that, after neurotrophin binding, the receptors translocate to ordered membrane microdomains, called lipid rafts, which have higher concentrations of cholesterol and sphingomyelin than plasma membranes.[112,114,116] These domains modulate activity of the NT receptors and are implicated in neurodegeneration.[107,111,117] This suggests that the lipid environment of the receptors is important for activity and thus in this chapter of the thesis, I created a model of a neuronal membrane. Then, in order to investigate the function of the receptors, I created models of full-length TrkA and TrkB receptor homodimers bound to the NGF and NT-4/5 NTs, respectively, and I embedded these models into the neuronal membrane. Some lipids, like GM1 and ceramide, have been found to activate the TrkA receptor.[119–122,124,125] Hence, I investigated the interactions of the receptors with the lipids of the membrane. Finally, I investigated the conformations of the proteins in their full-length and examined how these can be linked with receptor activity.

5.2 Methodology

Protein models

For the simulations of the transmembrane (TM), juxtamembrane (JM) and intracellular (IC) domains of the TrkA and TrkB monomers, I built continuous TM-JM-IC models embedded in neuronal membranes since experimental structures were only available for the TM helices and the kinase IC domains. For TrkA, I extracted the TM domain monomer from the homodimeric structure of the PDB ID: 2N90[60] NMR structure (the 1st model of the ensemble), while I built the IC kinase domain with the AutoModel class of MODELLER v.9.23[246] using the crystal structure 4F0I[72] of the TrkA IC kinase as a template. For the JM

domain, I built a part with AutoModel using the JM domain from the 5H3Q[70] crystal structure while I built the rest with the LoopModel class of MODELLER. For TrkB, I used the previously built model of the TM domain (presented in chapter 4), while for the kinase domain, I used the crystal structure 4AT5[72] as a template for AutoModel, with the activation loop and the 577-581 β sheet residues being built using the 4ASZ[72] and 4F0I crystal structures as templates, respectively. Then, I built the rest of the JM domain residues with LoopModel. In total, I built three different systems for both TrkA and TrkB, each one having a different orientation of the kinase domain with respect to the membrane plane in order to reduce bias from the input configuration.

For the full-length receptor models, I connected the extracellular (EC), transmembrane (TM) and intracellular (IC) domains with unstructured loops. The sequences of these linkers were not predicted to have any specific secondary structure from the NPSA-PRABI,[294] PSIPRED,[293] Jpred4,[292] and CFSSP[295] webservers and thus, I modeled them as random coils. I modeled the full-length systems as receptor homodimers bound to NGF and NT-4/5 neurotrophins, for TrkA and TrkB, respectively. For the EC dimers bound to NTs, I used the models already described in chapter 3, while for the TM dimers, I used the NMR structure (PDB ID: 2N90)[60] for TrkA and the model that I created for TrkB, as described in chapter 4. For the IC domain, I used the same model as for the TM-IC systems mentioned above. I built the EC and IC loops with the LoopModel class of MODELLER v.9.23. I inserted N-glycans of chemical type $Man_3GlcNAc_2$ with CHARMM-GUI[261] in the EC segments of the models at Asn 67, 95, 121, 188, 202, 253, 262, 281, 358 and 401 for TrkA, and at Asn 67, 95, 121, 178, 205, 241, 254, 338 and 412 for TrkB. These are the same glycosylation sites that were modeled in the EC domains of chapter 3, plus one additional glycan at the linker between the EC and TM domains. I built three different systems for each receptor, each one having different orientations of the kinase domains with respect to the membrane plane, to reduce bias from the input configuration.

Neuronal membrane model

I built an atomistic model of an asymmetric neuronal membrane in order to imitate the native environment of the lipid rafts, in which TrkA and TrkB reside in physiological conditions.[112,114,115] I used as reference the neuronal membrane model presented in reference [307], where the authors incorporated ~90 different lipid types into their neuronal membrane model. My aim was to reduce the number of the lipids, while maintaining the relative populations of the headgroups, as well as the overall unsaturation

of lipid tails (the relative ratios are shown in Figure 5.1). Also, within the different headgroups, I grouped lipids with similar tails into a single representative species that recapitulates the average number of unsaturated bonds per tail for that headgroup. After taking all these aspects into account, I created a model of the neuronal membrane with 19 constituents with the composition presented in Table 5.1. The percentages of each headgroup species, as well as the numbers of unsaturated bonds in the lipid tails are presented in Figure 5.1.

Table 5.1: Composition of model of neuronal membrane built for simulating full-length NT receptors. The lipid tail type is presented in the form "[C atoms of 1st tail]:[number of unsaturated bonds in the 1st tail]/ [C atoms of 2nd tail]:[number of unsaturated bonds in the 2nd tail]". The exact numbers of the lipids that were included in the outer and inner leaflets are presented in the 3rd and 4th columns. The lipids are presented in order from the most to the least abundant headgroup species. The lipid name codes are: 1,2-dipalmitoyl-sn-glycero-3-phosphocholine (DPPC), 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC), 1-palmitoyl-2-arachidonoyl-sn- glycero-3-phosphocholine (PAPC), 1,2-dipalmitoyl-sn-glycero-3-phosphoethanolamine (DPPE), 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoethanolamine (POPE), 2-linoleoyl-sn- glycero-3-phosphoethanolamine (LLPE), 1-palmitoyl-2-oleoyl-sn-glycero-3-phospho-L-serine (POPS), 1-palmitoyl-2-linoleoyl-sn-glycero-3-phospho-L-serine (PLPS), 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoinositol (POPI), 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoinositol 4,5-bisphosphate (POPI25 or PIP2), 1-stearoyl-2-arachidonoyl-sn-glycero-3-phosphoinositol (SAPI), N-stearoyl sphingomyelin (SSM), neutral sphingomyelin (NSM), ceramide 18:0 (CER180), monosialo-tetrahexosyl-ganglioside (GM1), monosialo-dihexosyl-ganglioside (GM3), galactosyl-ceramide 18:0 (GalCER180), galactosyl-ceramide 24:1 (GalCER241). The last two are cerebrosides.

Lipid	Lipid tails	Outer leaflet	Inner leaflet
Cholesterol	-	336	332
DPPC	16:0/16:0	18	12
POPC	16:0/18:1	108	56
PAPC	16:0/20:4	57	33
DPPE	16:0/16:0	27	51
POPE	16:0/18:1	21	39
LLPE	18:2/18:3	36	72
POPS	16:0/18:1	0	24
PLPS	16:0/18:2	0	48
POPI	16:0/18:1	0	15
POPI25 (PIP2)	16:0/18:1	0	15
SAPI	18:0/20:4	0	30

SSM	18:1/18:0	45	18
NSM	18:1/24:1	18	0
CER180	18:1/18:0	9	9
GM1	18:1/18:0	12	0
GM3	18:1/18:0	12	0
GalCER180	18:1/18:0	42	0
GalCER241	18:1/24:1	15	0

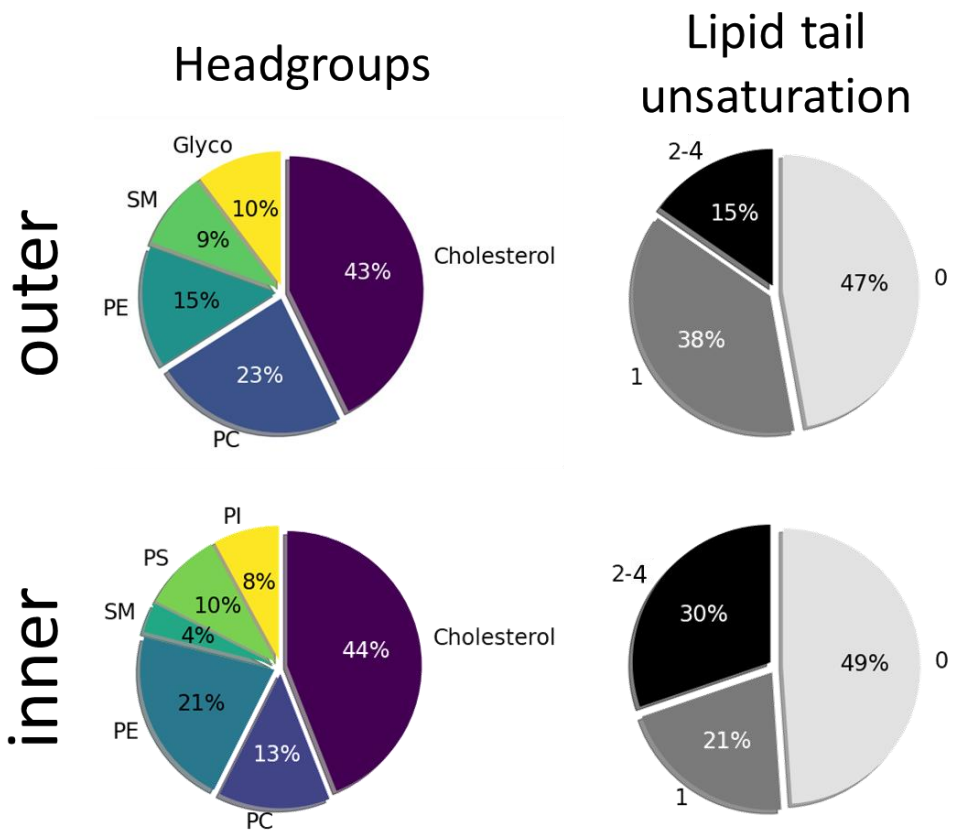


Figure 5.1: Pie charts with distributions of headgroups and numbers of unsaturated bonds in lipid tails, for the outer and inner leaflets of the model neuronal membrane. The headgroup types are cholesterol, phosphocholine (PC), phosphoethanolamine (PE), sphingomyelin (SM), glycolipids (Glyco) including GM1, GM3, GalCER180, GalCER241, phospho-L-serine (PS), phosphoinositol (PI). The right plots indicate the percentages for different numbers of unsaturated bonds in each lipid tail.

Molecular dynamics (MD) simulations

I embedded the models of the TM-IC monomers, as well as the full-length systems of TrkA and TrkB homodimeric receptors into the aforementioned neuronal membrane model with CHARMM-GUI.[310] I neutralized the charge of the periodic simulation box by adding Na^+ and Cl^- ions at 150 mM concentration. I ran all simulations with the GROMACS v.2020 MD engine,[262] using the CHARMM36m force field.[263] I used a many-step equilibration protocol, with gradual removal of restraints, to secure the correct behavior of the neuronal membrane. This was important for such a mixture of lipids, which are usually parameterized alone and not in the environment of other lipids. Thus, careful equilibration of the membrane was deemed necessary. I started each simulation with a two-step energy minimization protocol with steepest-descent and conjugate gradient, with the minimization converging once reaching a force of less than 100 kJ/(mol nm^2). Then, I performed one NVT equilibration and four NPT equilibration steps with gradual removal of restraints from protein and lipid atoms. The total equilibration time was 9.2 ns. Specifically, I ran the first NVT equilibration at 310 K for 200 ps with the restraints on the protein backbone, the side chains, the z-coordinate of the lipids, and some lipid improper dihedrals. The exact force constants of the restraints are presented in Table 5.2. For this step I used the Berendsen thermostat.[299] I ran the subsequent four (steps 4-7) NPT equilibrations at 310 K and 1 atm pressure for 1, 2, 4, 5 ns each with diminishing backbone (BB), side chain (SC) and lipid restraints, with force constants presented in Table 5.2. For all equilibration steps, I used the Berendsen thermostat and barostat. I used a coupling constant of 5 ps to maintain semi-isotropic pressure coupling with a compressibility of $4.5 \cdot 10^{-5}$. I applied a time step of 1 fs in the first NVT and NPT steps and then I increased the time step to 2 fs for the next steps. I treated the non-bonded electrostatic interactions with PME[268,269] with a cutoff distance at 1.2 nm. Non-bonded van der Waals interactions were calculated with a cutoff of 1.2 nm and a switching distance of 1.0 nm. I employed the LINCS algorithm[270] to fix the length of all hydrogen-containing bonds. I ran production simulations with the proteins and lipids completely unrestrained. I used the Nosé-Hoover thermostat[264,265] and the Parrinello-Rahman barostat.[266,267] The rest of the parameters were the same as for the equilibration steps. I simulated three replicas of the TM-IC TrkA and TrkB monomers for 1 μs each. For the full length systems, I ran 9 replicas for TrkA and 7 replicas for TrkB, of 500 ns each (Table 5.3). A rendering of the initial coordinates of the TM-IC and full-length systems is shown in Figure 5.2.

Table 5.2: Equilibration protocol of TM-IC and full-length systems in a model neuronal membrane, with gradual removal of restraints.

Step 1: minimization	10000 steps Steepest descent • Protein backbone restraints = 4000 kJ/(mol nm^2) • Protein side chains restraints = 2000 kJ/(mol nm^2) • Lipid restraints (on P atoms and only in z direction for the phospholipids, on a C atom in a similar position to P for gangliosides) = 1000 kJ/(mol nm^2) • Lipid improper dihedral restraints = 1000 kJ/mol/rad^2
Step 2: minimization	10000 steps Conjugate gradient • Protein backbone restraints = 4000 kJ/(mol nm^2) • Protein side chains restraints = 2000 kJ/(mol nm^2) • Lipid restraints (on P atoms and only in z direction for the phospholipids, on a C atom in a similar position to P for gangliosides) = 1000 kJ/(mol nm^2) • Lipid improper dihedral restraints = 1000 kJ/mol/rad^2
Step 3: equilibration (NVT)	200 ps Time step = 1 fs Thermostat = Berendsen T = 310 K • Protein backbone restraints = 4000 kJ/(mol nm^2) • Protein side chains restraints = 2000 kJ/(mol nm^2) • Lipid restraints (on P atoms and only in z direction for the phospholipids, on a C atom in a similar position to P for gangliosides) = 1000 kJ/(mol nm^2) • Lipid improper dihedral restraints = 1000 kJ/mol/rad^2
Step 4: equilibration (NPT)	1ns Time step = 1 fs Thermostat + Barostat = Berendsen T = 310 K P = 1 bar • Protein backbone restraints = 4000 kJ/(mol nm^2) • Protein side chains restraints = 2000 kJ/(mol nm^2) • Lipid restraints (on P atoms and only in z direction for the phospholipids, on a C atom in a similar position to P for gangliosides) = 1000 kJ/(mol nm^2) • Lipid improper dihedral restraints = 1000 kJ/mol/rad^2
Step 5: equilibration (NPT)	2 ns Time step = 2 fs Thermostat + Barostat = Berendsen T = 310 K P = 1 bar • Protein backbone restraints = 2000 kJ/(mol nm^2)

	• Protein side chains restraints = 1000 kJ/(mol nm^2) • Lipid restraints (on P atoms and only in z direction for the phospholipids, on a C atom in a similar position to P for gangliosides) = 500 kJ/(mol nm^2) • Lipid improper dihedral restraints = 500 kJ/mol/rad^2
Step 6: equilibration (NPT)	4 ns Time step = 2 fs Thermostat + Barostat = Berendsen T = 310 K P = 1 bar • Protein backbone restraints = 1000 kJ/(mol nm^2) • Protein side chains restraints = 500 kJ/(mol nm^2) • Lipid restraints (on P atoms and only in z direction for the phospholipids, on a C atom in a similar position to P for gangliosides) = 200 kJ/(mol nm^2) • Lipid improper dihedral restraints = 200 kJ/mol/rad^2
Step 7: equilibration (NPT)	5 ns Time step = 2 fs Thermostat + Barostat = Berendsen T = 310 K P = 1 bar • Protein backbone restraints = 1000 kJ/(mol nm^2) • Protein side chains restraints = 500 kJ/(mol nm^2) • Lipid restraints (on P atoms and only in z direction for the phospholipids, on a C atom in a similar position to P for gangliosides) = 0 kJ/(mol nm^2) • Lipid improper dihedral restraints = 0 kJ/mol/rad^2
Step 8: production	1 μs for TM-IC systems and 500 ns for full-length systems Time step = 2 fs Thermostat = Nose-Hoover Barostat = Parrinello-Rahman T = 310 K P = 1 bar • No restraints

Table 5.3: Summary of simulations TM-IC and full-length homodimeric TrkA and TrkB atomistic systems in a model neuronal membrane.

System No	Protein	Part of system	Replicas	Time/Replica (μs)
1	TrkA	TM-IC	3	1
2	TrkB	TM-IC	3	1
3	TrkA	full-length	9	0.5
4	TrkB	full-length	7	0.5

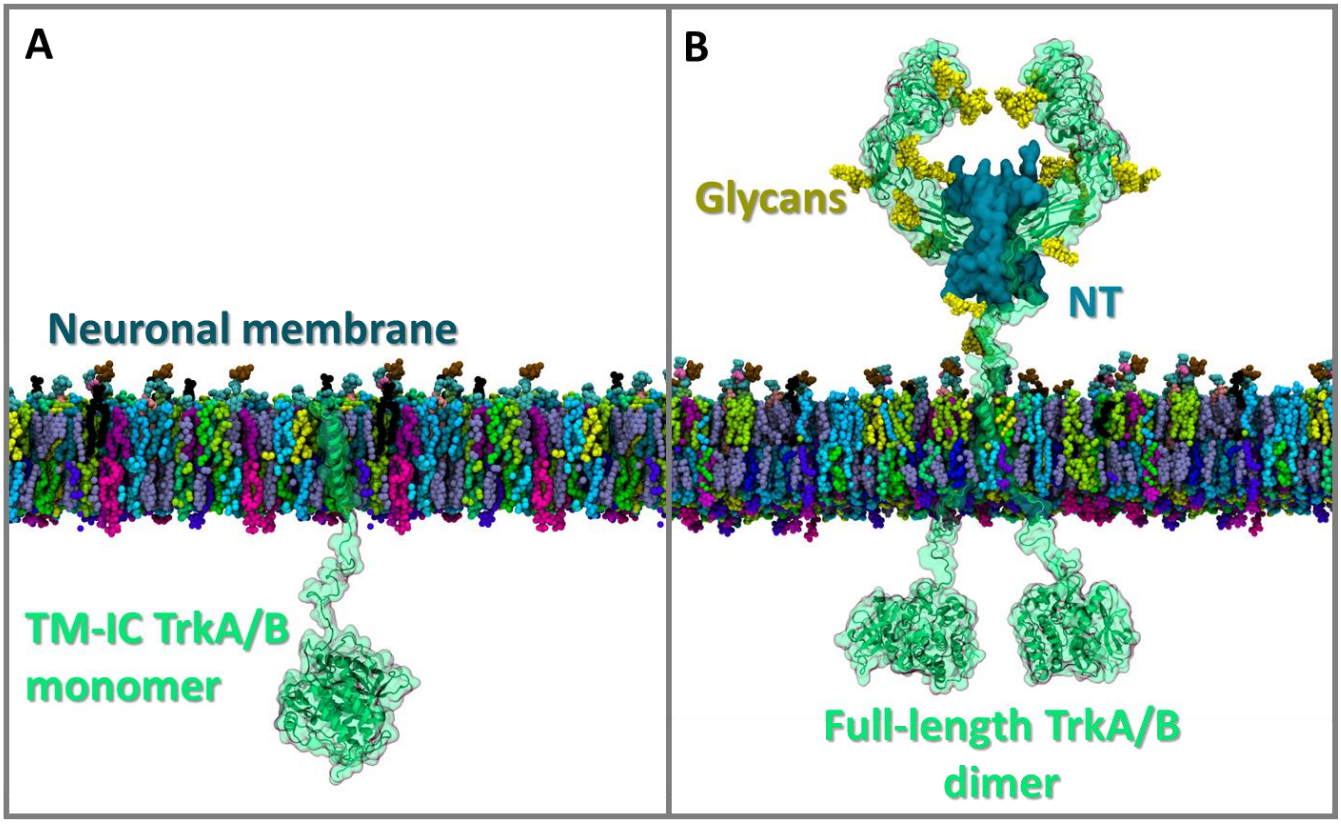


Figure 5.2: Systems whose simulation is described in this chapter. (A) TM-IC TrkA/B monomer embedded in neuronal membrane. (B) Full-length TrkA/B dimer bound to NT (NGF and NT-4/5) in the neuronal membrane.

Analysis of simulations

I post-processed all simulations with the *gmx trjconv* tool of GROMACS to make molecules whole, remove periodic boundary conditions (pbc) jumps and center the proteins in the simulation box. I saved frames every 100 ps. I also aligned the systems to the whole protein backbone throughout the trajectories. I performed the graphical rendering of protein structures from the trajectories with VMD 1.9.3.[271]

Root Mean Square Deviation (RMSD)

I calculated the Root Mean Square Deviation (RMSD) of the protein backbone for all full-length systems. Before the RMSD calculation, I aligned the proteins to their backbone throughout the simulation and then I calculated the RMSD of the backbone during the simulations with the MDAnalysis[272,273] python

library.[272,273] I calculated the RMSD in all frames of the trajectories and visualized the RMSD time evolution with the Matplotlib python library.[274] Additionally to the whole protein RMSD, I calculated the internal RMSD of all protein subdomains, by aligning the trajectory to the backbone of each subdomain before the RMSD calculation of that particular domain.

Root Mean Square Fluctuation (RMSF)

I calculated the Root Mean Square Fluctuation (RMSF) of the protein residues by aligning first the protein C_α atoms to an average structure from the simulations. I calculated the average structure with the align.AverageStructure class of MDAnalysis, which aligns the trajectory to the first frame and then averages the coordinates. Then I calculated the RMSF of the C_α atoms of each residue using MDAnalysis. I performed visualization of the results with Matplotlib.

Radius of gyration

I calculated the radius of gyration of the backbone of the full-length receptors, as well as separate domains, with MDAnalysis throughout the trajectories.

Distance and angle calculations

I calculated the distance between the z-coordinate of the center of mass (COM) of all NT atoms and the z-coordinate of the COM of the P atoms of the membrane, to estimate the distance of the NT from the plane of the membrane center. I also calculated the angle between the membrane normal and the vector defined by the center of geometry (COG) of the upper part of the NT and the COG of the lower part of the NT, with the lower part being the one close to the membrane. I did this to get an estimate of the orientation of the EC domain with respect to the membrane. I performed both calculations for all frames of the simulations, as well as for the frames after the first 150 ns, with MDAnalysis. I calculated and visualized the distributions of distance and angle with Seaborn.[275]

Secondary structure calculation

I calculated the secondary structure of the EC and IC (JM domain) linkers of the TrkA and TrkB receptors. The linkers were initially modeled as unstructured, as predicted from webservers, and thus I examined whether they remain like that during the simulations. For this calculation, I used the Timeline plugin of VMD for the residues of the EC and IC linkers. I did this analysis every 10 frames after the first 150 ns, during which the systems still equilibrate, as shown from the distance and angle calculations.

Contact analysis

For the TM-IC systems, I calculated the contact occupancies of each domain with the whole membrane and with each lipid species during the simulation time. For the full-length systems, I calculated the contacts between different species during the simulations: TM inter-helix contacts, kinase-kinase contacts, receptor domains-membrane contacts. I performed this calculation with MDAnalysis. I defined a contact as formed when any heavy (non-hydrogen) atom of a residue is within a distance of less than 6 Å of another heavy atom. I calculated the TM inter-helix contacts for all frames after the first 150 ns of all full-length replicas. I calculated the kinase-kinase and receptor-membrane contacts after the first 150 ns every 7 frames, because of the large number of residues involved in these calculations. Then, I post-processed the data with the Pandas library[278,279] and I visualized the results with Matplotlib.

TM helix order parameters

I calculated the order parameters that describe the various arrangements of the TM helices that were sampled during the simulations. These order parameters were (1) the distance between the centers of geometry (COG) of the two helices, (2) the crossing angle between the two helices defined as the dihedral angle formed by the COGs of the two helices and the COGs of the N-terminal halves of the two helices, (3) the phase and position which correspond to the rotation around each helix, defined as the dihedral formed by the N-terminal COG of one helix, the COG of the same helix, the COG of the second helix and one residue, which was Ala 428 for TrkA and Gly 444 for TrkB, (4) the distances between the respective N- and C-termini of the two helices, defined by their first and last residues. For the calculation of these order parameters, I used MDAnalysis[272,273] and the numpy.linalg.norm function of the Numpy library.[277] I did this analysis for the frames after the first 150 ns for all replicas.

Kernel Density Estimation (KDE)

I visualized the distribution densities of the aforementioned order parameters with Kernel Density Estimation (KDE), which smooths the density by summing kernels on the various sampled data points. I calculated and visualized the 2D KDE of two order parameters at a time with the seaborn.kdeplot function of Seaborn.[275]

Principal component analysis (PCA)

I performed principal component analysis (PCA) to identify the intrinsic motions of the full-length TrkA and TrkB receptors. I did this analysis for the frames after the first 150 ns, to calculate the motions of the equilibrated systems. For this purpose, I aligned the trajectories to the full receptor structures at 150

ns. After the alignment, I performed PCA in the full receptor coordinates using the built-in class in MDAnalysis. Plotting of the first and second principal component values was performed with Seaborn. I also calculated the dot products between PC1 and PC2 of different replicas with the numpy.dot function of the Numpy library.

Electrostatic potential

I performed electrostatic potential (ESP) calculations on the first and last frames of the TrkA kinase domain from the TM-IC system, to examine whether the interaction with the membrane in the last frame changes the ESP surface of the kinase. I used the Adaptive Poisson-Boltzmann Solver (APBS),[311] and I calculated the ESP surface and the inflated van der Waals-based ion-accessible surface. I multiplied the resulting two grids to visualize the ESP on the protein surface. I performed the visualization with VMD.

5.3 Results

TM-IC system simulations

The membrane environment can be crucial for the function of NT receptors as they translocate to lipid rafts upon NT binding.[112,114] This fact, together with evidence that certain lipids found in lipid rafts can activate the TrkA receptor,[119–121] indicates that receptor-lipid interactions are important for receptor activity. Therefore, I performed simulations of the TrkA and TrkB homodimeric receptors embedded in a model of a neuronal membrane which I created. The neuronal membrane model consisted of 19 lipid species. Due to the fact that the various lipids are usually separately parameterized in force fields and not extensively tested how they behave in a heterogeneous environment of lipid mixtures, I investigated the behavior of the membrane with several parameters.

First, I validated the neuronal membrane model by calculating parameters important for membrane models, such as the lipid density along the normal of the membrane and the membrane plane, the area per lipid, and the membrane thickness (Figure A.20 - Figure A.25 for the TM-IC systems). The distributions of the various lipid components in the x-y-z directions of the membrane are reasonable, while the membrane is thicker than commonly used for simulations model membranes due to the high ordering of the lipid tails induced by the high cholesterol concentration.

After the membrane validation, I investigated the interactions of the TM, JM and IC domains of TrkA and TrkB with the lipids of the neuronal membrane. Interestingly, the IC kinase domains of the TrkA and

TrkB receptors, which were initially placed in the bulk solvent at a distance from the membrane, approach and interact with the inner leaflet of the membrane during all replica simulations. The inner leaflet of the neuronal membrane has several lipid types with negatively charged headgroups. Analysis of the interactions between the kinase domain and the lipids showed that three regions of the TrkA kinase domain rich in positively charged residues interact with the inner leaflet of the membrane (Figure 5.3). The three regions of the TrkA kinase domain that interact with the membrane are magenta colored in Figure 5.3. These are located (1) in the end of the αC helix and its linker with the β3 strand, (2) in the linker between the β4 and β5 strands, and (3) in the activation loop (Figure 5.3). The positively charged residues K547, R554, R583 and R686 are responsible for the interactions with the membrane. Sequence analysis revealed that the positive charge in these positions is preserved in different species and Trk receptors, except for R583, which is an aspartic acid in most species (Figure 5.3).

A closer look into simulation snapshots revealed that these regions interact with negatively charged lipids of the inner leaflet (Figure 5.4). Electrostatic potential (ESP) calculations showed that the positive charge around these regions becomes exposed towards the end of the simulations compared to the first frame, as an induced effect from the interaction with the membrane. These interactions seem to stabilize the position of the kinase domain with respect to the membrane. This orientation of the kinase domain may facilitate the interaction of the two kinases in the homodimeric state of the receptor, as in the active state, the activation loop would face on the opposite site of the kinase, which does not interact with the membrane and could be accessible for phosphorylation by the other kinase monomer.

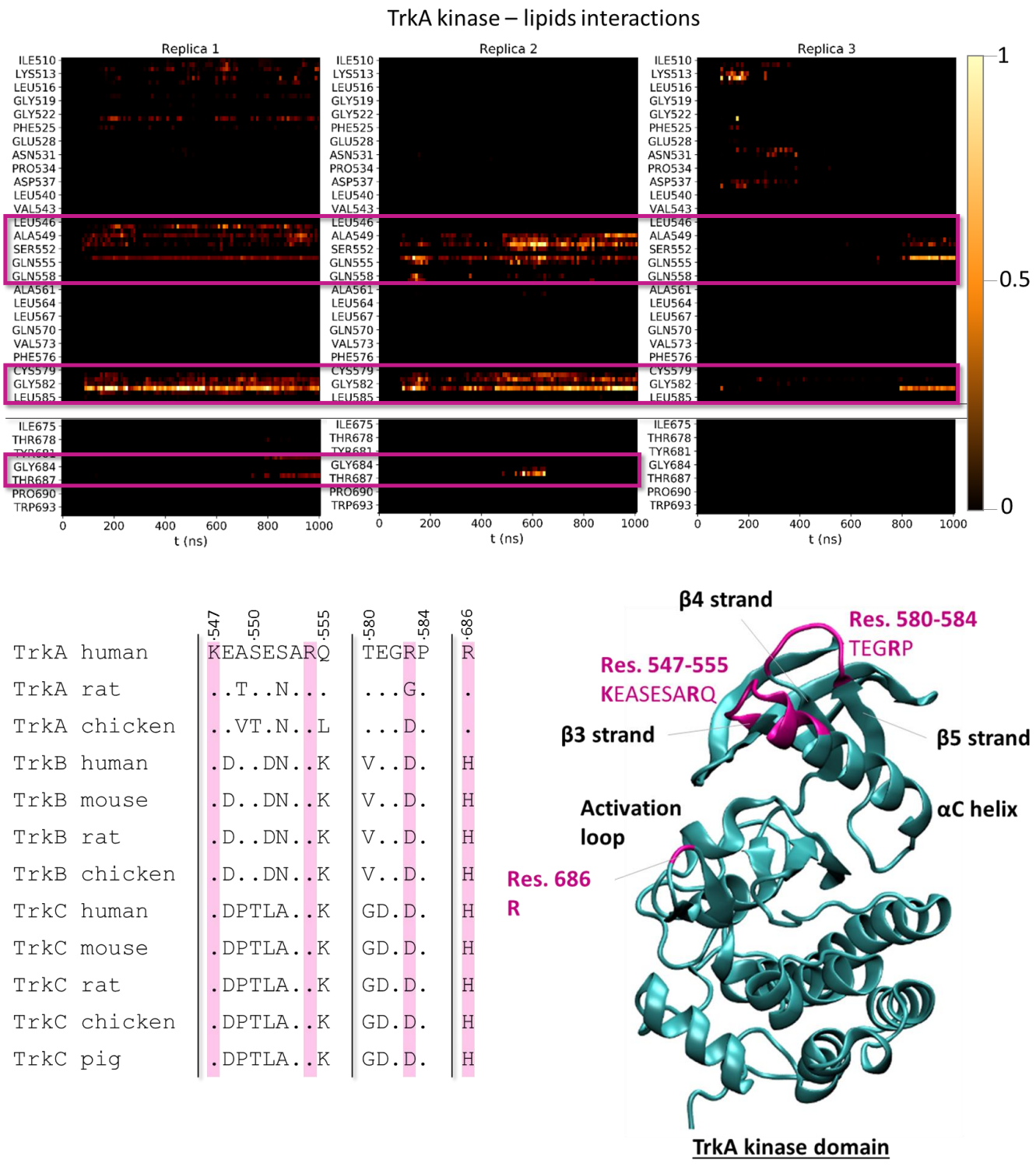


Figure 5.3: TrkA kinase – membrane interactions during the simulations. Upper plots: Contact occupancy between the TrkA kinase domain and the lipids of the neuronal membrane as function of time for the three replicas of the TM-IC system. The contact occupancy during the simulations are normalized from 0 to 1. Lower figure left: Sequence conservation of the three kinase regions that interact with the

membrane. Lower figure right: TrkA kinase domain with the regions that interact with the neuronal membrane colored in magenta. The residue numbers, as well as the residue types are indicated.

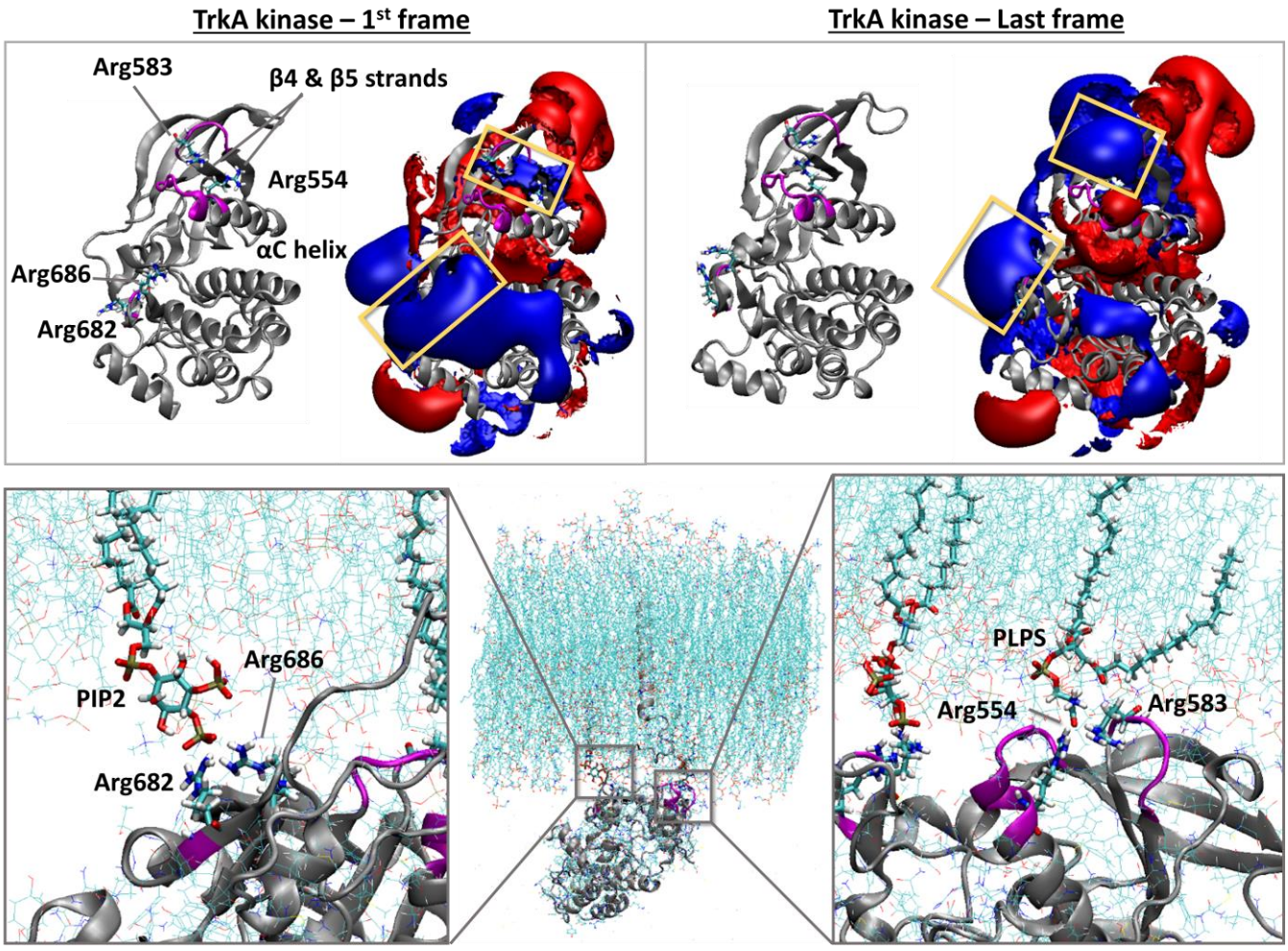


Figure 5.4: Interacting residues of the TrkA kinase domain with the membrane. Upper figure: Electrostatic isopotential surfaces at +1 and -1 charge of the first and last frames of one of the TM-IC replica simulations. The increase in the positive charge on the kinase interaction regions at the end of the simulation is indicated with yellow boxes. Lower figure: Simulation snapshot of the TrkA kinase domain interacting with the neuronal membrane. Insets show interactions between the positively charged residues of the kinase and the negatively charged lipids of the inner leaflet.

The kinase domain of the TrkB receptor also approaches and interacts with the membrane, however for TrkB no specific interaction patterns were identified (Figure A.26). I also performed interaction

analysis with the rest of the receptor domains and with specific membrane species. An important component of neuronal membranes is cholesterol, and thus I calculated its contacts with the TM helices of TrkA and TrkB (Figure 5.5). The analysis shows that TrkA TM helix interacts in its whole length with cholesterol, while TrkB seems to interact with its C-terminal half more than the N-terminal half, even though the latter contains a reversed Cholesterol Recognition Amino acid Consensus (CARC) motif (G**R**EHLSV**Y**A**VVV**IA).

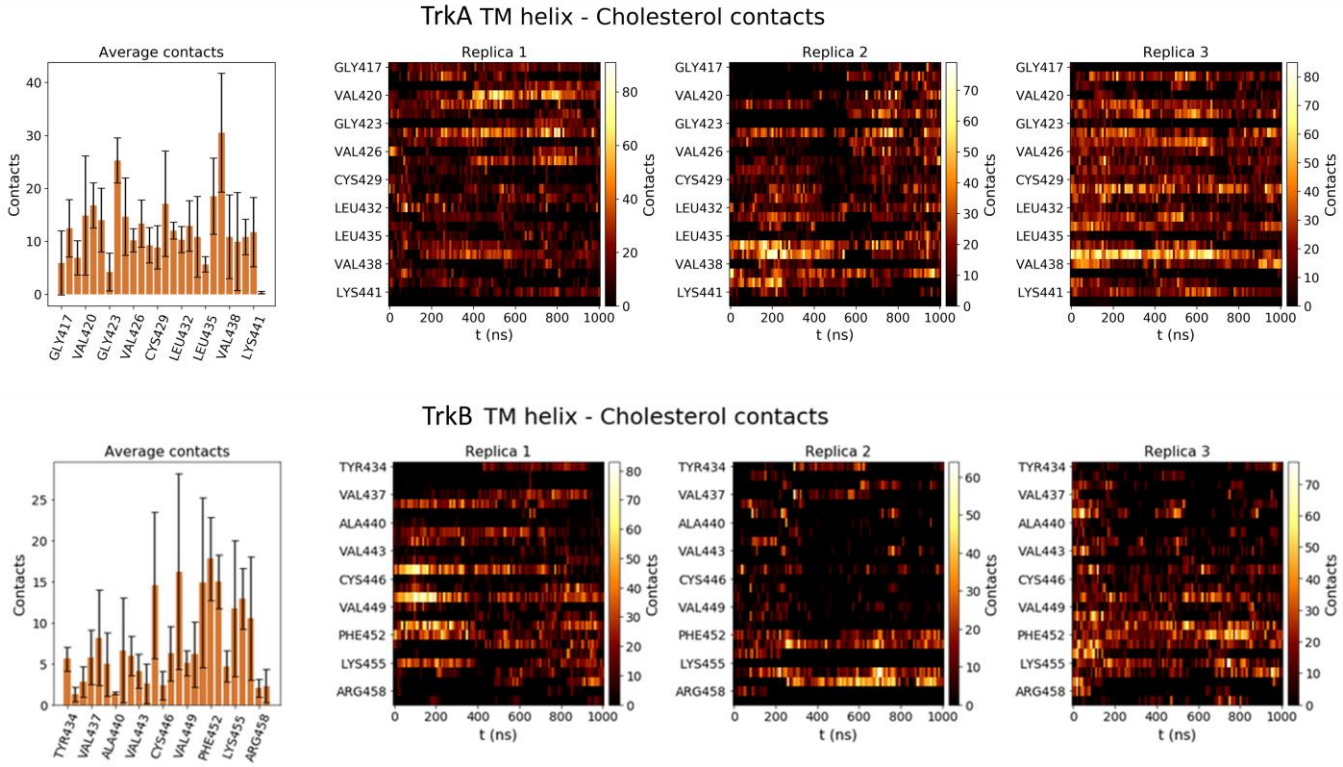


Figure 5.5: Number of contacts between the TM helices of TrkA and TrkB receptors and cholesterol. A contact is defined when two atoms from a residue and a cholesterol molecule are within 4.5 Å. The average contacts over the various replicas and the standard deviation are plotted in the first bar plots. The rest of the plots are heatmaps of the number of contacts over simulation time for each replica.

Another interesting point is the interaction of the JM domain of TrkA and TrkB with the inner leaflet of the membrane. While both JM domains interact with the membrane through their N-terminal region, which is right after the TM helix in sequence, TrkA-JM also interacts through its C-terminal part, right before the kinase domain starts (Figure A.27).

Full-length system simulations

After the initial test simulations of the TM-IC monomers in the neuronal membrane, I simulated the full-length TrkA and TrkB receptor dimers with NTs bound in their EC domains, embedded in the neuronal membrane. Specifically, I ran 9 replica simulations for TrkA and 7 replicas for TrkB. First, I validated the simulations by examining the stability of the initial structures. I performed RMSD calculations of the different domains of TrkA and TrkB receptors, with respect to the initial frame of the simulations, which corresponds to the experimental structures for the non-linker regions.

For TrkA, the RMSD calculations show high values for the EC domain, which seems to be due to hinge motions between the domains, since the intra-domain RMSD is low (Figure A.28). An exception to this is the D13 domain of TrkA in replica 9 for the 1st monomer and replica 1 for the 2nd monomer. The reason for this seems to be the partial unfolding of the N-terminal segment of the D13 domain (Figure 5.6), similarly to what was observed in the simulations of the EC domains in chapter 3. As already discussed in chapter 3, the D13 domain has quite high B-factors in the crystal structure with NGF bound (Figure A.3).

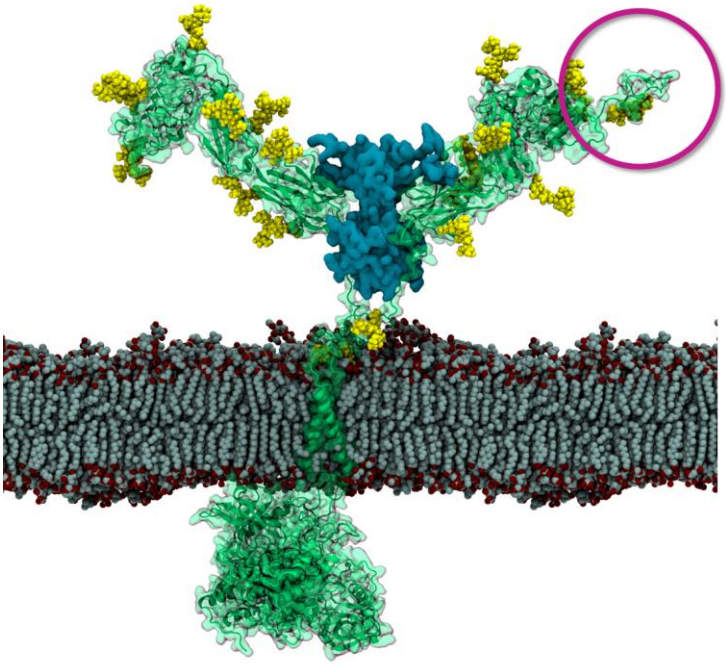


Figure 5.6: Snapshot from the end of replica 9 of TrkA full-length system. The red circle indicates the unfolded region of the N-terminal part of the D13 domain.

RMSF calculations also show high values for the N-terminal regions of the D13 domain for replica 1 and 9 (Figure A.29). The rest of the domains are quite stable, except for the EC and IC linkers, which started from extended conformations and move a lot as they equilibrate in the unrestrained production simulation. As indicated by the decreasing values of the radius of gyration, the linkers contract during the simulations and adopt more globular conformations (Figure A.30). Similar analyses were done for the TrkB simulations (Figure A.31 - Figure A.33).

After validating the simulations, I analyzed the overall behavior of the receptors during the simulations. Visualization of the trajectories revealed that the EC domains approach the membrane and bend towards it. Therefore, I decided to analyze the distance between the COM of the NT and the COM of the P atoms of the membrane phospholipids. This provided an estimate of how far away is the EC domain from the membrane. I also calculated the angle between the membrane normal (the vector vertical to the membrane) and a vector that connects the COGs of the upper and lower parts of the NT, the lower part being the one close to the membrane. This should provide a metric of the approximate orientation of the EC domain with respect to the membrane normal.

The results of this analysis are presented in Figure 5.7 for TrkA. The calculations for the whole production replicas (for all 500 ns) show that overall the NT-membrane distance decreases and the angle with the membrane normal increases. In order to compute the distribution of the distances and angles, I removed the first 150 ns of the simulations, during which the systems still equilibrate. The distributions of the distance for the different replicas have a wide range of distance values, between 60-100 Å. On the other hand, the angle between the NT and the membrane normal is at around 30-40^{o}, with only one replica reaching angles around 80^{o}.

I performed the same analysis on the TrkB simulations (Figure 5.8). As for TrkA, the NT-membrane distance has a wide distribution, with values ranging between 65-90 Å. The angle between NT and membrane normal seems to be slightly higher than TrkA, with values ranging mostly between 40-60^{o}. This could be due to the longer EC linker of the TrkB receptor, with 49 residues, compared to 33 residues in TrkA. Since I modeled the EC linkers as unstructured at the beginning of the simulations, as predicted by webservers, I examined whether they form any secondary structure during the simulations. I did this analysis for both EC and IC linkers (Figure A.34 and Figure A.35). The analysis showed that in some cases beta sheet elements are formed by a maximum of three residues, which means that they remain mostly unstructured.

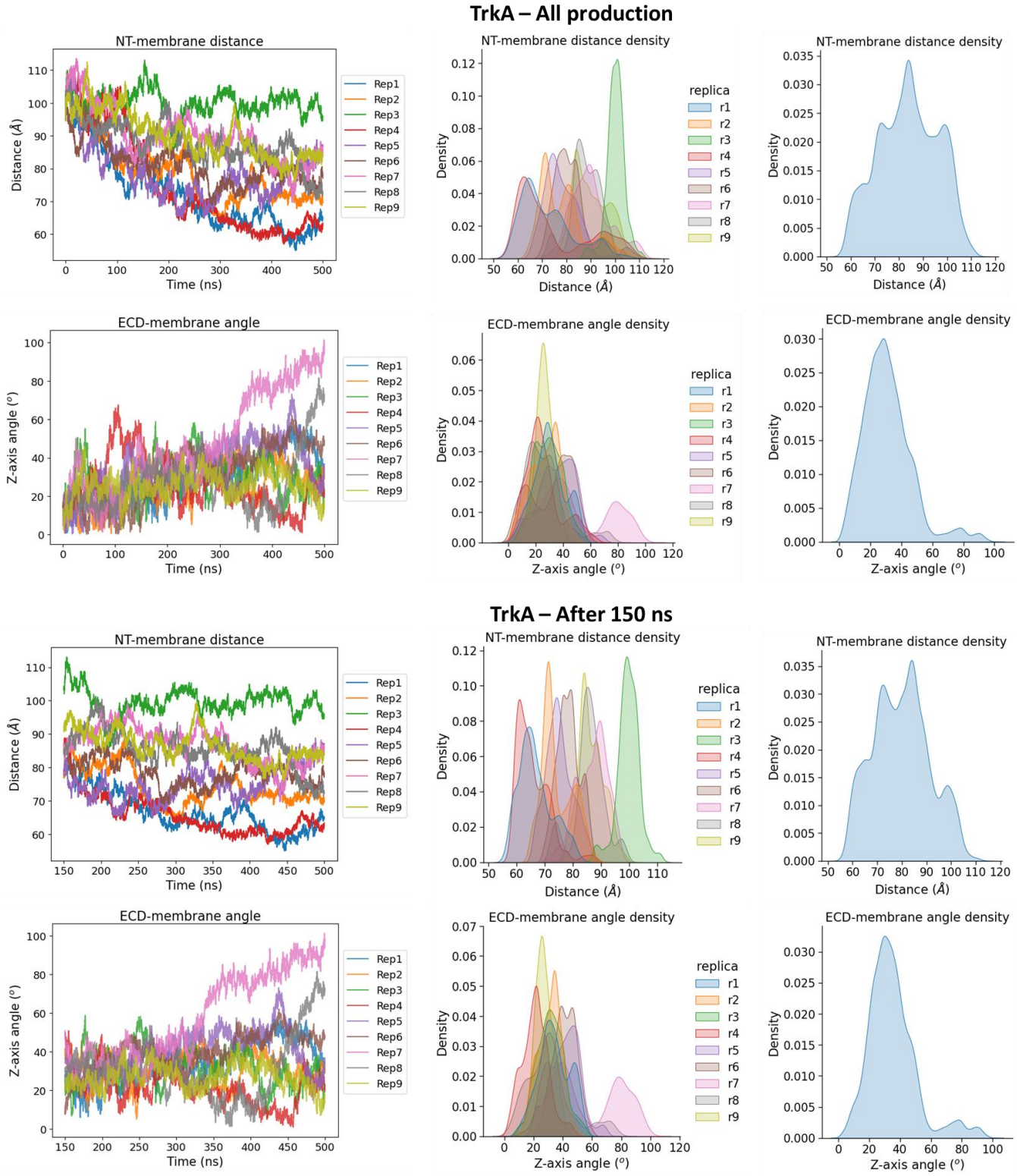


Figure 5.7: NT-membrane distance and EC domain-membrane normal angle calculations for the whole length of the simulations and after the first 150 ns of the full-length TrkA systems. The distribution densities of the distance and angle are also plotted per replica and all together.

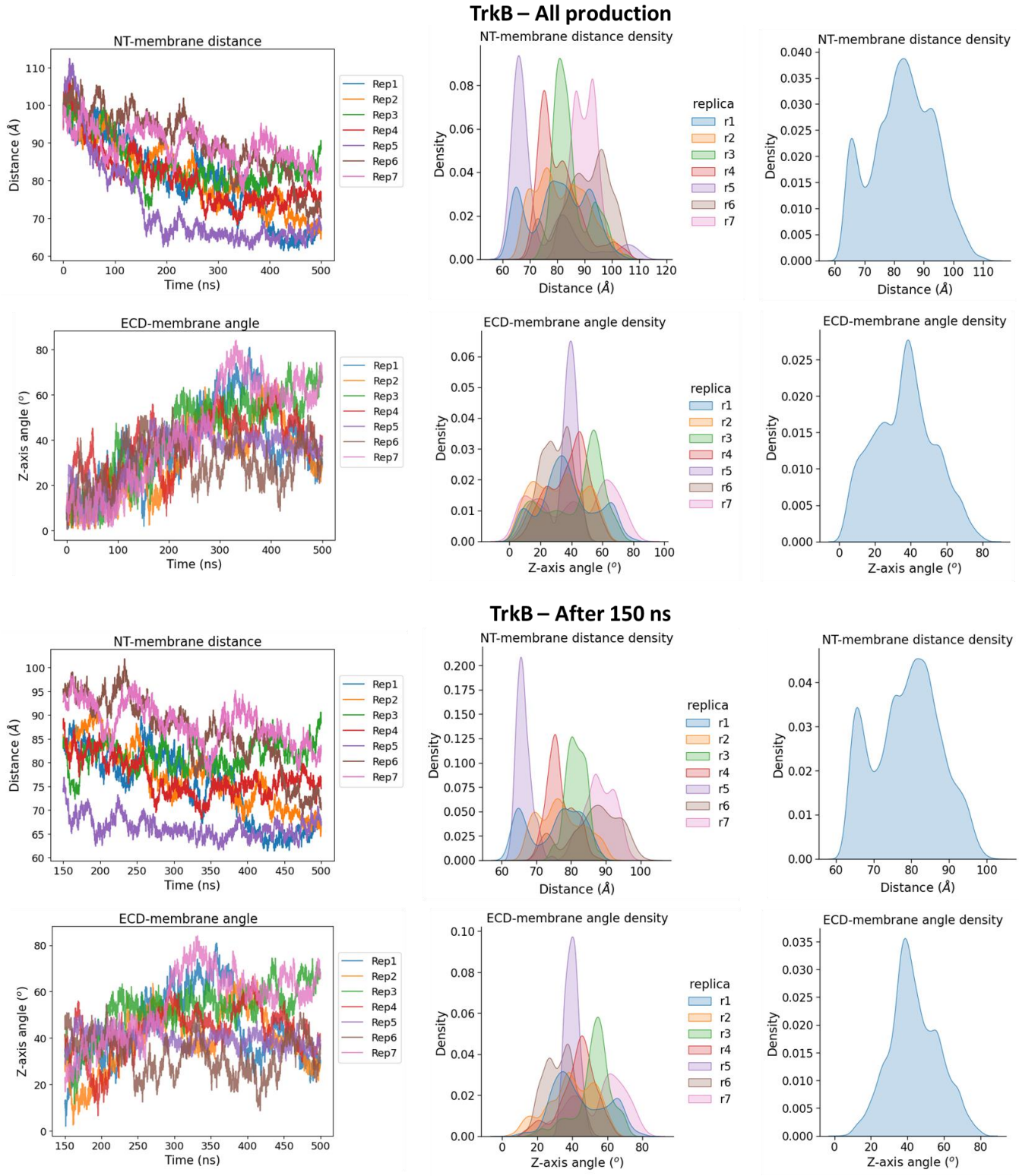


Figure 5.8: NT-membrane distance and EC domain-membrane normal angle calculations for the whole length of the simulations and after the first 150 ns of the full-length TrkB systems. The distribution densities of the distance and angle are also plotted per replica and all together.

After having analyzed the large motions of the full-length systems, I focused on the behavior of the TM helices in the neuronal membrane environment. I calculated the contact maps between the TM helices for each replica of the TrkA and TrkB systems. The results for TrkA are presented in Figure 5.9 and they show that the helices interact mostly through their inactive interface.[60] This means that the helices retain the inactive NMR interface from which they start, even though the NT is bound in the EC domain.

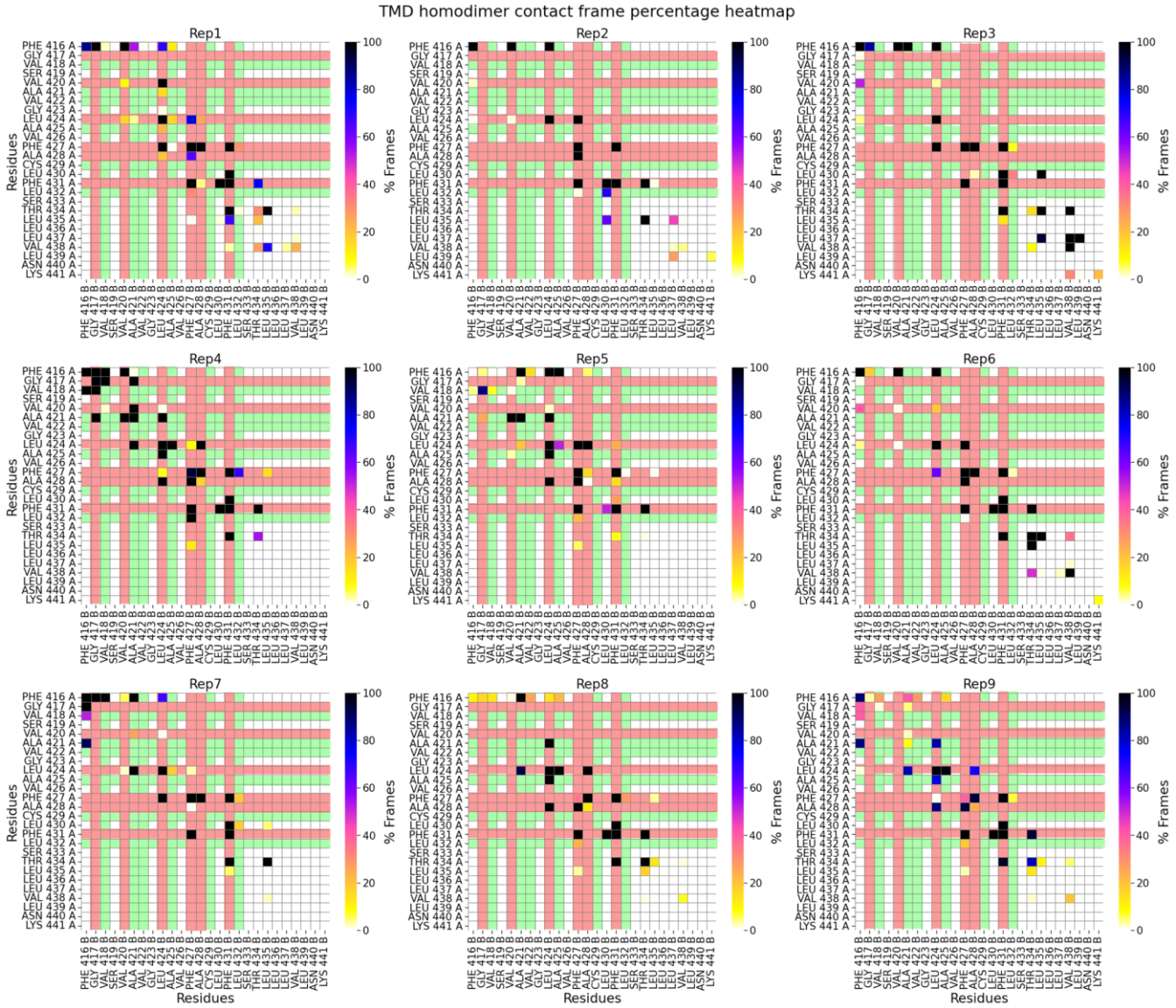


Figure 5.9: Heatmaps with the percentage of inter-helix contacts of the TM domain of TrkA during the different replica simulations. The residues that have been identified to be part of the active and inactive interface are indicated with green and red columns and rows, respectively.

The inability of the TrkA TM helices to adopt the active configuration may be due to the highly ordered environment of the neuronal membrane, and thus longer simulation times would be needed to sample the activation of the TM helices. I performed the same analysis for the TM helices of the TrkB receptor (Figure 5.10). In this case, I examined whether the TM helices interact through the AxxxG motif. The results show that the motif is partially positioned at the interface in replicas 1 and 4, but not in the rest of the replicas. In fact, in replicas 3 and 5 the TM helices seem to dissociate and interact only through their N-terminus.

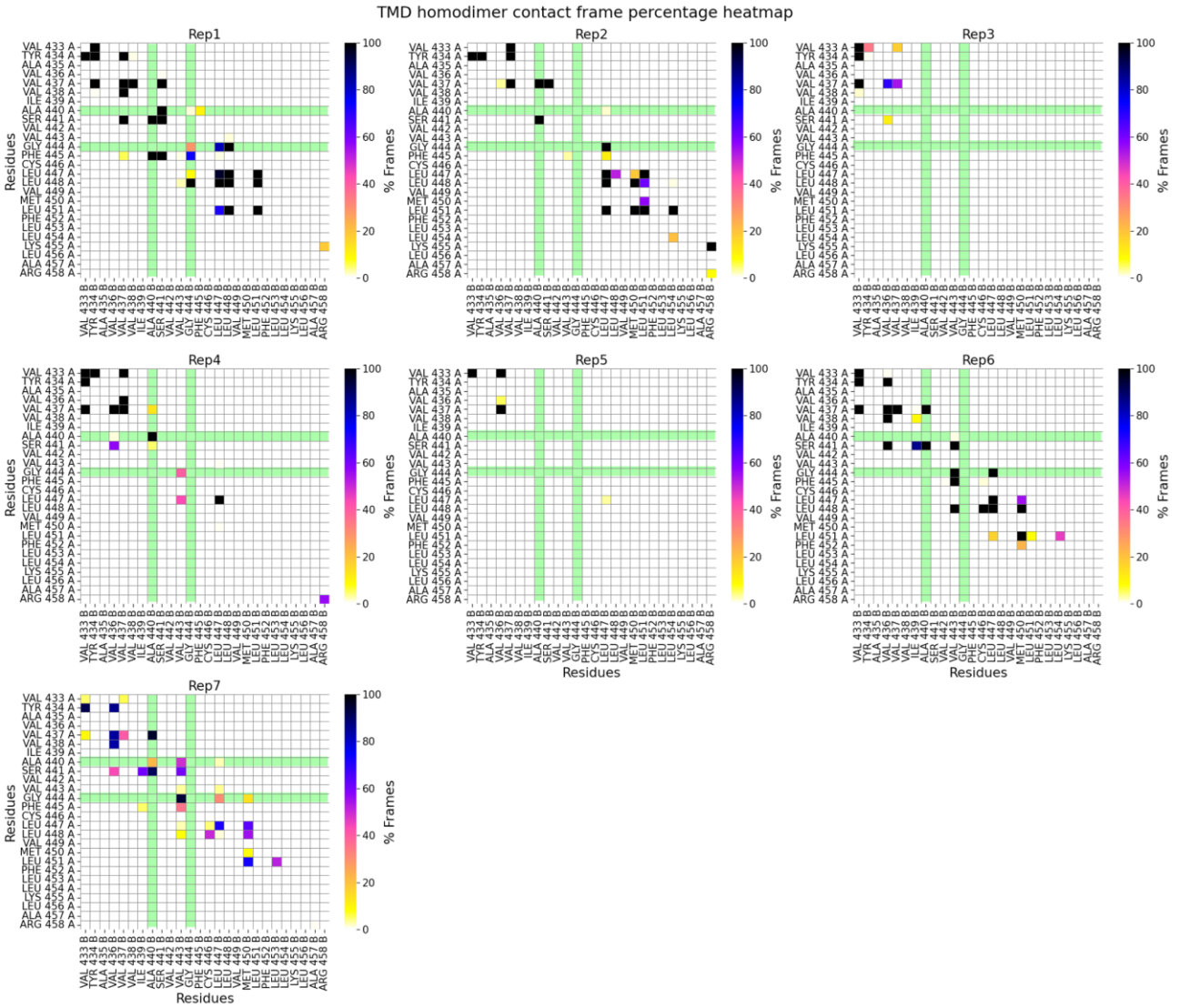


Figure 5.10: Heatmaps with the percentage of inter-helix contacts of the TM domain of TrkB during the different replica simulations. The AxxxG motif, which has been suggested to be part of the active interface, is indicated with green columns and rows.

I continued the analysis of the full-length systems, by examining the interactions of the various receptor domains with lipids of the neuronal membrane. As mentioned in the methods section, the neuronal membrane has different percentages of lipids and thus lipids in higher concentration are more probable to interact with the receptor. The relative populations of the lipids in the membrane are presented in the bar plot of Figure 5.11. In order to reduce the influence of the starting placement of the lipids in the membrane, I remodeled the membrane for each replica. This way, the lipids are placed in different positions at the beginning of the simulations.

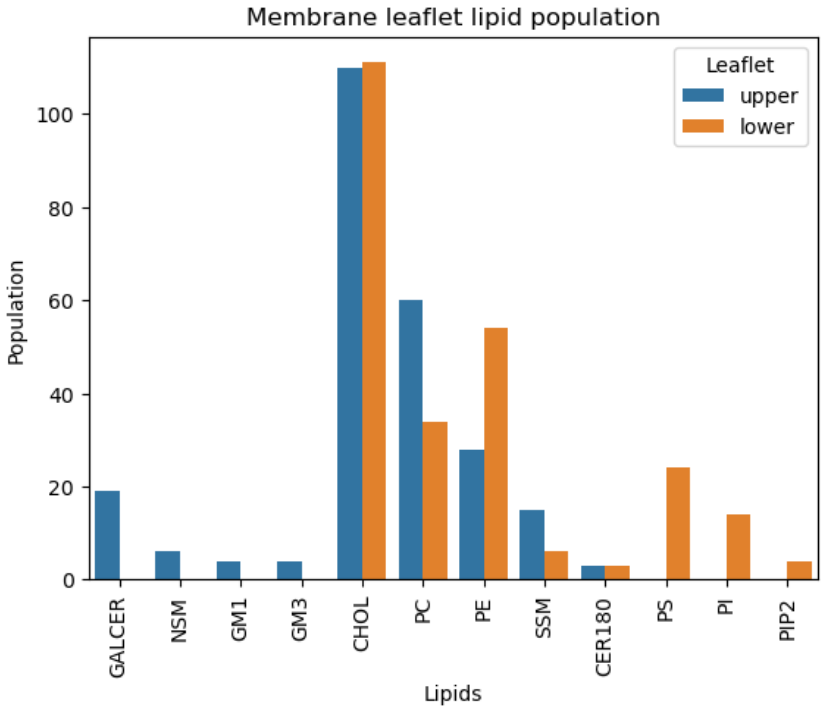


Figure 5.11: Neuronal membrane composition per leaflet. The bars show the ratios of lipid molecules per headgroup type. Lipid species at higher percentage are more probable to interact with the receptors.

The contact analysis between the membrane lipids and the different segments of TrkA and TrkB receptors is presented in Figure 5.12 and Figure 5.13, respectively. The analysis showed that the EC domain of TrkA interacts more frequently with the membrane than TrkB. Specifically, it interacts mostly with the GM1 ganglioside, which has been proven to activate the TrkA receptor experimentally,[119,121,125] and with the choline portion of PC lipids, which is positively charged. In the two replicas that the TrkB-

EC domain interacts with the membrane (replicas 4 and 5), it seems to interact with the same lipids but also with the GM3 ganglioside.

The EC linker of TrkA, interacts throughout the simulations with the PC and PE headgroups of lipids but also with the glycolipids GalCER, GM1 and GM3. The EC linker of TrkB interacts with the same lipids, although it seems to interact more often with GM3 that GM1. A difference compared to TrkA, is that the EC linker of TrkB interacts in almost all simulation frames with cholesterol. This may be due to the cholesterol recognition motif, which starts at the end of the EC linker and continues in the TM helix.

The TM domain of both TrkA and TrkB interacts a lot with cholesterol, PC, and PS, while the IC linker (JM domain) of both receptors interacts with the same lipids but also PI and PIP2. TrkA seems to interact in the biggest part of the simulation time with the negatively charged PS, PI and PIP2. Also, the kinase IC domain of TrkA interacts with the negatively charged lipids of the inner leaflet PS, PI and PIP2, as shown for the monomer simulations presented before. The kinase domain of TrkB also interacts with these lipids, but in some replicas to a lower percentage of the trajectory. Finally, the NT bound to TrkA, which is NGF, interacts more often with the membrane than NT-4/5, which is bound to TrkB. NGF interacts with GM1, PC and PE.

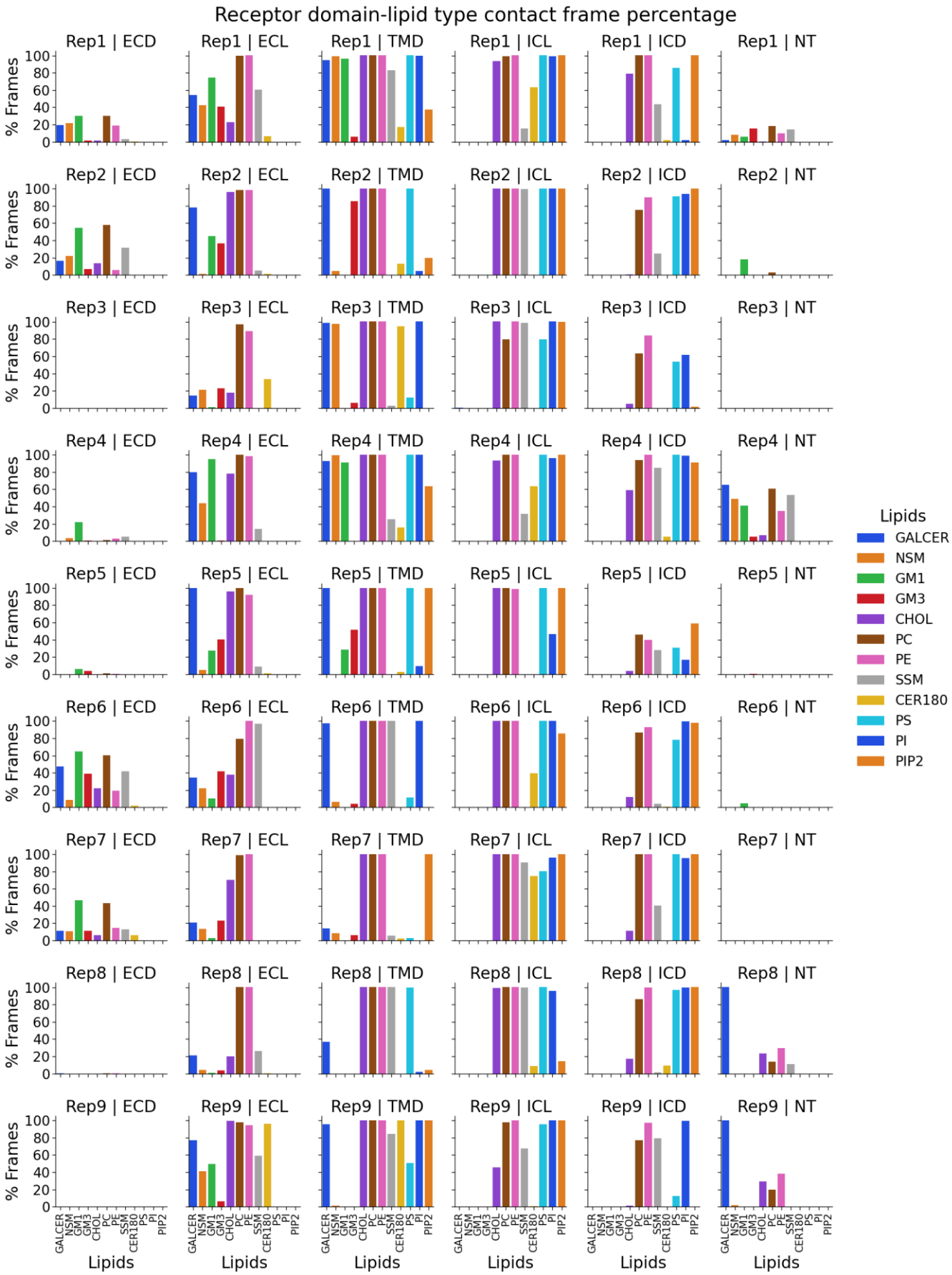

Receptor domain-lipid type contact frame percentage
Rep1 | ECD
Rep1 | ECL
Rep1 | TMD
Rep1 | ICL
Rep1 | ICD
Rep1 | NT
Rep2 | ECD
Rep2 | ECL
Rep2 | TMD
Rep2 | ICL
Rep2 | ICD
Rep2 | NT
Rep3 | ECD
Rep3 | ECL
Rep3 | TMD
Rep3 | ICL
Rep3 | ICD
Rep3 | NT
Rep4 | ECD
Rep4 | ECL
Rep4 | TMD
Rep4 | ICL
Rep4 | ICD
Rep4 | NT
Rep5 | ECD
Rep5 | ECL
Rep5 | TMD
Rep5 | ICL
Rep5 | ICD
Rep5 | NT
Rep6 | ECD
Rep6 | ECL
Rep6 | TMD
Rep6 | ICL
Rep6 | ICD
Rep6 | NT
Rep7 | ECD
Rep7 | ECL
Rep7 | TMD
Rep7 | ICL
Rep7 | ICD
Rep7 | NT
Rep8 | ECD
Rep8 | ECL
Rep8 | TMD
Rep8 | ICL
Rep8 | ICD
Rep8 | NT
Rep9 | ECD
Rep9 | ECL
Rep9 | TMD
Rep9 | ICL
Rep9 | ICD
Rep9 | NT
% Frames
0
20
40
60
80
100
GALCER
NSM
GM1
GM3
CHOL
PC
PE
SSM
CER180
PS
PI
PIP2
Lipids
Lipids
GALCER
NSM
GM1
GM3
CHOL
PC
PE
SSM
CER180
PS
PI
PIP2

Figure 5.12: Bar plots with contacts between the neuronal membrane lipids and the different segments of TrkA receptor, as well as the bound NT. The receptor was split into: ECD: extracellular domain, ECL: extracellular linker, TMD: transmembrane domain, ICL: intracellular linker, ICD: intracellular domain.

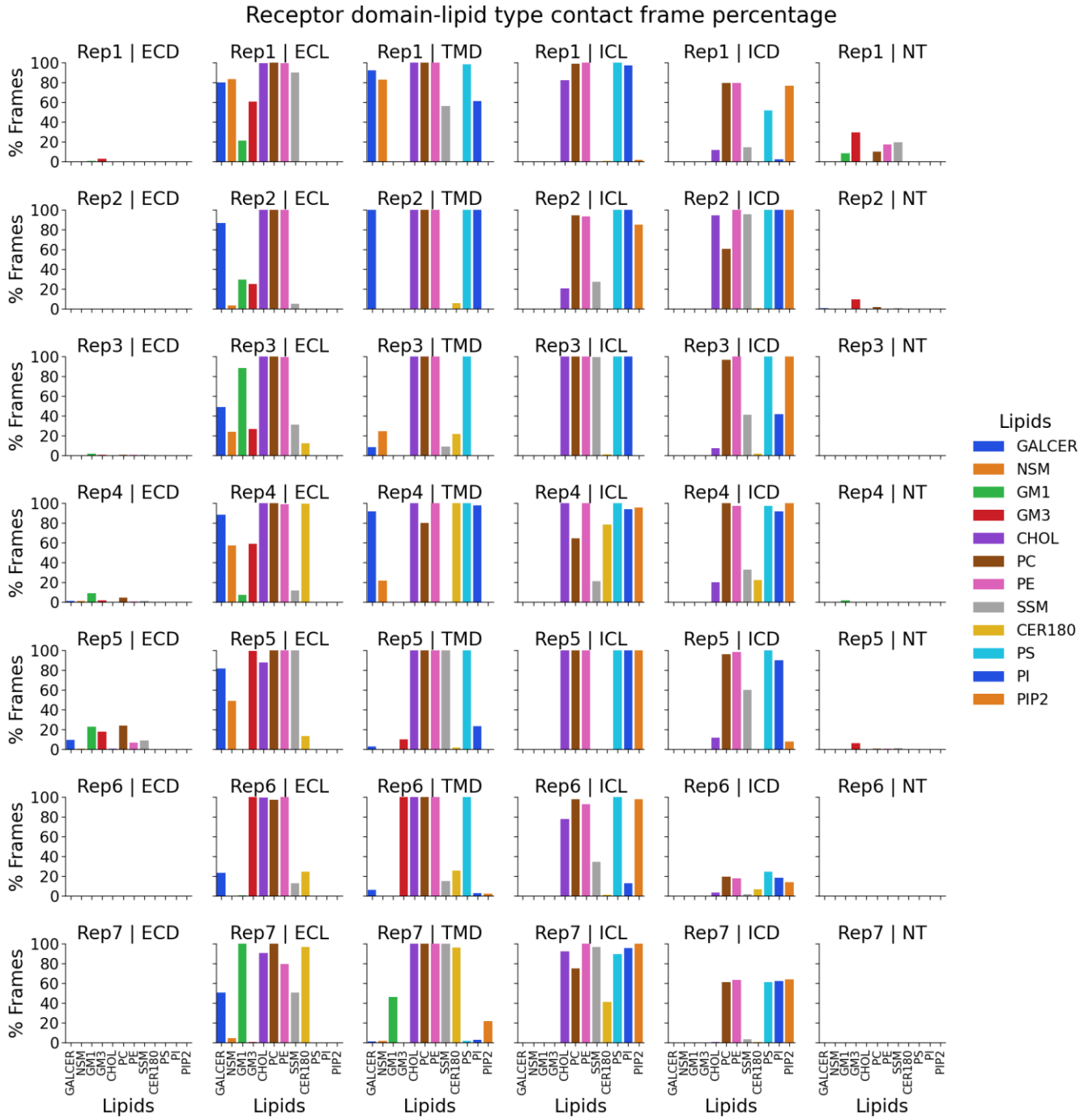


Figure 5.13: Bar plots with contacts between the neuronal membrane lipids and the different segments of TrkB receptor, as well as the bound NT. The receptor was split in: ECD: extracellular domain, ECL: extracellular linker, TMD: transmembrane domain, ICL: intracellular linker, ICD: intracellular domain.

5.4 Conclusions

In this chapter of the thesis, I investigated the dynamical behavior of the full-length TrkA and TrkB homodimeric receptors embedded in a model neuronal membrane. As shown experimentally, the NT receptors translocate to lipid rafts upon NT binding.[112,114] Moreover, certain lipids found in lipid rafts of neuronal membranes, like the ganglioside GM1, can activate the TrkA receptor.[119–121] For this reason, I also focused on how the TrkA and TrkB receptors behave in the neuronal membrane and which are their interactions with it.

I created a model of neuronal membrane consisting of 19 lipid species, by retaining important features of a reference neuronal membrane.[307] I performed test simulations of the membrane with the TM-IC TrkA and TrkB receptor monomers, to examine the correct behavior of the membrane and its interactions with the receptors. This step was important, as force field parameters for lipids are usually developed in a homogeneous lipid environment rather than in lipid mixtures. I validated the membrane model by achieving reasonable values for parameters, such as lipid density, area per lipid and membrane thickness, which were close to values reported in the literature.[307]

After the membrane validation, I analyzed the interactions between the TM-IC receptor monomers and the membrane. The kinase IC domain of both TrkA and TrkB, which was initially placed at a distance from the membrane, approached and interacted with the inner leaflet of the neuronal membrane in all replica simulations. Contact analysis revealed three specific regions in the kinase domain of TrkA which interacted with the membrane in all replicas of the TM-IC and full-length systems. These regions contain positively charged residues, which interact with the negatively charged lipids of the inner leaflet. The kinase domain of TrkB also interacted with the membrane, however no specific interaction patterns were identified. This specific orientation of TrkA with respect to the membrane leaves exposed the area in which the activation loop would reside in its active conformation, and thus would be accessible for phosphorylation by the other kinase monomer.

After the test simulations with the TM-IC systems, I modelled the full-length TrkA and TrkB receptor homodimers bound to the NGF and NT-4/5 NTs, respectively. I embedded the full-length receptors in the neuronal membrane and simulated them. During these simulations, the EC domains approached the membrane and bent towards it. The TrkB-EC domain came closer to the membrane than TrkA, probably because of its longer EC linker. However, the EC domain of TrkA forms more interactions with the membrane and particularly with the GM1 ganglioside, which activates the TrkA receptor

experimentally.[119,121,125] Also, the JM and kinase domains of TrkA interacted for a long simulation time with negatively charged lipids of the inner leaflet. The regions of TrkA kinase that interacted with the membrane were the same as those observed for the TM-IC systems, which means that this arrangement is also possible in the dimeric state of the receptor. However, the two kinases were not in an arrangement where one could phosphorylate the other. For this to happen, each kinase domain should adopt its active conformation, which involves a big movement of the activation loop. Considering that both kinase monomers start from an inactive conformation, much longer simulation times would be required for them to adopt an active conformation. Consequently, the currently sampled arrangements probably show how the kinases behave in the inactive state of the receptor. An approach to study the activation of the kinase domains would be the use of enhanced sampling methods, which would bias the activation pathway. Another possible solution to this would be the simulation of the kinase domains starting from a modeled active state, based on homology modeling using the active state of other crystallized kinases. Besides, the active conformation of kinases is quite conserved among different types of kinases.[66]

A difference between the TrkA and TrkB receptors was that the kinase domains of TrkB interacted in more than one way with the membrane, with at least two regions in opposite sides of the kinases domain creating contacts with lipids. This shows that the TrkB kinase does not have a dominant arrangement intracellularly, which might be the reason why the TrkB kinase has a 2-fold lower catalytic efficiency than TrkA and becomes autophosphorylated *in vitro* more slowly than TrkA.[312] The difference in the orientation of TrkA and TrkB kinase domains also suggests that different areas of the kinase can be exposed for interactions with binding partners which can lead to different signaling cascades. Indeed, activation of TrkA and TrkB in the same cell type can yield very different cellular outcomes.[312] For example, activation of TrkA in neuroblastoma cells can inhibit cell growth, but activation of TrkB does not.[313] The distinct mechanism of kinase interaction with the membrane might be the first step for understanding differences in the signaling of TrkA and TrkB receptors.

6 Investigation of neurotrophin mimetics' mechanism of action

6.1 Introduction to the study of neurotrophin mimetics and its challenges

Neurotrophins have been tested as a potential treatment for neurodegenerative disorders, since some of these diseases are linked to low levels of neurotrophin expression.[183–185] The administration of neurotrophins is restricted by the blood brain barrier (BBB), due to their size, and by the difficulty to selectively target the affected neurons. As a solution, invasive procedures, such as intracerebroventricular (ICV) injection,[187,188] have been used for the administration of neurotrophins, which have been reported to cause side effects, such as neuropathic pain.[189–195]

An alternative solution to neurotrophins, is the development of small molecule neurotrophin mimetics, which can activate the neurotrophin receptors. Small molecules generally have better pharmacokinetic properties and administration is easier. Previous work by the collaborators of the EuroNeurotrophin consortium has identified two steroidal agonists of the TrkA and p75 neurotrophin receptors, dehydroepiandrosterone (DHEA) and its C17-spiroepoxy derivative BNN-27, that are able to promote neuronal survival, without causing hyperalgia.[202] Therefore, the EuroNeurotrophin network focused on the development of new DHEA derivatives, with the aim to selectively activate each neurotrophin receptor.

In this chapter, I will present the molecular modeling of the neurotrophin mimetics developed by the consortium bound to the receptors, and the efforts to understand their mechanism of action through molecular dynamics simulations. The possibility of the existence of more than one binding site on the receptors has also been examined and is presented here.

6.2 Methodology

Protein modeling

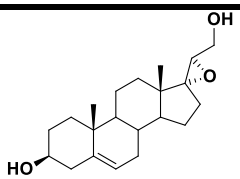
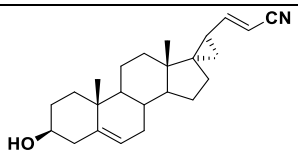
For the modeling studies of this chapter, I used the crystal structure of the complex of TrkA-D5 homodimer with the mature NGF homodimer (PDB ID: 1WWW)[29] and the crystal structure of the complex of the TrkB-D5 homodimer with the mature neurotrophin-4/5 (NT-4/5) homodimer (PDB ID: 1HCF).[260] Some residues were missing from these crystal structures and thus I modeled them. Specifically, the structure of NGF had the following missing residues: Ser 122, Pro 182 – Ser 187, Ala 237 – Ala 241 (residue numbers correspond to the UniProt sequence of human NGF, UniProt code: P01138). The structure of NT-4/5 had the following missing residues: Asn 145 – Gly 150 and Gly 207 – Ala 210 in chain A and Gly 81 – Glu 84, Asn 145, Arg 209 and Ala 210 in chain B (residue numbers correspond to the UniProt sequence of human NT-4/5, UniProt code: P34130). I built the missing residues with the AutoModel class of MODELLER v.9.23.[246] Also, the first three residues of TrkB-D5 were cloning artefacts and thus I mutated them to the wild-type sequence. For NGF, the linker corresponding to residues Pro 182 – Ser 187 has a conserved conformation in existing crystal structures and therefore I built it using the structure of human NGF from PDB ID: 5JZ7[259] as a template. For NT-4/5, I built the linker corresponding to residues Asn 145 – Gly 150 using the structure of human NT-4/5 from PDB ID: 1B98[23] as a template. I prepared the final models with the Protein Preparation Wizard[314,315] of Schrödinger Suite v.2020. Specifically, I followed a procedure, during which bond orders were assigned, hydrogens were added, disulfide bonds were created, protonation states were assigned at pH 7 with PROPKA,[316] and optimization of the H-bond network and restrained energy minimization using the OPLS3e force field[317] with heavy atoms restricted to a maximum root mean squared deviation (RMSD) from the initial structure of 0.3 Å were performed. I also used the TrkA and TrkB transmembrane (TM) helical dimers, which were presented in chapter 4, for molecular dynamics (MD) simulations with the compounds.

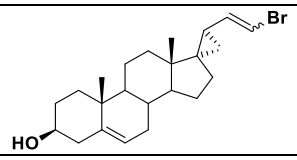
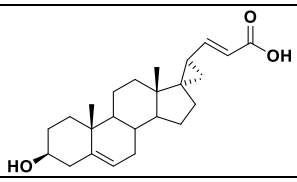
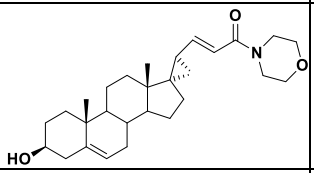
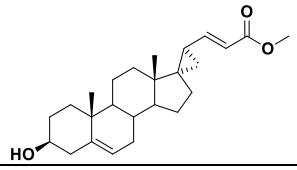
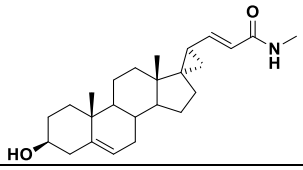
Molecular docking

I performed docking calculations of the EuroNeurotrophin consortium compounds to the TrkB receptor with NT-4/5 bound. The same calculations were performed in the TrkA receptor with NGF by Alexandros Tsengenes from the Molecular and Cellular Modeling (MCM) group at the Heidelberg Institute for Theoretical Studies (HITS). I used the SiteMap tool[318–320] of the Schrödinger Suite for the identification

of possible binding sites in the TrkB – NT-4/5 structure. From the predicted binding sites for docking, I selected the sites 1a and 1b, which were previously described as the possible binding sites of the TrkA agonist BNN27, which belongs to the same congeneric series of compounds.[202] I built the compounds shown in Table 6.1, which are the compounds experimentally verified to be active for the TrkA or TrkB receptors by PhD students in the EuroNeurotrophin consortium, with Maestro[321] and prepared them with the LigPrep tool[322] of the Schrödinger Suite. The compounds were docked to the two symmetric sites 1b of the TrkB - NT-4/5 complex using the standard precision (SP) protocol of the Glide tool[247–250] of the Schrödinger Suite. Grids were generated with Glide to describe the physicochemical properties of the protein pockets. The hydroxyl groups of serine, threonine and tyrosine residues and the thiol groups of cysteine residues, located within the grid box that describes the binding sites, were allowed to rotate during the docking calculation, to capture all possibilities for making hydrogen bonds with the ligands. For site 1a, which is narrower than 1b, ligand poses with favorable scores could not be obtained with rigid body docking and therefore, the Induced Fit Docking (IFD) tool[323–326] of the Schrödinger Suite was used; this utilizes Glide for docking and Prime[327–329] for protein remodeling. I also performed docking calculations of the ENT-A011, ENT-A013E/Z and BNN-27 compounds to the TM domain of the TrkB receptor, at the interface of the N-terminal part of the two TM helices. Graphical analysis of the docking poses was carried out with VMD v.1.9.3.[271]

Table 6.1: Small molecule neurotrophin mimetics discovered by the EuroNeurotrophin consortium. The biological tests (unpublished except for ENT-A011[330] and ENT-A013E/Z[204], provided by Alessia Latorrata, National Hellenic Research Foundation, NHRF, Greece) are the following for TrkA expressing cells: phosphorylation of TrkA (pTrkA), AKT (pAKT) and ERK1/2 (pERK1/2) and cell survival. Similar for TrkB expressing cells.

		TrkA				TrkB		
Compound code	**Structure**	**pTrkA**	**pAKT**	**pERK1/2**	**Cell survival**	**pTrkB**	**pAKT**	**Cell survival**
BNN-27		yes	yes	yes	yes	no	no	no
ENT-A011[330]		no	no	no	no	yes	yes	yes

ENT-A013E/Z[204]		yes	yes	yes	yes	no	no	no
ENT-A040		yes	yes	yes	yes	no	no	no
ENT-A044		no	no	no	no	yes	yes	yes
ENT-A045		no	no	no	no	yes	yes	yes
ENT-A061		no	no	no	no	yes	yes	yes

Molecular dynamics (MD) simulations

I performed MD simulations with the ENT-A011, ENT-A013E/Z and BNN27 compounds bound to the TrkB receptor. The same simulations were performed to TrkA by Alexandros Tsengenes. I performed simulations with the extracellular (EC) and the transmembrane (TM) domains of the TrkB receptor with the compounds bound. For the simulations with the EC domain, I used the docking poses presented in the previous section, with two compounds, each bound at the two interfaces of NT with the D5 domain of the receptors (Table 6.2). I tested compounds bound in the two different pockets in sites 1a and 1b at distinct replica simulations. For the simulations with the TM domains, I performed two sets of simulations. The first set had one compound molecule docked at the interface of the N-terminal part of the TM helices, while the second set had several compounds manually placed around the TM helices in the membrane environment (Table 6.3). I also performed some sets of simulations with the active states of the TM helices as identified by the MD simulations that were presented in chapter 4.The TM systems were embedded in pure POPC or POPC:cholesterol 6:4 bilayers with CHARMM-GUI.[310]

I simulated the EC models with the compounds bound to them with the same protocol that I used for the simulations with the glycosylation in chapter 3. For the TM systems, I used the same simulation

protocol as for the full-length atomistic systems in a neuronal membrane presented in chapter 5. I ran three replica simulations for ~300 ns for each system.

I also performed membrane permeability simulations of the small-molecules, using the same equilibration protocol that I presented in chapter 5 for the atomistic simulations in membrane. I set up the systems by placing one molecule at the outer extracellular region of the membrane and performed MD simulations to observe if and how the compounds enter the membrane. I tested the permeability in both pure POPC or POPC:cholesterol 6:4 bilayers. Additionally to the aforementioned compounds, I also tested some other compounds for their membrane permeability, like cholesterol, DHEA and DHEAS. I ran three replica simulations for 500 ns for each system. The same simulations, but with ten molecules instead of one next to the membrane, were performed by Alexandros Tsengenes. A summary of all these simulations is shown in Table 6.4. I carried out the graphical analysis of the docking poses with VMD v.1.9.3.[271]

Table 6.2: Summary of simulated TrkA and TrkB EC systems with NTs and two compounds bound.

Receptor domain	Binding site	NT	Compound
TrkA-D5	site 1a	NGF	2xENT-A011
TrkA-D5	site 1a	NGF	2xENT-A013E
TrkA-D5	site 1a	NGF	2xENT-A013Z
TrkA-D5	site 1a	NGF	2xBNN27
TrkA-D5	site 1b	NGF	2xENT-A011
TrkA-D5	site 1b	NGF	2xENT-A013E
TrkA-D5	site 1b	NGF	2xENT-A013Z
TrkA-D5	site 1b	NGF	2xBNN27
TrkA-D5	-	NGF	-
TrkB-D5	site 1a	NT4/5	2xENT-A011
TrkB-D5	site 1a	NT4/5	2xENT-A013E
TrkB-D5	site 1a	NT4/5	2xENT-A013Z
TrkB-D5	site 1a	NT4/5	2xBNN27
TrkB-D5	site 1a	NT4/5	2xENT-A045
TrkB-D5	site 1b	NT4/5	2xENT-A011
TrkB-D5	site 1b	NT4/5	2xENT-A011 - IFD
TrkB-D5	site 1b	NT4/5	2xENT-A013E
TrkB-D5	site 1b	NT4/5	2xENT-A013Z
TrkB-D5	site 1b	NT4/5	2xBNN27
TrkB-D5	site 1b	NT4/5	2xENT-A045
TrkB-D5	-	NT4/5	-

Table 6.3: Summary of simulated TrkA and TrkB TM systems with one or ten compound molecules. The inactive dimers correspond to the NMR structure of TrkA and to the model of TrkB based on the TrkA NMR structure. The active state of the dimers was modeled by manually rotating the helices. "Dist. Rest." refers to distance restraints.

Receptor domain	Dimer state	Compound	# of Compounds	Environment
TrkA-TM	inactive	ENT-A011	10	POPC
TrkA-TM	inactive	ENT-A013E	10	POPC
TrkA-TM	inactive	ENT-A013Z	10	POPC
TrkA-TM	inactive	BNN27	10	POPC
TrkA-TM	inactive	cholesterol	10	POPC
TrkB-TM	inactive	ENT-A011	10	POPC
TrkB-TM	inactive	ENT-A013E	10	POPC
TrkB-TM	inactive	ENT-A013Z	10	POPC
TrkB-TM	inactive	BNN27	10	POPC
TrkB-TM	inactive	cholesterol	10	POPC
TrkA-TM	active	ENT-A011	1	POPC
TrkA-TM	active	ENT-A013E	1	POPC
TrkA-TM	active	ENT-A013Z	1	POPC
TrkA-TM	active	BNN27	1	POPC
TrkB-TM	active	ENT-A011	1	POPC
TrkB-TM	active	ENT-A013E	1	POPC
TrkB-TM	active	ENT-A013Z	1	POPC
TrkB-TM	active	BNN27	1	POPC
TrkA-TM	active	ENT-A011	1	POPC:cholesterol 6:4
TrkA-TM	active	ENT-A013E	1	POPC:cholesterol 6:4
TrkA-TM	active	ENT-A013Z	1	POPC:cholesterol 6:4
TrkA-TM	active	BNN27	1	POPC:cholesterol 6:4
TrkB-TM	active	ENT-A011	1	POPC:cholesterol 6:4
TrkB-TM	active	ENT-A013E	1	POPC:cholesterol 6:4
TrkB-TM	active	ENT-A013Z	1	POPC:cholesterol 6:4
TrkB-TM	active	BNN27	1	POPC:cholesterol 6:4
TrkB-TM	active – Dist. Rest.	ENT-A011	1	POPC
TrkB-TM	active – Dist. Rest.	ENT-A013E	1	POPC
TrkB-TM	active – Dist. Rest.	ENT-A013Z	1	POPC
TrkB-TM	active – Dist. Rest.	BNN27	1	POPC
TrkB-TM	active – Dist. Rest.	ENT-A011	1	POPC:cholesterol 6:4
TrkB-TM	active – Dist. Rest.	ENT-A013E	1	POPC:cholesterol 6:4
TrkB-TM	active – Dist. Rest.	ENT-A013Z	1	POPC:cholesterol 6:4
TrkB-TM	active – Dist. Rest.	BNN27	1	POPC:cholesterol 6:4

Table 6.4: Summary of simulated systems for membrane permeability of the compounds.

Compound	# of compounds	Environment
BNN27	1	POPC
ENT-A011	1	POPC
ENT-A013E	1	POPC
ENT-A013Z	1	POPC
cholesterol	1	POPC
DHEA	1	POPC
DHEAS	1	POPC
BNN27	1	POPC:cholesterol 6:4
ENT-A011	1	POPC:cholesterol 6:4
ENT-A013E	1	POPC:cholesterol 6:4
ENT-A013Z	1	POPC:cholesterol 6:4
cholesterol	1	POPC:cholesterol 6:4
DHEA	1	POPC:cholesterol 6:4
DHEAS	1	POPC:cholesterol 6:4
BNN27	10	POPC
ENT-A011	10	POPC
ENT-A013E	10	POPC
ENT-A013Z	10	POPC
cholesterol	10	POPC
BNN27	10	POPC:cholesterol 6:4
ENT-A011	10	POPC:cholesterol 6:4
ENT-A013E	10	POPC:cholesterol 6:4
ENT-A013Z	10	POPC:cholesterol 6:4
cholesterol	10	POPC:cholesterol 6:4

6.3 Results

The primary goal of the EuroNeurotrophin consortium was the design of small-molecule NT mimetics, able to activate the NT receptors. This effort built on previous successful work of some of the EuroNeurotrophin partners, who identified small-molecule neurosteroids that can act as NT mimetics by activating the NT receptors. Specifically, they have identified that DHEA and its C17-spiroepoxy derivative, BNN-27, are able to activate the TrkA and p75 NT receptors and promote neuronal survival, without causing hyperalgia.[202] In the same work, Saturation Transfer Difference Nuclear Magnetic Resonance (STD-NMR) experiments suggested that these compounds bind at the interface of the NT with the D5 domain of the TrkA receptor, thus bridging the heterodimer, and two possible binding sites at the interface of the two proteins were proposed computationally to be the most probable.[202]

In this chapter of the thesis, I performed docking of the EuroNeurotrophin compounds that were identified by the consortium to activate the TrkA or TrkB receptors (Table 6.1). These are the compounds that were identified by the consortium to phosphorylate, i.e. to activate the TrkA and TrkB receptors, as well as downstream signaling proteins, in TrkA and TrkB expressing cells, respectively. The cell survival was also verified for the compounds that activated the receptors.[204,330]

I docked the compounds presented in Table 6.1 to the two sites (site 1a and 1b) at the interface of the TrkB receptor with the NT-4/5 (Figure 6.1A), while the same docking calculations were performed at the TrkA-D5 with NGF by Alexandros Tsengenes from the MCM group at the Heidelberg Institute for Theoretical Studies (HITS) and I present them here as well for comparison. The docked poses of the compounds are shown in Figure 6.2 and Figure 6.3 for the compounds that bind selectively on TrkA and TrkB, respectively. The results for ENT-A013E/Z were published in REF.[204] The binding poses of the compounds in the two symmetric sites present complementarity to the binding pockets. For site 1b, which is partially solvent-exposed, the docking poses have the 3β-hydroxyl group pointing towards the solvent region and the C17 steroid substituents facing towards the interface of TrkA-D5 and NGF, thus providing a possible explanation for the selectivity of the substituents. The docking poses in site 1a present various possible arrangements of the steroidal core for the different compounds, indicating structurally less stable binding.

Additionally to the docking calculations, I initiated MD simulations from these poses to examine the stability of the docked compounds in the binding sites. Interestingly, the compounds detached from the binding sites during several simulation replicas in the first 100 ns without the protein structure being affected by the ligand unbinding, indicating that either the initial binding modes were not so stable, or there may be another binding site(s), to which the compounds bind with higher affinity.

Since the compounds have a steroidal core that is the same as that of cholesterol and the TM domain of TrkB has previously been identified as a drug target for anti-depressant drugs, such as fluoxetine,[91] I performed simulations to test the membrane permeability of the compounds. I placed the molecules close to POPC or POPC:cholesterol 6:4 membranes to test their tendency to enter into the membrane (Figure 6.1B). I did these simulations with one compound close to the membrane (Figure 6.1B left), and the same simulations with 10 molecules close to the membrane were performed by Alexandros Tsengenes (Figure 6.1B right).

Simulations with one compound close to the membrane showed that it can insert spontaneously in all replicas sometimes in the first 10 ns, while when 10 compound molecules were placed close to the

membrane, they created aggregates which can also insert into the membrane and then diffuse in it (Figure 6.4). This shows that the compounds could also be found in the membrane environment and possibly interact with the TM domain of the receptors. Thus, I performed simulations with one compound docked to the interface of the TM helices, as shown for fluoxetine,[91] or with 10 compounds around the TM domain (Figure 6.1C). However, no specific binding was observed in either case, with the compounds either dissociating from the interface of the helices in the first 100 ns or forming transient interactions with the length of the helices. This means that there might be more than one binding site for these compounds on the receptors, which can then act synergistically and lead to receptor activation. Also, since their activity against the NT receptors might be an indirect effect, direct binding assays for the ENT compounds to the receptors would be advisable.

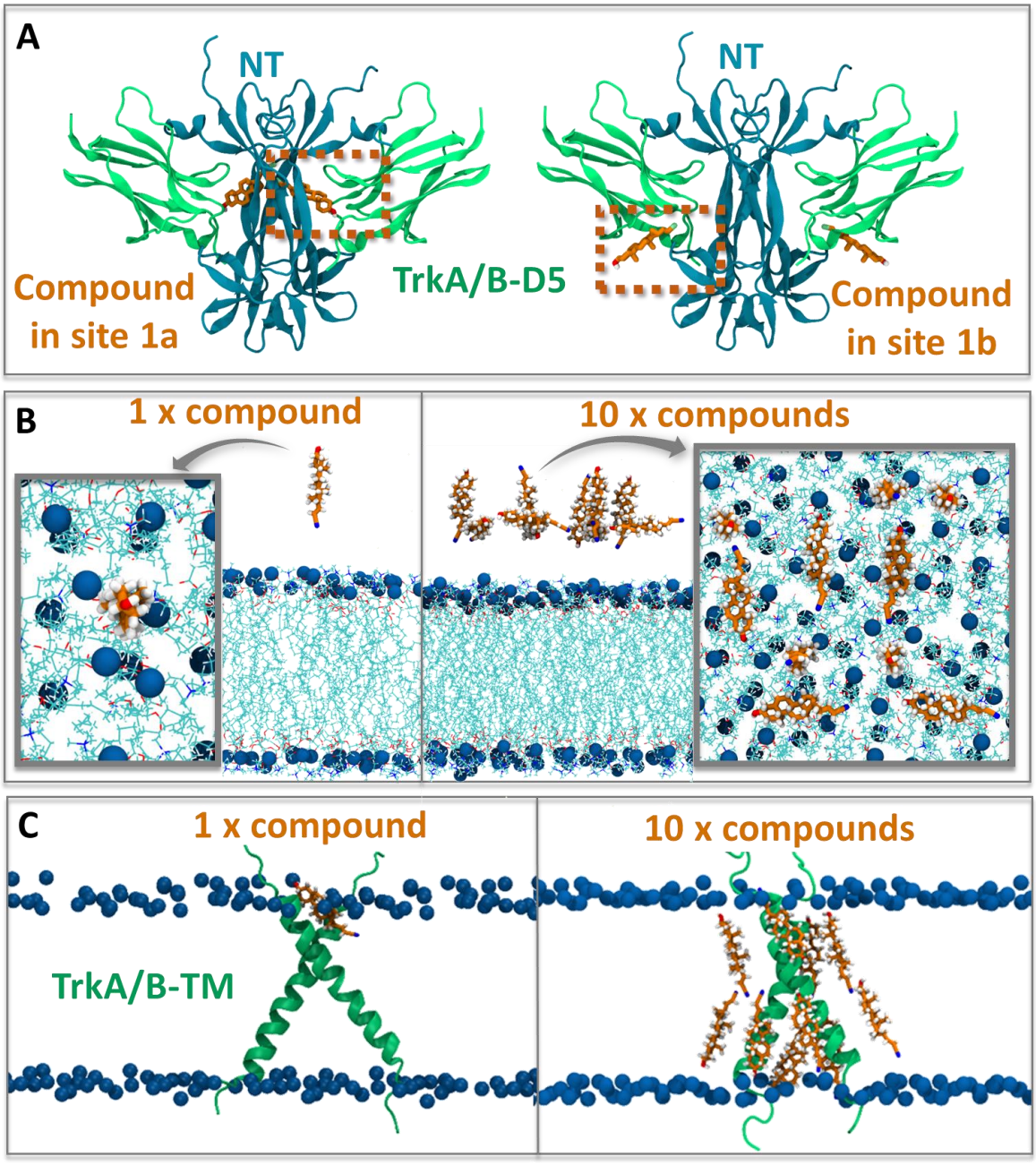


Figure 6.1: Summary of the initial configurations of the simulated systems with the neurotrophin mimetics. One figure is shown for each system, but the same systems were tested for both TrkA and TrkB receptors. (A) Compounds bound at the interface of the D5 domain of TrkA and TrkB receptors with the NTs NGF and NT-4/5, respectively. (B) One and ten compound molecules placed close to the bilayer for membrane permeability simulations shown in perpendicular views. (C) One or ten compounds placed at the interface or close to the TM helices of TrkA and TrkB receptors. The P atoms of the POPC lipids are shown in blue vdW representation.

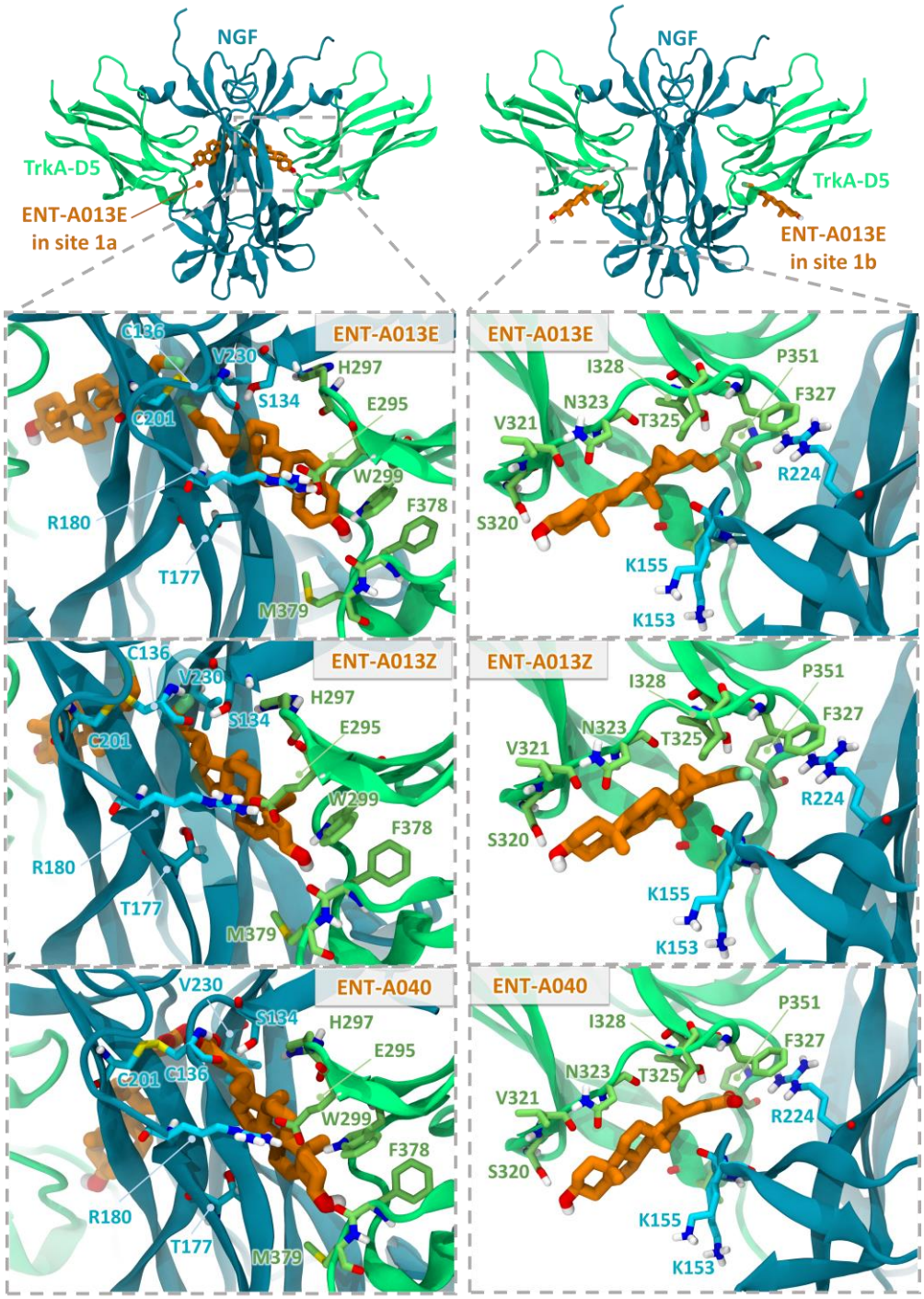


Figure 6.2: Docking poses of two compounds, ENT-A013E/Z and ENT-A040, in sites 1a and 1b at the two symmetry-related TrkA-D5 – NGF interfaces of the neurotrophin-receptor complex. The close-up views show residues lining the two sites. TrkA-D5 and NGF are shown in green and blue new cartoon representation, respectively, with selected residues in stick representation colored by atom type. The compounds are shown in orange stick representation. The figure was made by Alexandros Tsengenes.

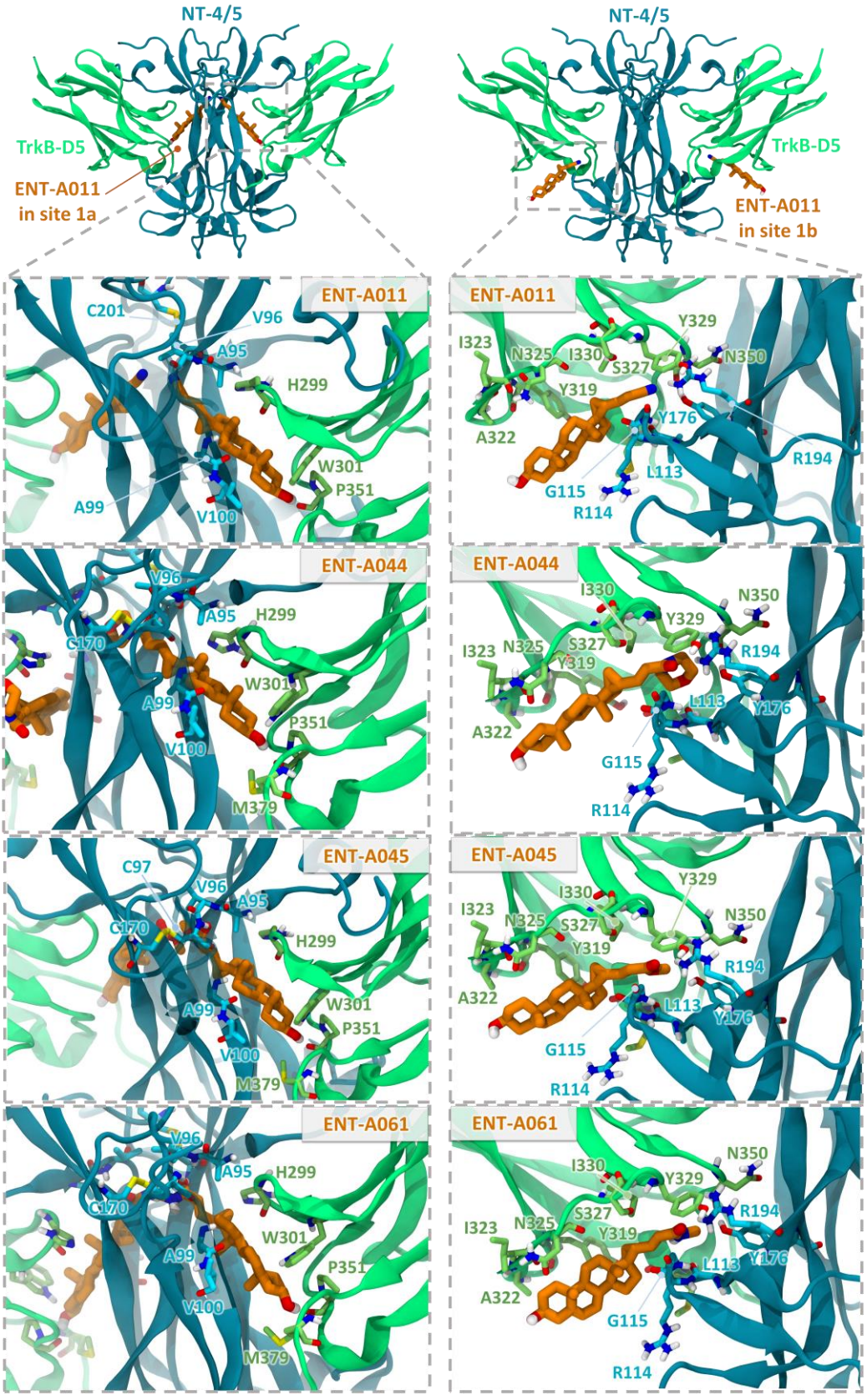


Figure 6.3: Docking poses of four compounds, ENT-A011, ENT-A044, ENT-A045 and ENT-A061, in sites 1a and 1b at the two symmetry-related TrkB-D5 – NT-4/5 interfaces of the neurotrophin-receptor

complex. The close-up views show residues lining the two sites. TrkB-D5 and NT-4/5 are shown in green and blue new cartoon representation, respectively, with selected residues in stick representation colored by atom type. The compounds are shown in orange stick representation.

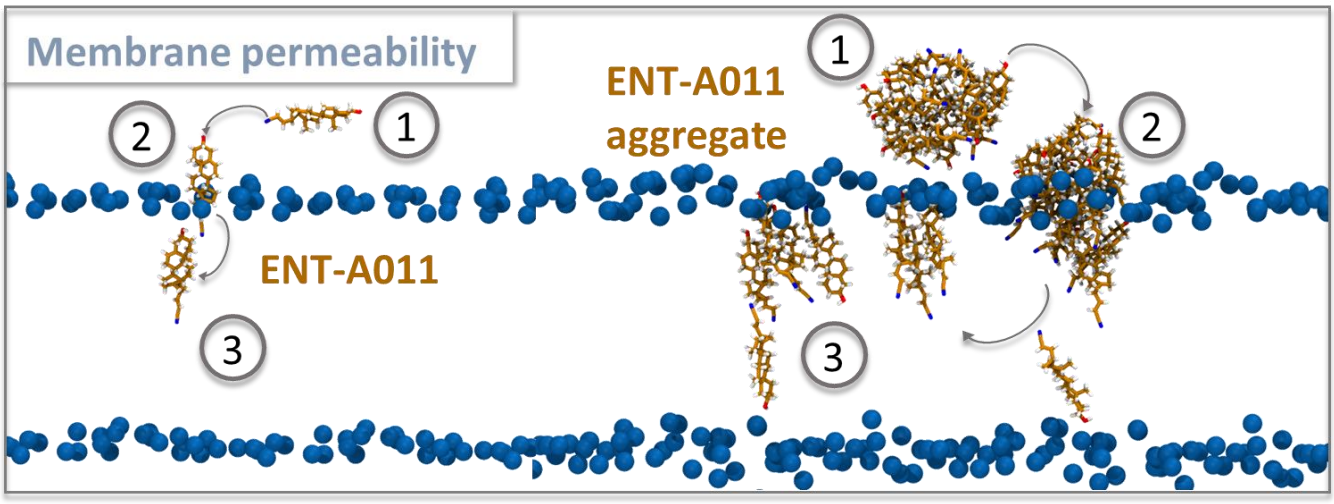


Figure 6.4: Membrane permeability simulations for one and ten molecules placed close to a membrane. In the figure, an example with compound ENT-A011 is shown. Specifically, simulation snapshots are shown with compound ENT-A011 entering at consecutive simulation times a POPC membrane as a single molecule (left) and as an aggregate of ten molecules (right). The P atoms of the POPC lipids are shown in blue vdW representation.

6.4 Conclusions

Small organic molecules that can act as neurotrophin mimetics and activate the neurotrophin receptors have the potential to provide an alternative to neurotrophin administration for neurodegenerative diseases. In the EuroNeurotrophin consortium, small-molecule derivatives of the neurosteroid DHEA have been identified to activate the TrkA and TrkB receptors, as well as downstream signaling proteins, and induce cell survival.

In this chapter of the thesis, I investigated the potential interactions of these compounds with the neurotrophin receptors with molecular docking and MD simulations. Specifically, I docked the compounds to the interface of D5 domain of the receptors with the neurotrophins, in agreement to the

binding sites previously identified by STD-NMR.[202] The binding poses could provide a rationale for the mode of action of the compounds. However, MD simulations starting from these poses showed that the compounds can quite easily dissociate form these binding sites, indicating weak binding. However, the dissociation could be due to incorrect force field parameters for the small molecules, although the similarity of the compounds with cholesterol, which is well parameterized in the CHARMM36 force field, render this scenario less probable. Another possible reason for the quick dissociation of the compounds is incorrect initial binding poses in the receptors. Even a slightly wrong orientation of the compounds in the binding sites could omit important interactions with the receptor that could stabilize the complex with the compounds. In either case, the docking models were not able to explain the selectivity of the compounds for TrkA or TrkB. This alone indicates a weakness of these docking models and could be the reason for dissociation.

Since these molecules have a steroidal core, identical to that of cholesterol, and since the TM domain of TrkB has been previously identified to be a drug target for anti-depressant drugs, such as fluoxetine,[91] I investigated the possibility that these compounds act through binding to the TM domain of the NT receptors. Thus, I performed simulations with the compounds close to the membrane, which resulted in the molecules spontaneously entering the membrane, either as single molecules or as aggregates. This shows that the compounds can easily penetrate the cell membrane and possibly interact with the TM region of the receptors. Of course, the NT receptors, as discussed in the previous chapter, lie in neuronal membranes, in which penetration of the compounds would be more difficult because neuronal membranes are much more rigid than the membranes tested here. Simulations of the compounds with neuronal membrane models would most probably lead to higher times for small molecule insertion in the membrane. However, the successful insertion of the compounds to POPC:cholesterol 6:4 membranes indicates that even more rigid membrane can accommodate the NT mimetics.

Therefore, I performed simulations with one compound molecule bound at the interface of the TM helices of the receptors, as fluoxetine,[91] or with several molecules around the helices. However, as for the simulations with the EC domains, the compounds either detached from the interface of the helices or created transient interactions with them. This suggests that either more than one binding site exists that can act synergistically to invoke activation, or the activity of the compounds is an indirect effect of binding in other types of membrane receptors that can then activate the NT receptors. In either case, more extensive investigation of the mechanism of action of these neurosteroids is deemed necessary for the design of more potent and selective NT receptor agonists.

7 Final conclusions and future perspectives

In this thesis, I have provided an overview of the computational work that I carried out for the investigation of the mechanism of action of TrkA and TrkB neurotrophin (NT) receptors and several NT mimetics. I have split the work done on the receptors based on their extracellular (EC), transmembrane (TM), and intracellular (IC) segments, as well as the full-length receptors. Specifically, I studied with molecular dynamics simulations: (i) the effects of glycosylation on the dynamics of the EC domains of the receptors, (ii) the dimerization interface of the TrkA and TrkB TM homodimers with all-atom (AA) and coarse grained (CG) unbiased and biased MD simulations, (iii) the conformations that are adopted by the full-length TrkA and TrkB homodimers, as well as their interactions with a neuronal membrane model and finally, (iv) the potential binding modes of NT mimetics to the receptors and their membrane permeability. The final conclusions and future perspectives on this work are discussed below.

Glycosylation effects on the conformation of the EC domains of the NT receptors

I modeled the N-glycosylation of the full EC domains of the TrkA and TrkB domains by attaching the first five oligosaccharides of the N-glycan core at specific Asn residues based on experimental information from mass spectrometry and crystallography. I simulated with MD the EC domains of TrkA and TrkB receptors in their glycosylated and non-glycosylated form at three different oligomerization states: as (i) dimers bound to NGF and NT-4/5 neurotrophins, respectively, (ii) as monomers bound to the NTs, and (iii) as monomers alone. The results showed that the D5 domain of TrkB is less shielded by glycans compared to TrkA, which might be important for NT binding. Also, the accessible surface area of the receptors in the presence of the glycans that can be accessed by large molecules is smaller compared to small molecules. This may be relevant for protection of the receptors from proteolytic cleavage by e.g. secretases, which is something that happens physiologically for the p75 receptor.[280] Additionally, the glycans were identified to form interactions with the bound NTs and increase the contact area between receptor and NT.

An important finding of the MD simulations was the ability of the glycans to affect the conformation of the EC domains of the receptors. Large hinge motions were detected between the D13 and D4 domains of the receptors, with the presence of glycans promoting more extended conformations. In

fact, in some replica simulations without glycans, the D13 domain bent and interacted with the D5 domain. This happened in both TrkA and TrkB receptors. This could affect NT binding to the D5 domain, by making the binding site less accessible. Also, in some simulations with the receptor dimers, the D13 domains from each monomer interacted with each other in a closed configuration which could also hinder NT binding to the D5 domain. The ability of the glycans to promote extended, open conformations of the receptors is mostly due to steric hindrance of the closed conformations when the glycans are present. This is in agreement with the fact that many of the glycans are attached at the linkers between the subdomains or close to them. This way, the glycans can limit the hinge motions of the receptors and keep them in more extended conformations, which might facilitate neurotrophin binding.

In future work, the findings in this thesis on the importance of the glycans in preserving open NT receptor configurations could be corroborated by enhanced sampling simulations, which could increase the statistical sampling of the open and closed states of the receptors and improve the sampling of relevant configurations. Also, enhanced sampling methods could determine the relative free energy of the different states and provide the transition paths from one state to the other. This is quite challenging because appropriate reaction coordinates along the slow degrees of freedom need to be determined. Experimental validation of the open and closed states is also very important, which could be achieved with structural biology techniques. Most importantly, the results presented here highlight the importance of taking glycosylation into account in either experimental or computational studies of the NT receptors.

Arrangements of TM helices of NT receptors

The TM domains of NT receptors, as for other receptor tyrosine kinases, are responsible for conveying the activation signal from the EC to the IC domains. This is achieved by a rearrangement of the TM helices in the dimeric state of the receptors, which leads to a transition from the inactive to an active arrangement. Mutagenesis experiments have suggested that, for TrkA, the two states correspond to two distinct interfaces, with the inactive arrangement being solved with NMR.[60] In this thesis, I investigated the arrangements of the TM helical homodimers of TrkA and TrkB receptors with AA and CG unbiased and MD simulations in micelles and metadynamics simulations in bilayers. The goal was to identify all possible arrangements of the TM helices, which can be crucial for the design of small molecules that bind at the interface of these helices, as recently discovered for the TrkB receptor.[91]

The simulations revealed that the CG simulations provide better sampling and more plausible TM arrangements with much fewer computational resources. For this reason, the CG simulations were further analyzed. For TrkA, various different arrangements occurred which differ from the NMR structure, even at varying simulation conditions, which raises questions about the conditions under which the NMR structure was retrieved. From the sampled arrangements of TrkA, the most dominant one has the helices interacting through residues that have been identified to be part of the active interface of the dimer,[60] which suggests that this configuration might correspond to the active state of the receptor. This arrangement could be targeted with structure-based drug design techniques to develop TrkA agonists that stabilize the active state of the TM domain, similarly to the recently identified TrkB agonists.[91] For TrkB, three dominant arrangements were sampled, with the most populated having the AxxxG motif at the interface, and thus it could also correspond to the active arrangement of the TrkB receptor. Additional enhanced sampling CG metadynamics simulations were performed, which allowed the calculation of the free energy difference of the various states, revealing a stronger binding for the TrkA helices of 10 kJ/mol compared to TrkB.

The arrangements of the TM helices of the TrkA and TrkB receptors that have been identified in this thesis might depend also on other additional factors, such the micelle environment or the composition of the membrane. Here, the unbiased simulations were performed in micelles, which might favor different arrangements from those that occur in a cell membrane. Also, the metadynamics simulations were performed in a simple POPC membrane whose properties can differ significantly from those of a complex, asymmetric neuronal membrane. The coarse grained nature of the Martini force field could potentially lead to inaccurate description of steric effects in the binding of the helices. Additional atomistic simulations in various membrane environments could shed light on the sensitivity of the TM arrangements on these factors. Structure-based drug design of small molecules that can target the active TM arrangements of the TrkA helices could potentially lead to novel agonists, as it was discovered for TrkB.[91] However, the few residues that compose the interface of the TM helices would make it hard to design specific TrkA agonists due to the limited available residues for selective interactions. This would probably lead to low affinity agonists, which could potentially offer though a solution to NT receptor activation.

Full-length NT receptors in the neuronal membrane

The TrkA and TrkB NT receptors are TM receptors, with EC, TM and IC domains. While there are experimental structures of the three different domain segments, there is no model, experimental or computational, of their full-length structures. In this thesis, I have created models of the full structures of the TrkA and TrkB receptors in their glycosylated homodimeric state bound to the NTs NGF and NT-4/5, respectively. These receptors localize in neuronal membranes and specifically in membrane microdomains, called lipid rafts, which are rich in cholesterol and sphingomyelin.[112,114] Lipids found in neuronal membranes, like gangliosides and ceramides, have also been found to activate the NT receptors.[119–121] This shows the importance of the neuronal membrane as the environment of these receptors, and thus I created here a model of neuronal membrane.

I tested the behavior of the neuronal membrane by simulating it first with a smaller test protein system containing the TM and IC domains of TrkA and TrkB as monomers. These test simulations verified the normal behavior of the neuronal membrane model, and they also revealed some interesting interactions between the kinase IC domains and the neuronal membrane. The kinase IC domain of both TrkA and TrkB, which was initially placed at a distance from the membrane, approached and interacted with the inner leaflet of the neuronal membrane in all replica simulations. Interaction analysis revealed that three positively charged regions in the kinase domain of TrkA interacted with negatively charged lipids of the inner leaflet of the membrane. For the kinase domain of TrkB no specific interaction patterns were identified. This specific orientation of TrkA with respect to the membrane leaves exposed the area in which the activation loop would reside in its active conformation, and thus would be accessible for phosphorylation by the other kinase monomer.

Then, I performed MD simulations of the full-length TrkA and TrkB models embedded in the neuronal membrane model. During these simulations, the EC domains approached the membrane and bent towards it. The TrkB-EC domain came closer to the membrane than TrkA, probably because of its longer EC linker. Interactions between the EC domain of TrkA and the GM1 ganglioside were identified, which might explain the ability of GM1 to activate TrkA.[119,121,125] The kinase domains in the full-length systems interacted with the membrane, similarly to what was observed in the TM-IC systems. The regions of TrkA kinase that interacted with the membrane were the same as those observed for the TM-IC systems, which means that this arrangement is also possible in the dimeric state of the receptor. However, the two kinases were not in an arrangement where one could phosphorylate the other, and thus the currently sampled arrangements probably show how the kinases behave in the inactive state of the

receptor. The TrkA kinase domains interacted in more than one way with the membrane. The lack of a dominant arrangement might be the reason why the TrkB kinase has a 2-fold lower catalytic efficiency than TrkA and becomes autophosphorylated *in vitro* more slowly than TrkA.[312] This also leaves different regions on the kinase domain exposed for interactions with binding partners which can lead to different signaling cascades. Indeed, activation of TrkA and TrkB in the same cell type can yield very different cellular outcomes.[312]

The results in this thesis reveal some interesting features of the conformations that can be adopted by the full-length NT receptors in a neuronal membrane and their interactions with it. In order to increase the statistical significance of the observed events, enhanced sampling techniques could be used. Additionally, in the current thesis, the simulations started from the inactive state of the receptors, and due to incomplete sampling, the active state was never sampled. Perhaps extensive enhanced sampling could give a solution to this as well, or initiating future simulations from the already active state would be more efficient. To date, there is no experimental information on the conformations adopted by the NT receptors in a membrane. New information provided by structural biology techniques would be valuable for the validation of the current results and for providing new directions in the study of the mechanism of action of NT receptors. This would not be without challenges though, as the membrane environment of the receptors and their flexible linkers cause difficulties and can lower the resolution.

Small molecule NT mimetics

The work presented here has been part of the EuroNeurotrophin Innovative Training Network (ITN), which has focused on the design of small molecule NT mimetics that can act as agonists of the NT receptors. In the current thesis, I have investigated small-molecule derivatives of the neurosteroid DHEA that have been identified by the consortium to activate the TrkA and TrkB receptors, as well as downstream signaling proteins, and induce cell survival. Specifically, I examined the interactions of the compounds with the EC domains of the receptors through molecular docking and MD simulations. Even though docking poses that have been previously supported by STD-NMR[202] were also identified here, the simulations showed that the poses were not stable, with the compounds dissociating from the receptors. This could be due to incorrect initial binding poses, which were also not able to explain the

selectivity of the compounds for TrkA or TrkB. This shows the need for crystal or NMR structures that would clearly show the binding mode of the compounds.

Due to the cholesterol core of these compounds, I tested the ability of the compounds to enter the membrane. Simulations with the compounds close to the membrane were initiated, which resulted in the molecules spontaneously entering the membrane, either as single molecules or as aggregates. This shows that the compounds can easily penetrate the cell membrane and possibly interact with the TM region of the receptors. Thus, I investigated the possibility that these compounds act through binding to the TM domain of the NT receptors. Simulations started with the compounds either close to the TM domains within the membrane bilayer, or at the interface of the TM helices. However, no specific interactions between compounds and receptors were identified in this case either.

These results suggest that either more than one binding site exists that can act synergistically to invoke activation, or the activity of the compounds is also an indirect effect of binding to other types of membrane receptors that can then activate the NT receptors. In either case, more extensive investigation of the mechanism of action of these neurosteroids with STD-NMR, crystallography and mutagenesis assays is deemed necessary for the design of more potent and selective NT receptor agonists.

8 Acknowledgments

I would like to thank everyone who supported me professionally and personally throughout the past years and enabled me to perform the work presented here, as well as write this dissertation. First of all, I would like to thank my supervisor Prof. Dr. Rebecca Wade, who gave me the opportunity to work in her group and in this interesting and challenging research project. Most of all, I would like to thank her for being an inspiring role model and mentor for me and for showing me the way to independent scientific thought.

I would also like to thank the person with whom I collaborated and spend most of the PhD time with, Alexandros Tsengenes, who also did his PhD on the EuroNeurotrophin project. I would like to thank him for the endless discussions on the difficulties and future directions in our project. I would not have managed to produce all the results in this thesis without his catalytic help. Mostly, I would like to thank him for his support on a personal level and for giving me strength to continue during the challenging times of the project.

Another person with whom I collaborated a lot is Ainara Claveras Cabezudo, who did her Masters' thesis on the EuroNeurotrophin project. Working and discussing with her on the transmembrane work of the project enabled me to think about new aspects of the project, while her hard work helped progress this challenging project. Most importantly, I would like to thank her for her positive attitude in all aspects of work, and for all the fun time that we had in the lab.

I further would like to thank my second supervisor and TAC member Prof. Dr. Ursula Kummer and my third TAC member Dr. Daniela Mauceri for all their helpful feedback in my TAC meetings. I would also like to thank the 3rd and 4th commission members, Prof. Dr. Claudio Joazeiro and Dr. Stefan Pfeffer, for agreeing to be part of my examination committee.

I would like to thank all the members of the Molecular and Cellular Modeling (MCM) group at the Heidelberg Institute for Theoretical Studies (HITS) in the years 2018-2022. They made my PhD time more interesting with scientific and various funny discussions. Special thanks to Dr. Stefan Richter for all the technical support, Dr. Goutam Mukherjee for his help in coarse-grained simulations and for always supporting me in any possible way, Dr. Giulia Paiardi for her help in glycosylation modeling, Dr. Ariane Nunes Alves for her mentoring and Jonathan Teuffel for his help with the german translation of the thesis

summary. I would also like to thank Dr. Giulia D'Arrigo and Marco Rizzo for all the fun time we had in the lab.

My PhD was part of the EuroNeurotrophin consortium, and I would like to thank all the partners for the collaboration in this challenging project. I would like to acknowledge my PhD work was funded from the European Union's Horizon 2020 research and innovation programme "Euroneurotrophin" under the Marie Skłodowska-Curie grant agreement No. 765704 and the Klaus Tschira Foundation. The computations presented in this thesis were supported by the state of Baden-Württemberg through bwHPC and the German Research Foundation (DFG) through Grant INST 35/1134-1 FUGG.

Θα ήθελα τέλος να ευχαριστήσω την οικογένειά μου και τους πιο στενούς μου φίλους για την υποστήριξη και το κουράγιο που μου προσέφεραν όλα τα χρόνια των σπουδών μου, ακόμα κι όταν ήμουν μακριά τους. Σας ευχαριστώ όλους!

9 Appendix

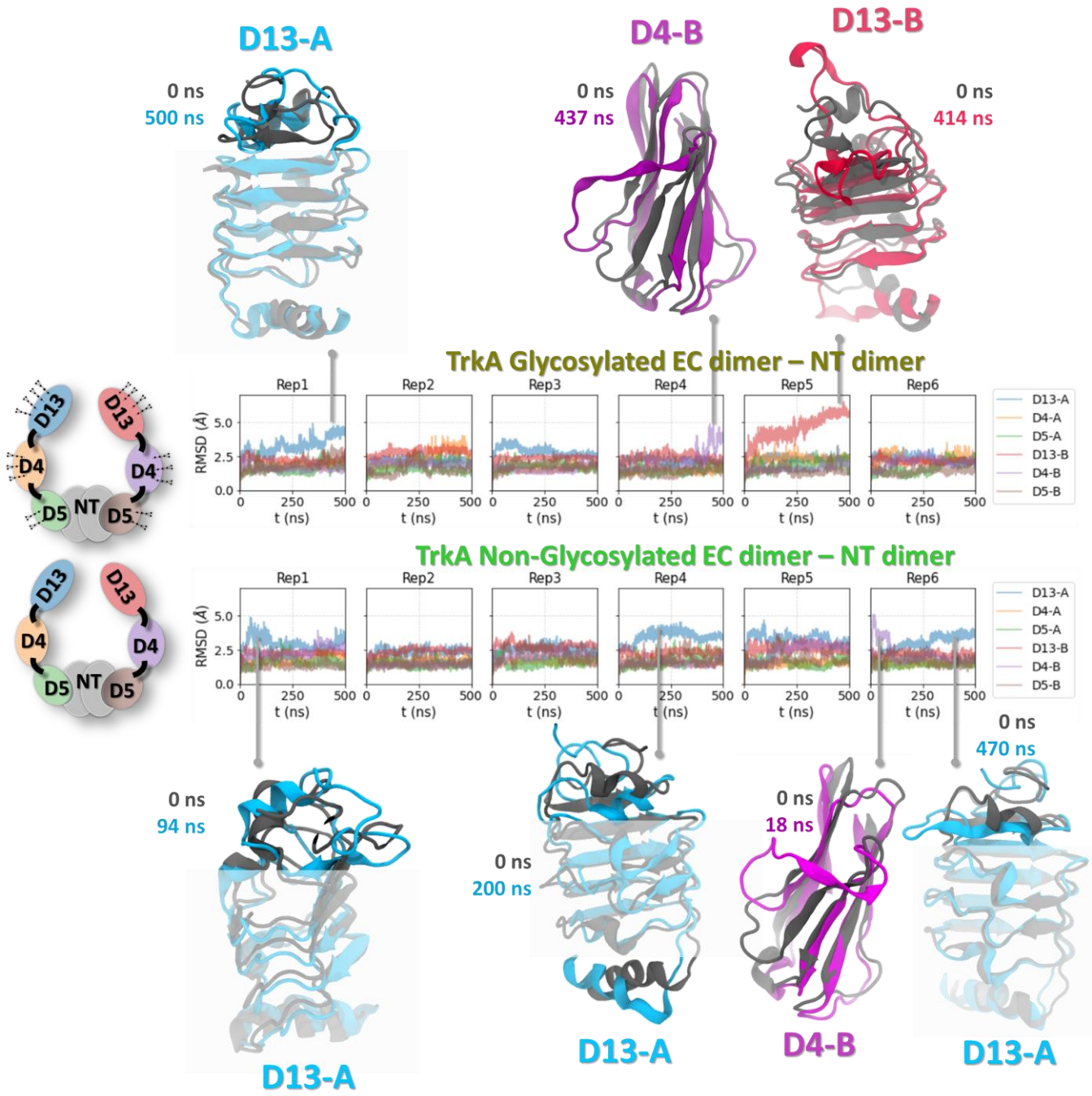


Figure A.1: RMSD calculations for each domain of the glycosylated and not TrkA-EC homodimer bound to NGF system. For each system the time-dependent RMSD of the domain backbone compared to the first frame is shown for each replica. Before the calculation of the internal RMSD of each domain, I performed alignment in the respective domain at each simulation snapshot. Schematic representation

of each system, with color coding of the domains is shown at the left. In cases of high RMSD values, the conformations of the domains are shown in cartoon representation aligned to the first frame.

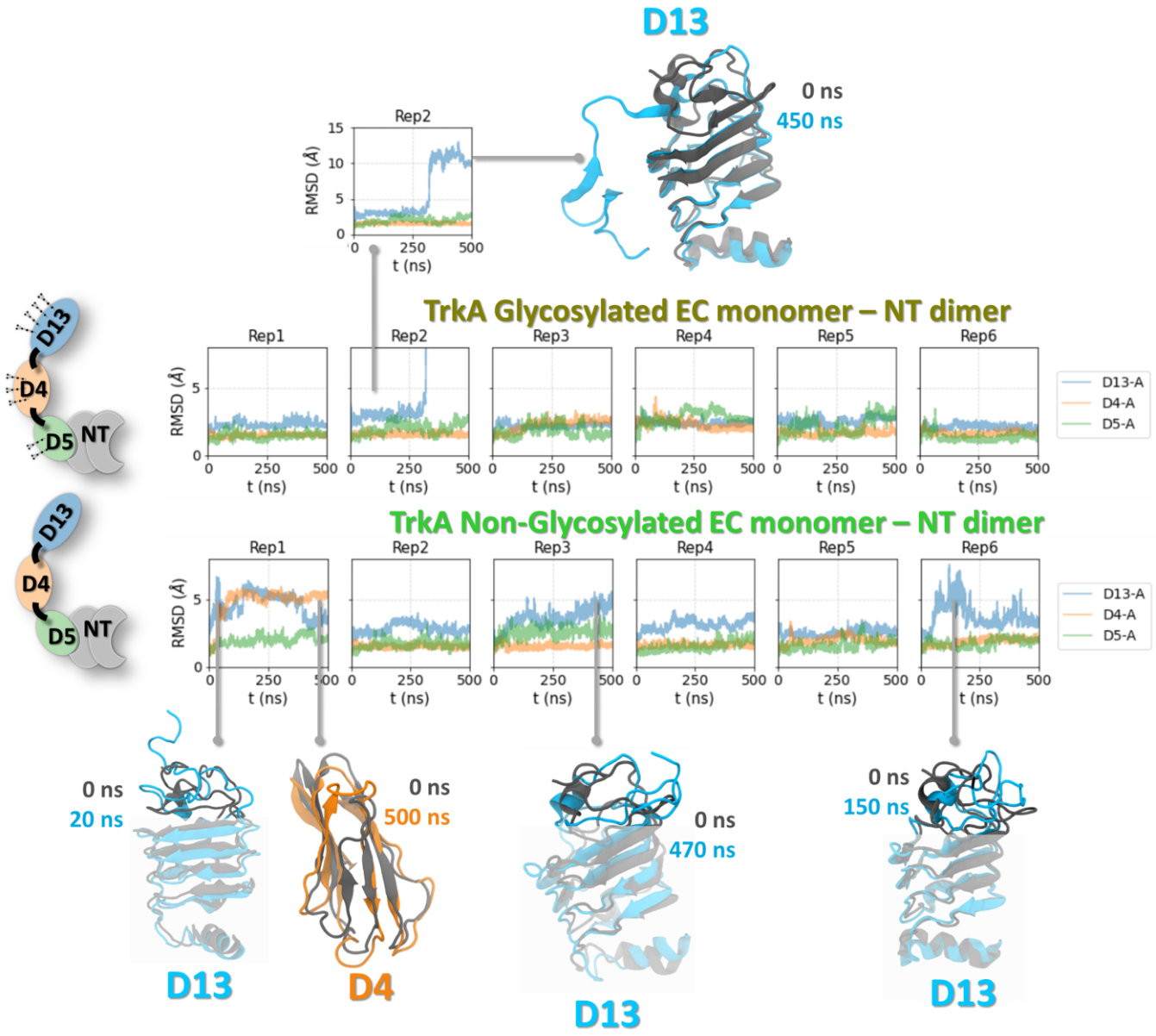


Figure A.2: RMSD calculations for each domain of the glycosylated and not TrkA-EC monomer bound to NGF system. For each system the time-dependent RMSD of the domain backbone compared to the first frame is shown for each replica. Before the calculation of the internal RMSD of each domain, I performed alignment in the respective domain at each simulation snapshot. Schematic representation of each

system, with color coding of the domains is shown at the left. In cases of high RMSD values, the conformations of the domains are shown in cartoon representation aligned to the first frame.

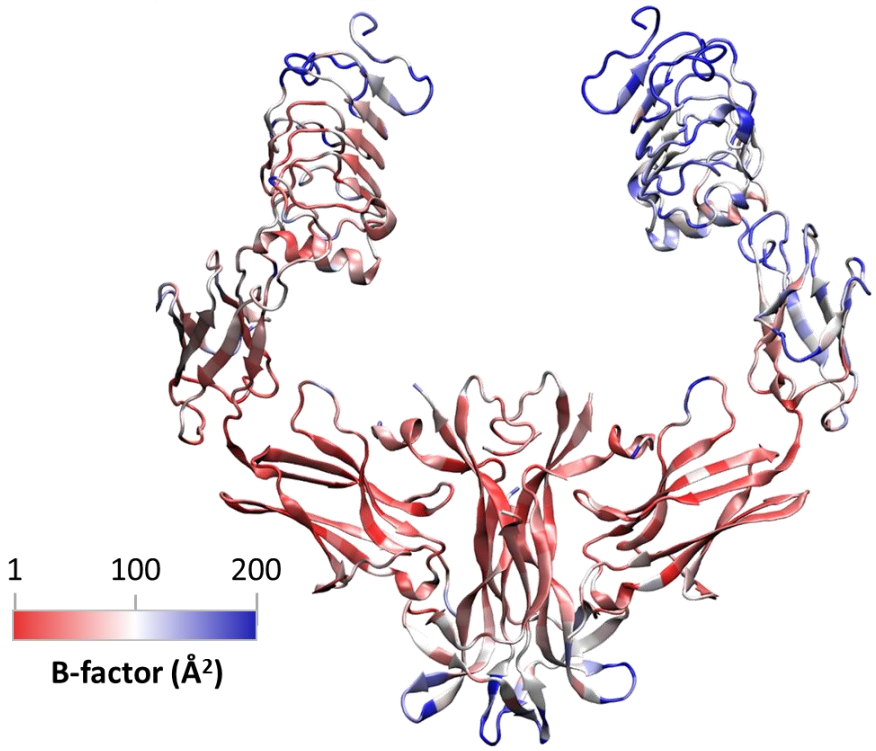
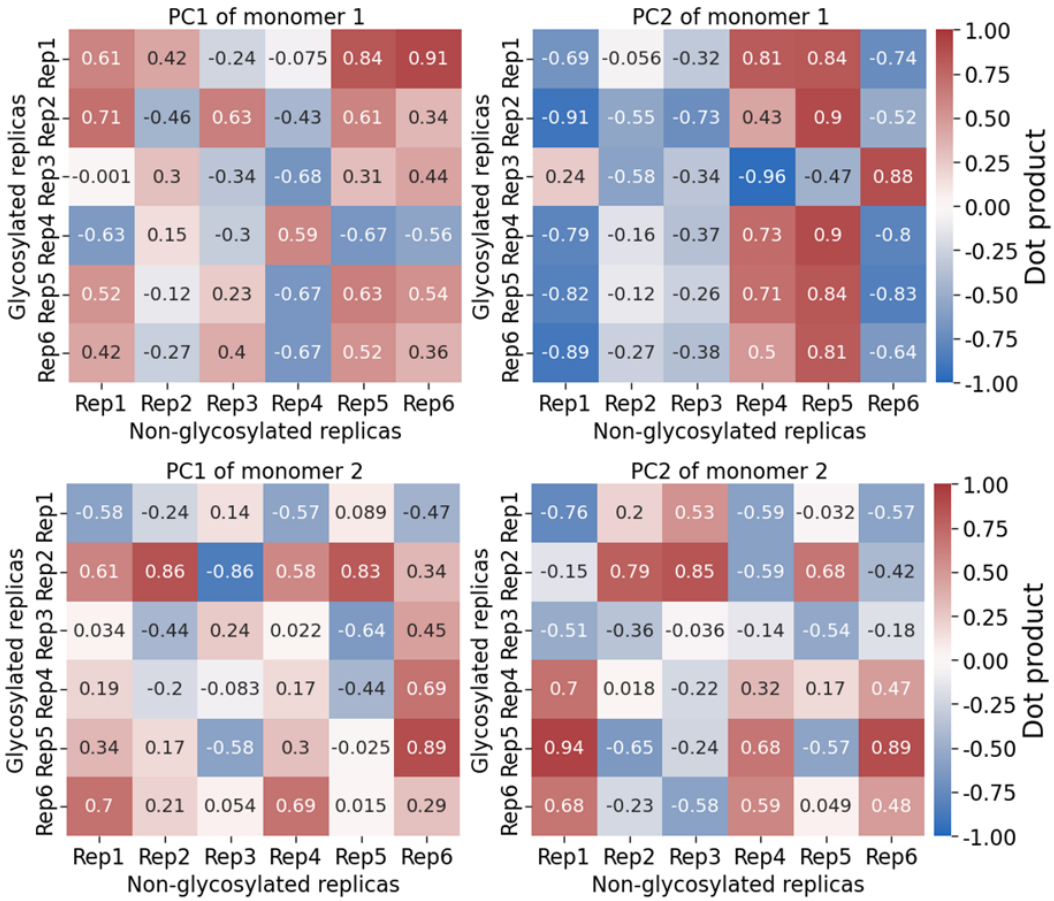
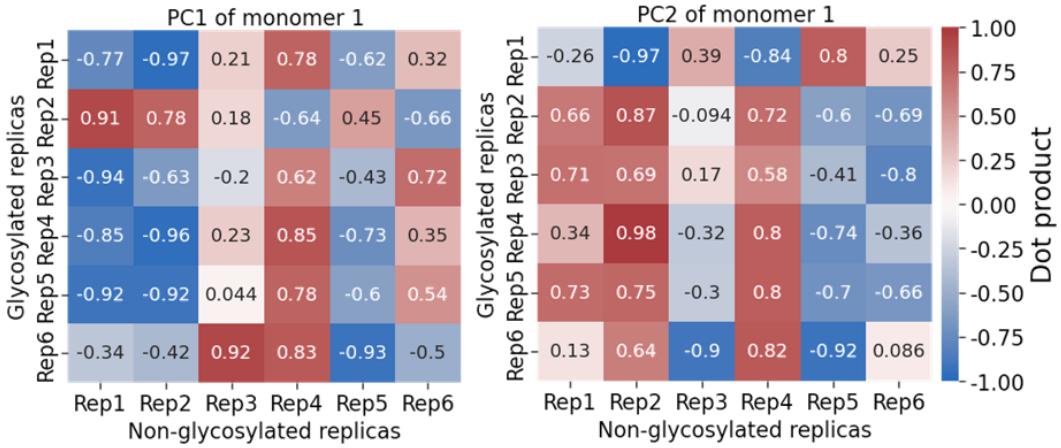


Figure A.3: B-factors of the TrkA-EC bound to NGF from the crystal structure (PDB ID: 2IFG).[28]

TrkA EC dimer – NT dimer



TrkA EC monomer – NT dimer



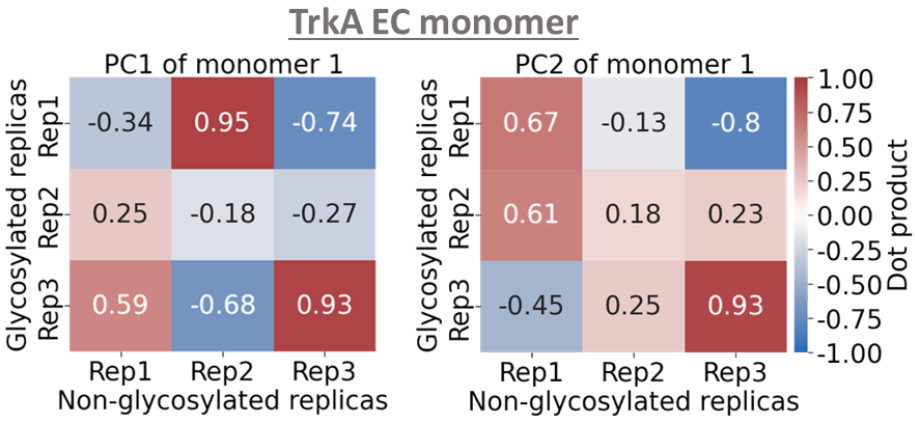
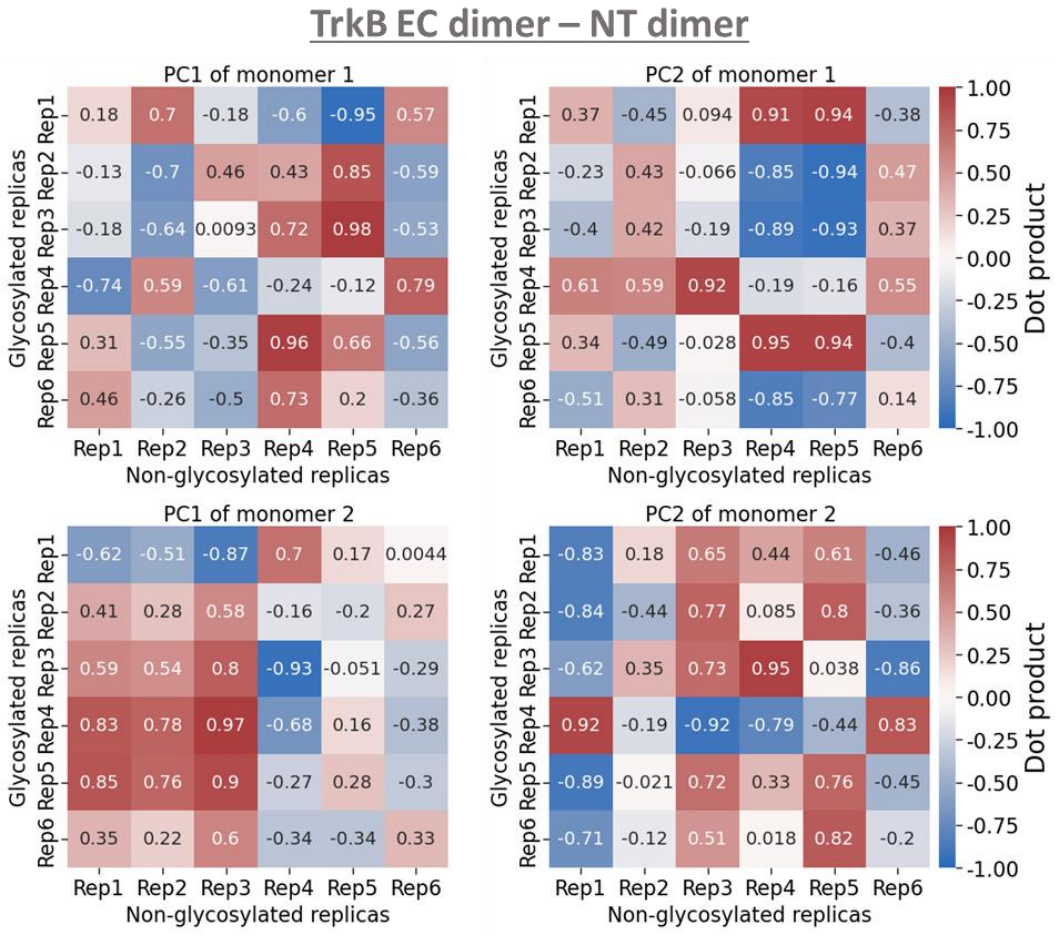


Figure A.4: Calculation of the dot products of the first (PC1) and second (PC2) principal components between glycosylated and non-glycosylated replica pairs of TrkA systems. The PCA was performed on the D13 domain only, after alignment of the trajectories for the D4-D5 domains. In the systems of the receptor dimers, the PC1 and PC2 dot products were calculated for each monomer separately. The dot products are visualized as heatmaps, and the values mean: [1] parallel vectors, [-1] anti-parallel vectors and [0] vertical (or orthogonal) vectors.



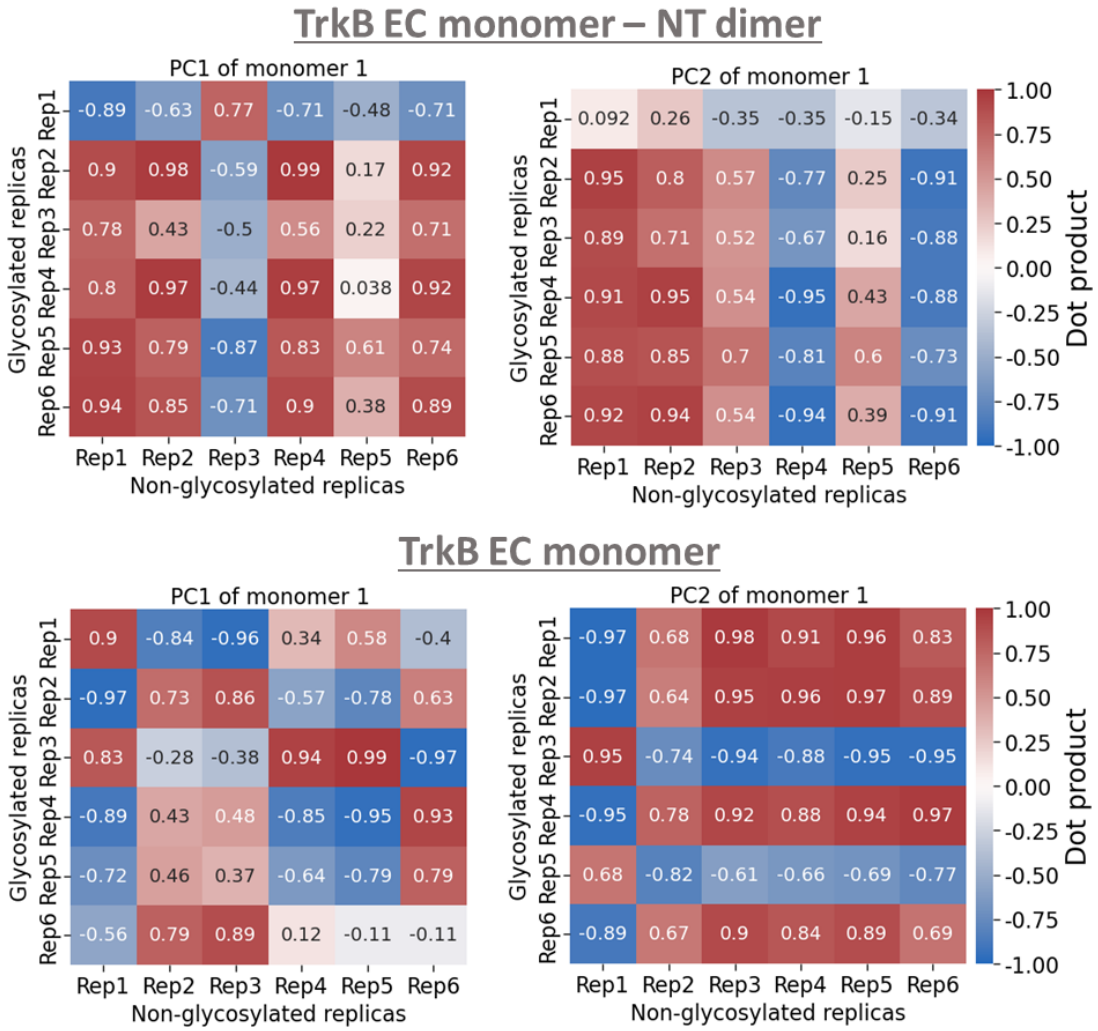


Figure A.5: Calculation of the dot products of the first (PC1) and second (PC2) principal components between glycosylated and non-glycosylated replica pairs of TrkB systems. The PCA was performed on the D13 domain only, after alignment of the trajectories in the D4-D5 domains. In the systems of the receptor dimers, the PC1 and PC2 dot products were calculated for each monomer separately. The dot products are visualized as heatmaps, and the values mean: [1] parallel vectors, [-1] anti-parallel vectors and [0] vertical (or orthogonal) vectors.

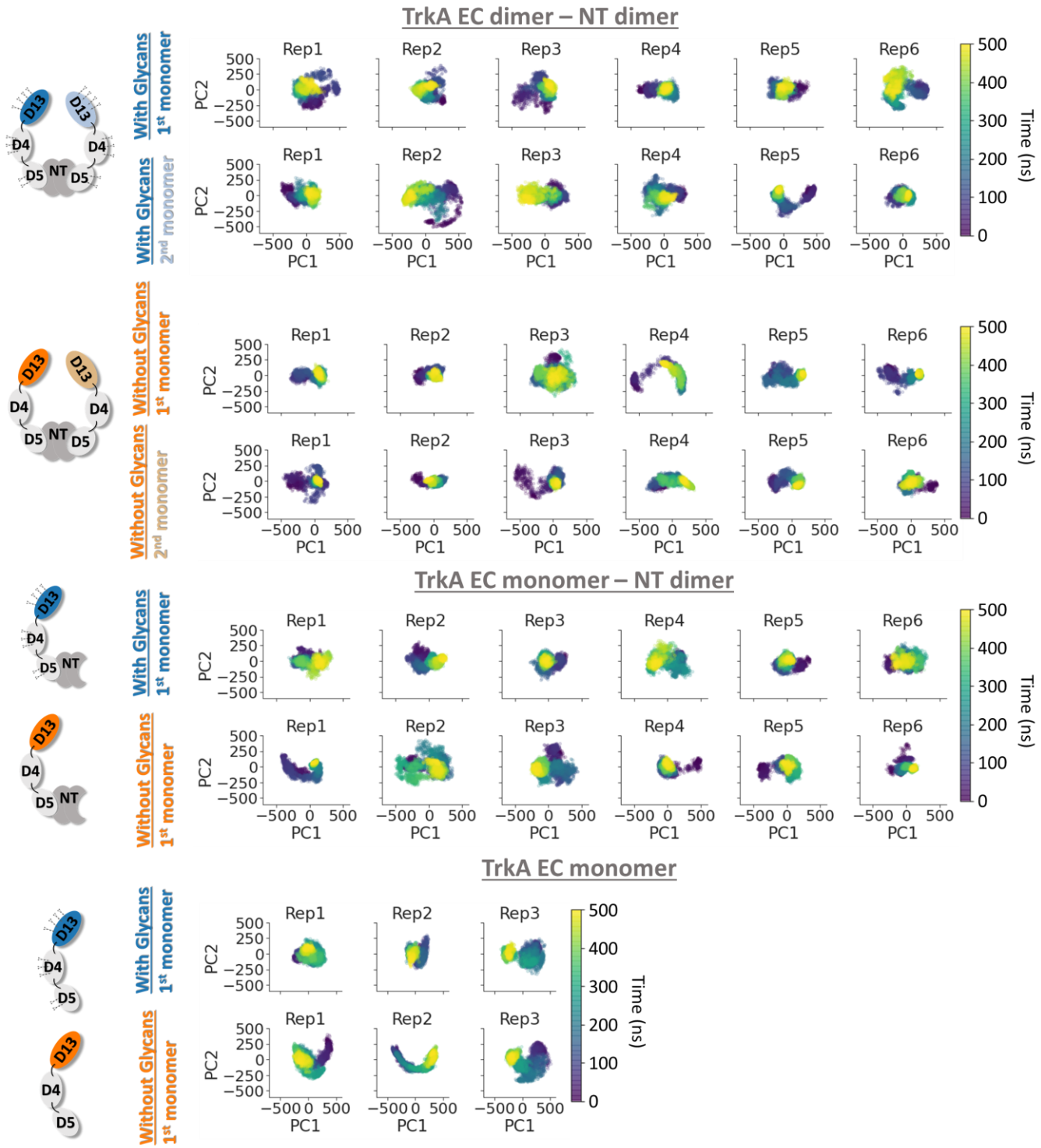


Figure A.6: Time evolution of the first and second principal components of the D13 domain during the simulations of the glycosylated and not TrkA systems. PCA analysis was performed to the D13 domains after alignment of each EC segment to the D4 and D5 domains. The color coding of each domain is shown in the schematic representations of the systems at the left of the plots.

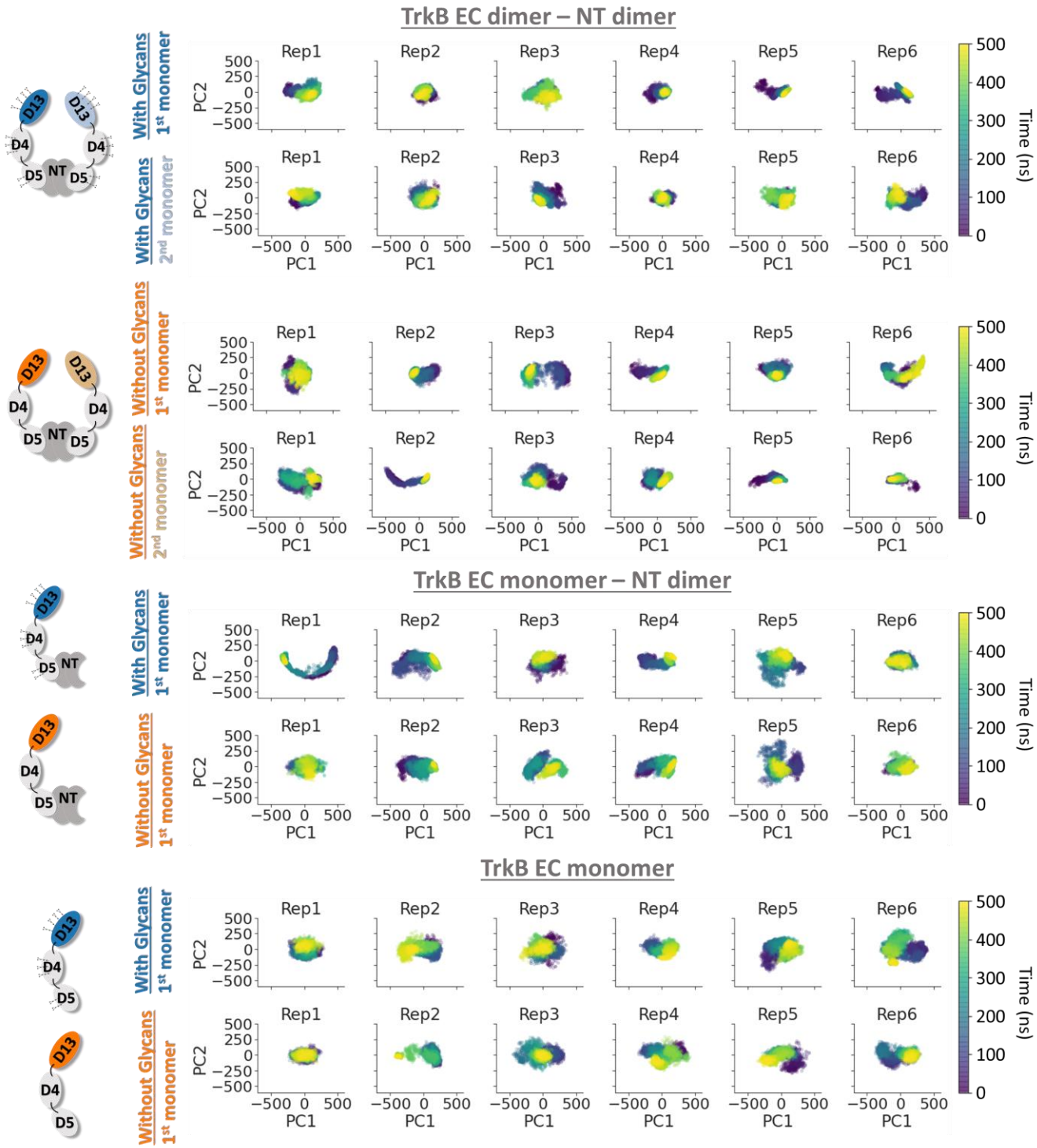


Figure A.7: Time evolution of the first and second principal components of the D13 domain during the simulations of the glycosylated and not TrkB systems. PCA analysis was performed to the D13 domains after alignment of each EC segment to the D4 and D5 domains. The color coding of each domain is shown in the schematic representations of the systems at the left of the plots.

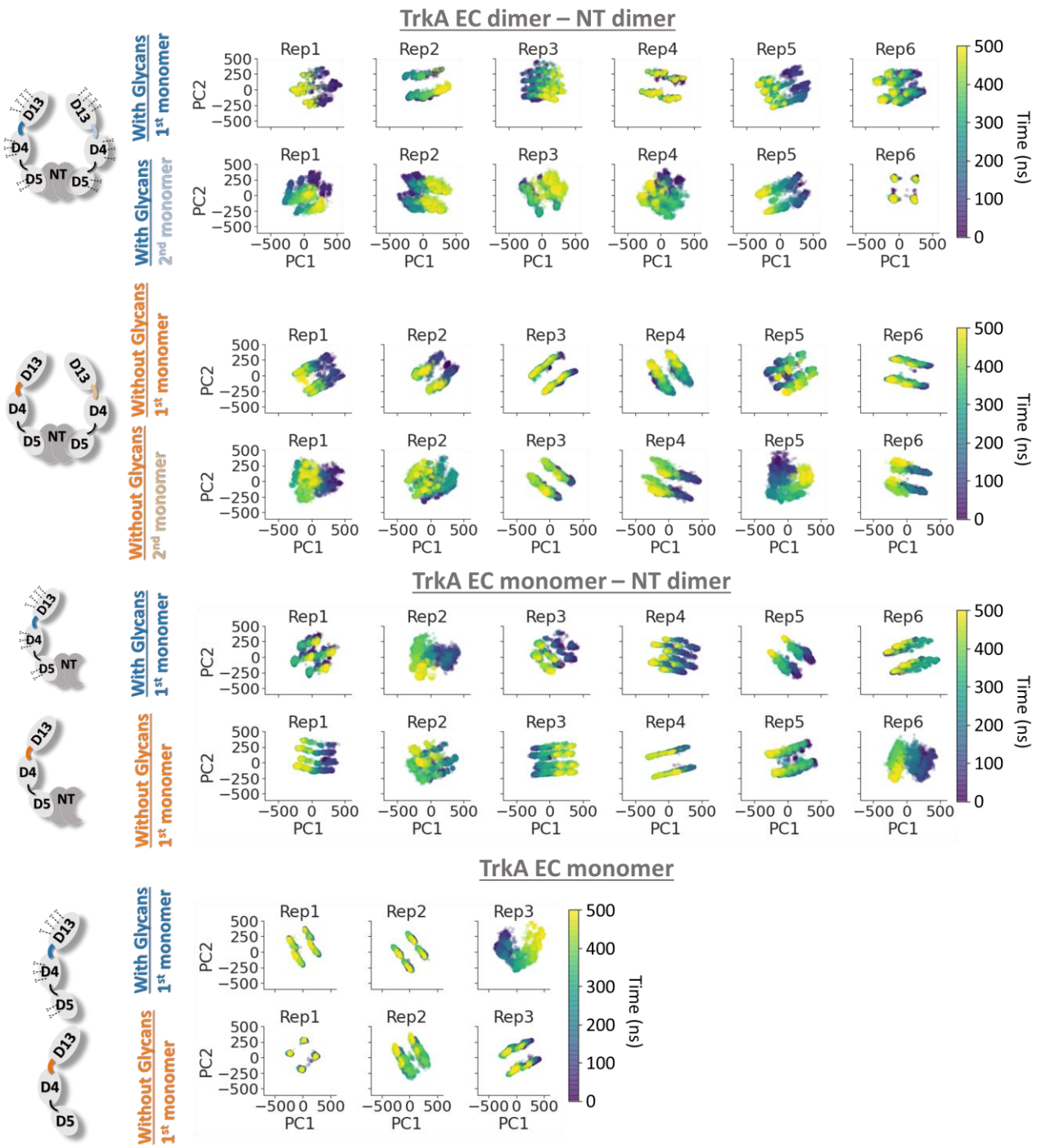


Figure A.8: Time evolution of the first and second principal components of the linker between the D13-D4 domains phi and psi dihedrals during the simulations of the glycosylated and not TrkA systems. Dihedral PCA analysis was performed to the linker after alignment of each EC segment to the D4 and D5 domains. The color coding of each domain is shown in the schematic representations of the systems at the left of the plots.

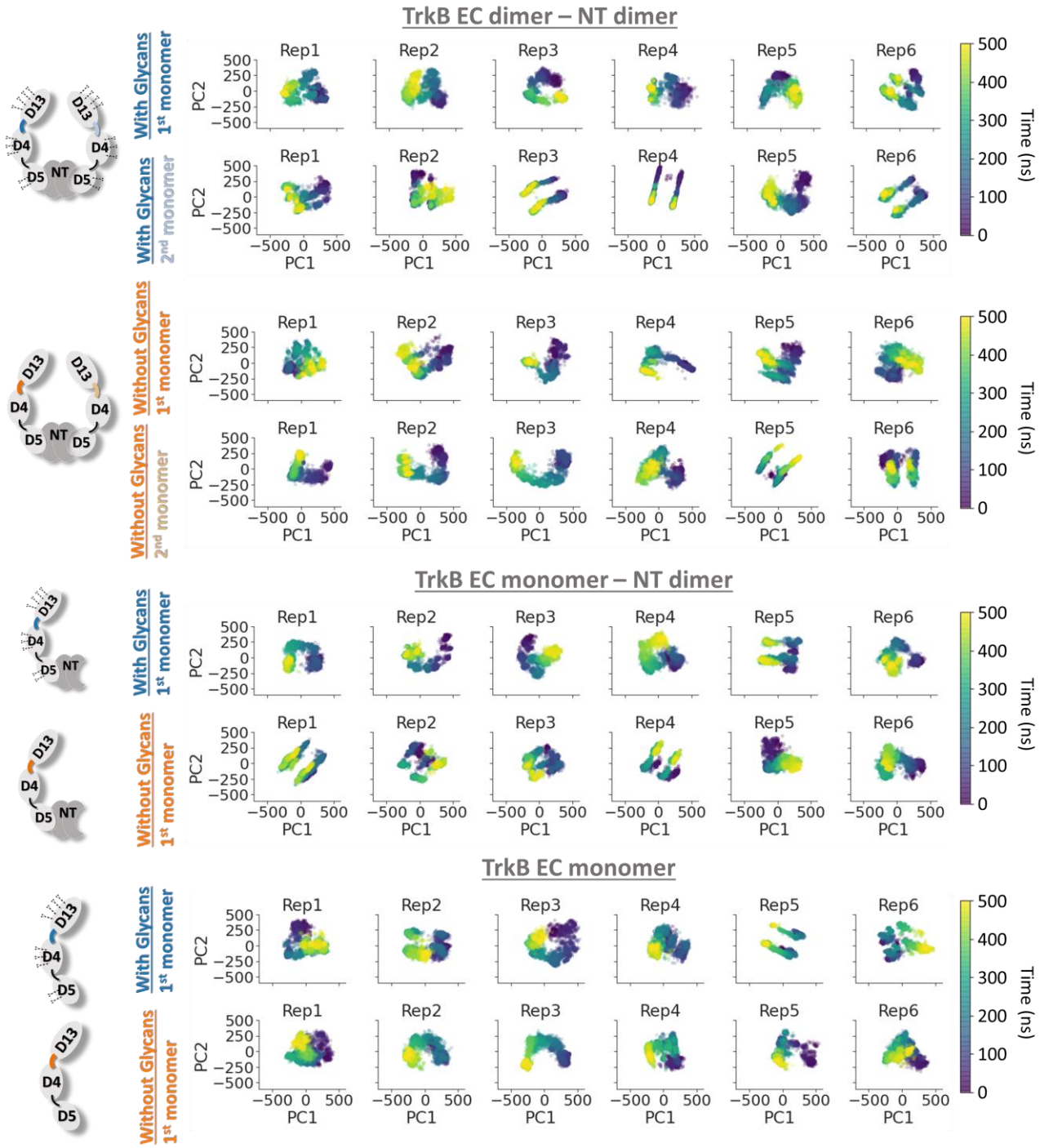


Figure A.9: Time evolution of the first and second principal components of the linker between the D13-D4 domains phi and psi dihedrals during the simulations of the glycosylated and not TrkB systems. Dihedral PCA analysis was performed to the linker after alignment of each EC segment to the D4 and D5 domains. The color coding of each domain is shown in the schematic representations of the systems at the left of the plots.

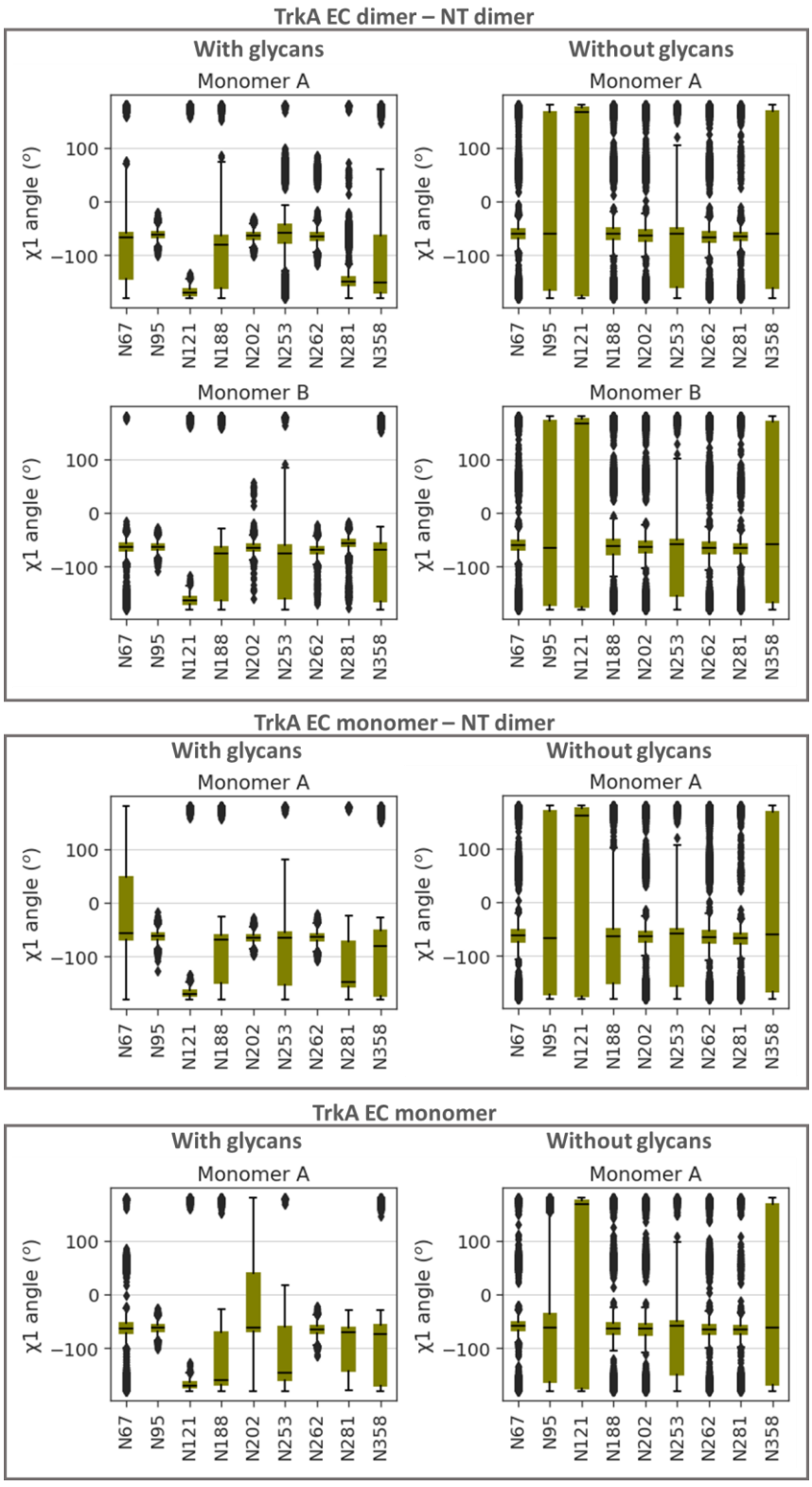


Figure A.10: Box plots of the distribution of χ1 dihedral angle of glycosylated Asn residues. A comparison of the χ1 distributions is shown between glycosylated and non-glycosylated TrkA systems.

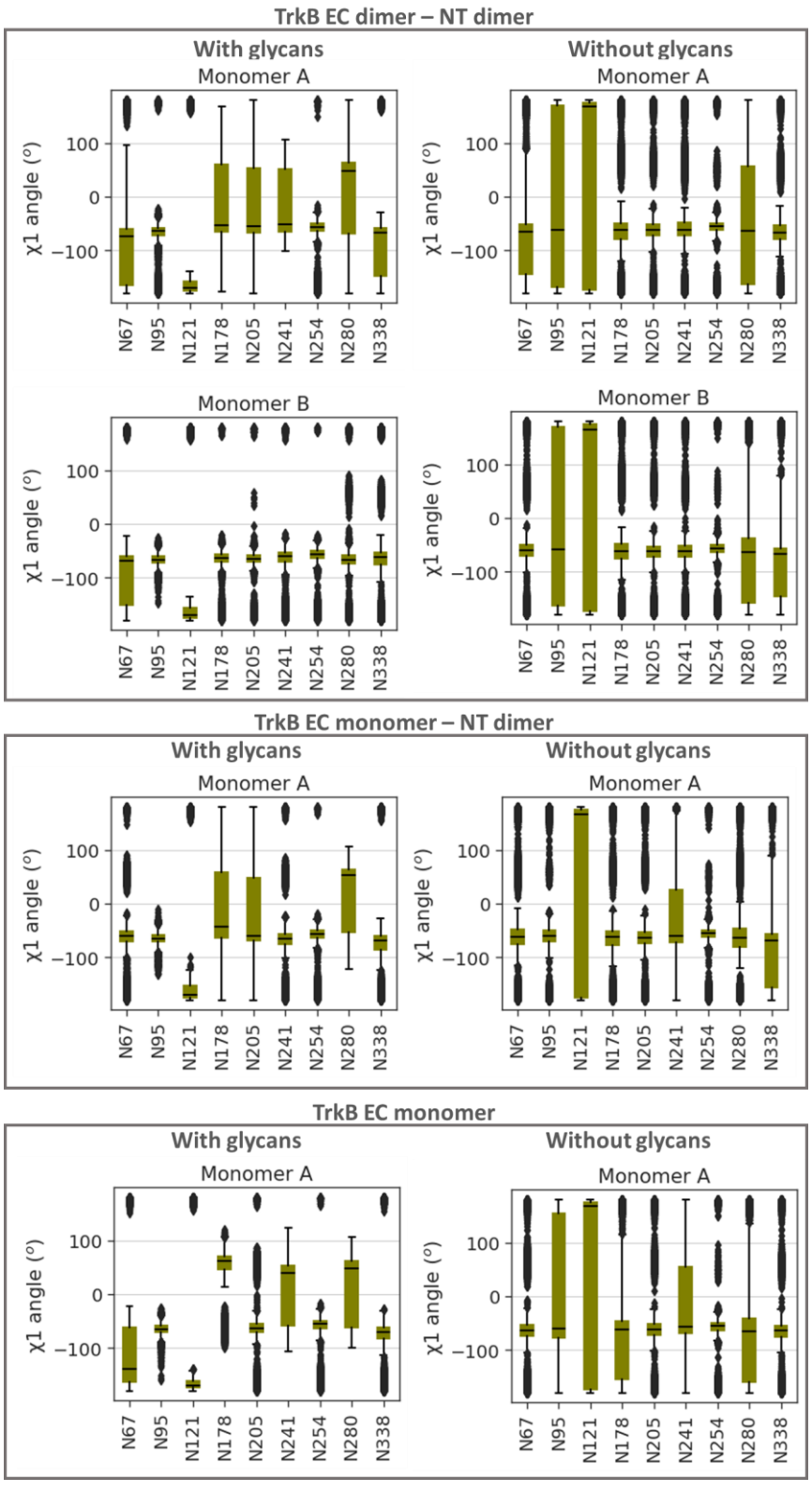


Figure A.11: Box plots of the distribution of χ1 dihedral angle of glycosylated Asn residues. A comparison of the χ1 distributions is shown between glycosylated and non-glycosylated TrkB systems.

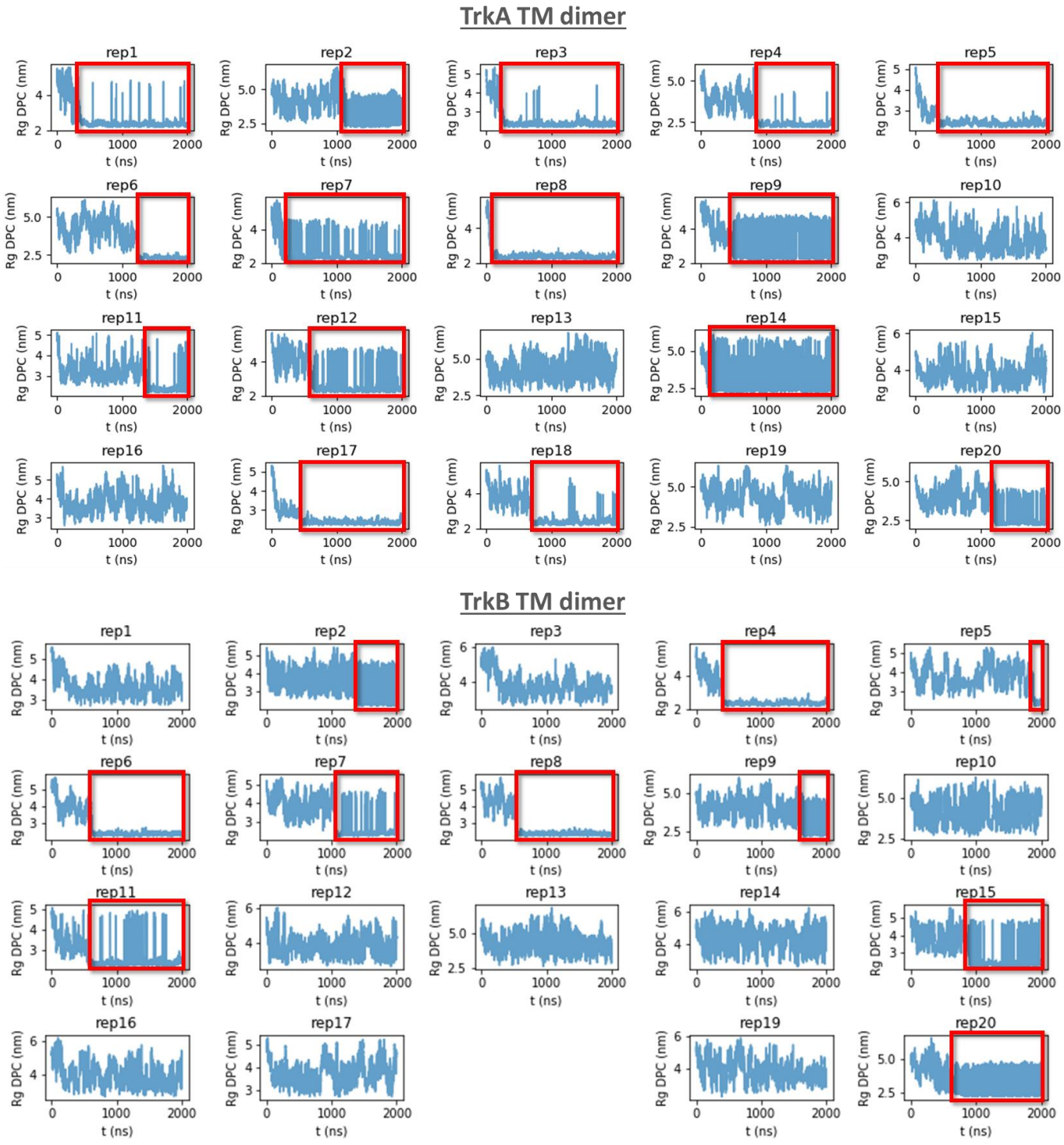


Figure A.12: Monitoring of radius of gyration (Rg) of DPC lipids during the 20 equilibration micelle self-assembly replica simulations of the TrkA and TrkB TM dimer CG systems. Low and constant values of Rg indicate formation of a micelle. Because in some simulations the formed micelle crosses the boundaries of the box, the Rg displays "jumps" as well. In these cases, a constant value of Rg was considered to indicate an equilibrated micelle. All simulation trajectories were also examined visually. The red boxes indicate the parts of the equilibration simulations where a micelle has been formed. The last frames of

replicas 1-9 and 11 from the TrkA systems and replicas 2, 4-9, 11, 15 and 20 from the TrkB systems, i.e. 10 replicas from each system, were used for production runs. Replica 18 from the TrkB set of simulations did not run properly and thus it is not included in the plots.

```
[ molecule ]
DPC
[ martini ]
NC3 PO4 C1 C2 C3
[ mapping ]
charmm36
[ atoms ]
; Terminal head group (choline)
   1     N   NC3
   2   C13   NC3
   3  H13A   NC3
   4  H13B   NC3
   5  H13C   NC3
   6   C14   NC3
   7  H14A   NC3
   8  H14B   NC3
   9  H14C   NC3
  10   C15   NC3
  11  H15A   NC3
  12  H15B   NC3
  13  H15C   NC3
  14   C12   NC3 NC3 NC3 PO4
  15  H12A   NC3 NC3 NC3 PO4
  16  H12B   NC3 NC3 NC3 PO4
  17   C11   NC3 PO4
  18  H11A   NC3 PO4
  19  H11B   NC3 PO4
; Phosphate group
  20     P   PO4
  21   O13   PO4
  22   O14   PO4
  23   O11   PO4 PO4 C1
  24   O12   PO4 PO4 PO4 NC3
; Acyl chain
  25   C31   C1 PO4
  26   H1X   C1 PO4
  27   H1Y   C1 PO4
  28   C32   C1 C1 PO4
  29   H2X   C1 C1 PO4
  30   H2Y   C1 C1 PO4
  31   C33   C1 C1 C1 PO4
  32   H3X   C1 C1 C1 PO4
```

```
  33   H3Y   C1 C1 C1 PO4
  34   C34   C1
  35   H4X   C1
  36   H4Y   C1
  37   C35   C1 C1 C2
  38   H5X   C1 C1 C2
  39   H5Y   C1 C1 C2
  40   C36   C2 C1
  41   H6X   C2 C1
  42   H6Y   C2 C1
  43   C37   C2 C2 C1
  44   H7X   C2 C2 C1
  45   H7Y   C2 C2 C1
  46   C38   C2
  47   H8X   C2
  48   H8Y   C2
  49   C39   C2 C2 C3
  50   H9X   C2 C2 C3
  51   H9Y   C2 C2 C3
  52  C310   C3 C2
  53  H10X   C3 C2
  54  H10Y   C3 C2
  55  C311   C3 C3 C2
  56  H11X   C3 C3 C2
  57  H11Y   C3 C3 C2
  58  C312   C3
  59  H12X   C3
  60  H12Y   C3
  61  H12Z   C3
;;;making a choline group
[out]
C14 N C13 C12
H14A N C13 C12
H14B N C13 C12
H14C N C13 C12
[ chiral ]
C15 N C12 C13 C14
H15A N C12 C13 C14
H15B N C12 C13 C14
H15C N C12 C13 C14
```

Figure A.13: Map between coarse-grained beads and atoms for a DPC lipid molecule. This was used in the backward.py script[301] for conversion of the CG DPC micelle to atomistic representation.

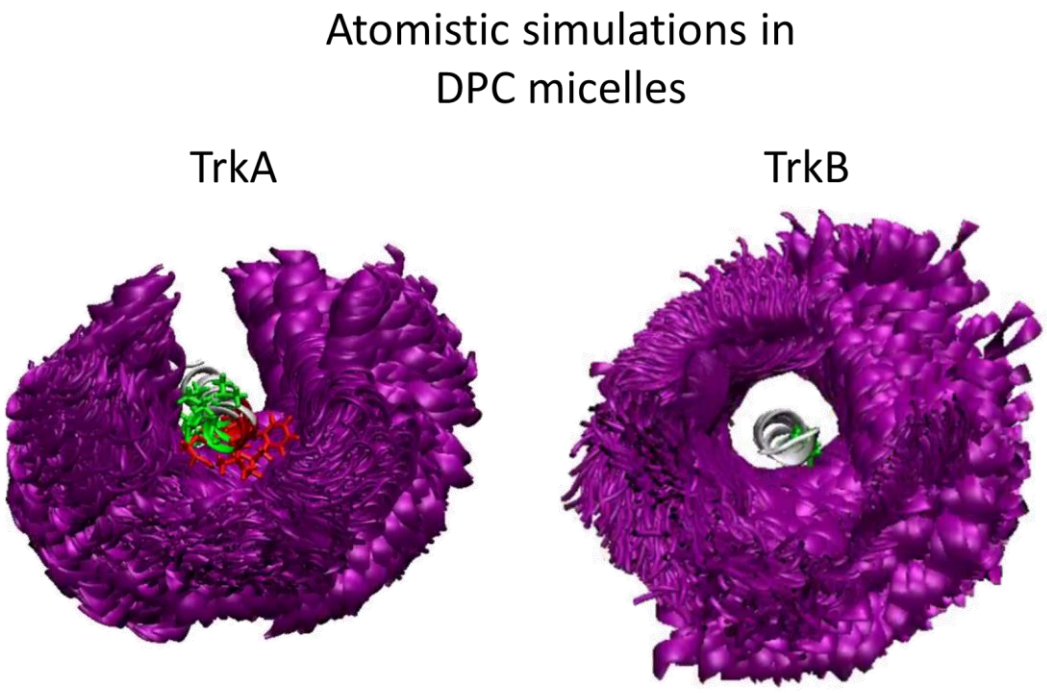


Figure A.14: TrkA and TrkB configurations of TM dimers sampled during the atomistic simulations in DPC micelles. I have performed alignment of all simulation frames to the first helix, shown here in white cartoon representation. The second helix of the dimer is shown in purple cartoon representation in all frames of the simulations. The residues of TrkA that have been indicated as active or inactive from mutagenesis studies[60] are shown in green and red stick representation, respectively. The AxxxG motif of TrkB, which has been suggested to be part of the active interface[91] is also shown in green stick representation.

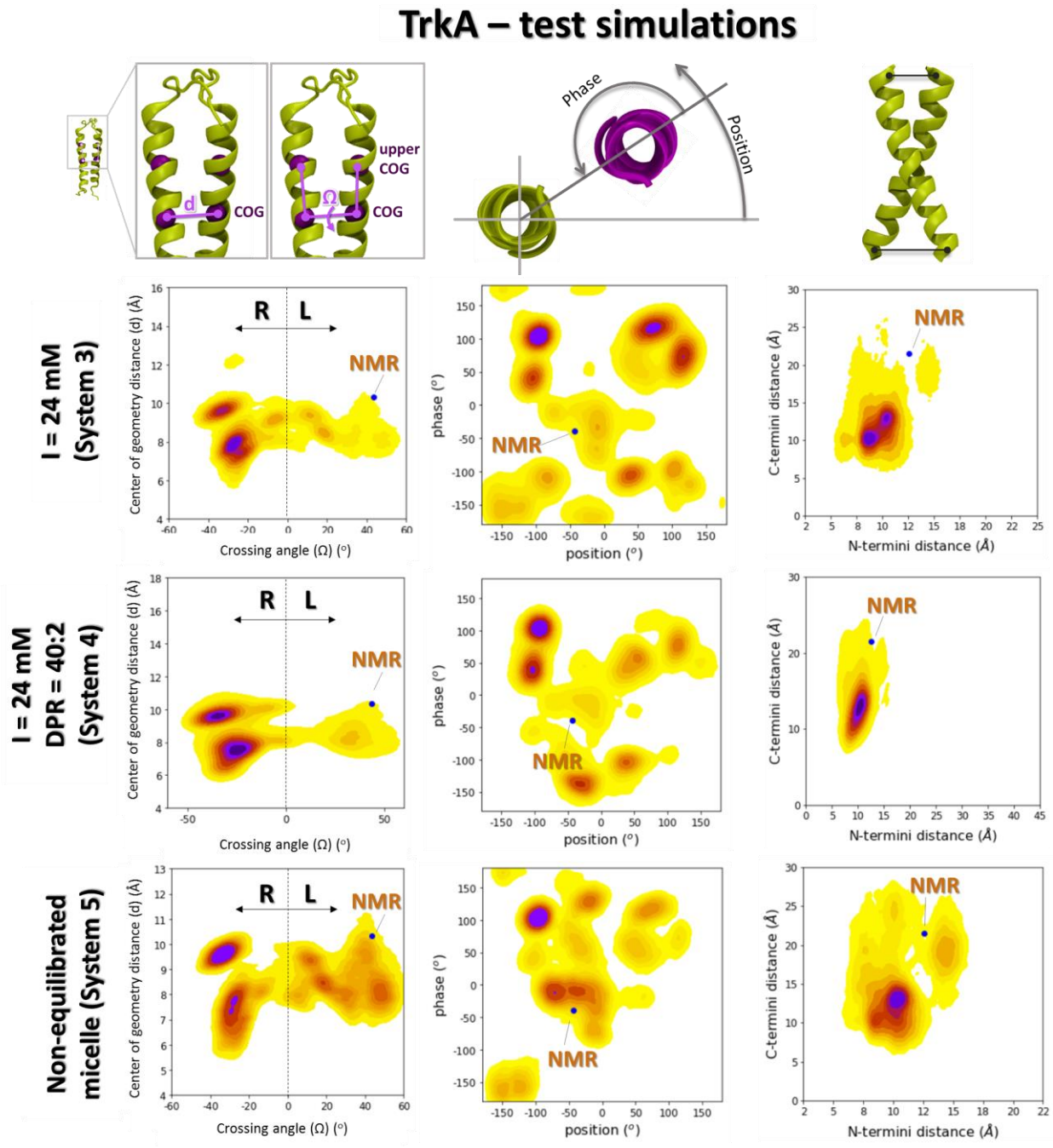


Figure A.15: Comparison of sampling in different tests of CG simulations for the TrkA TM homodimer. I calculated the KDE plots for six different order parameters: interhelical distance (d), interhelical crossing angle (Ω), phase (rotation around 1st helix), position (rotation around 2nd helix), N- and C- termini distances. The parameters are also visualized schematically above the plots. Right- and left-handed arrangements are indicated with (R and L). The NMR state is shown with a blue dot. The "I" stands for ionic strength and "DPR" for detergent-to-protein molar ratio.

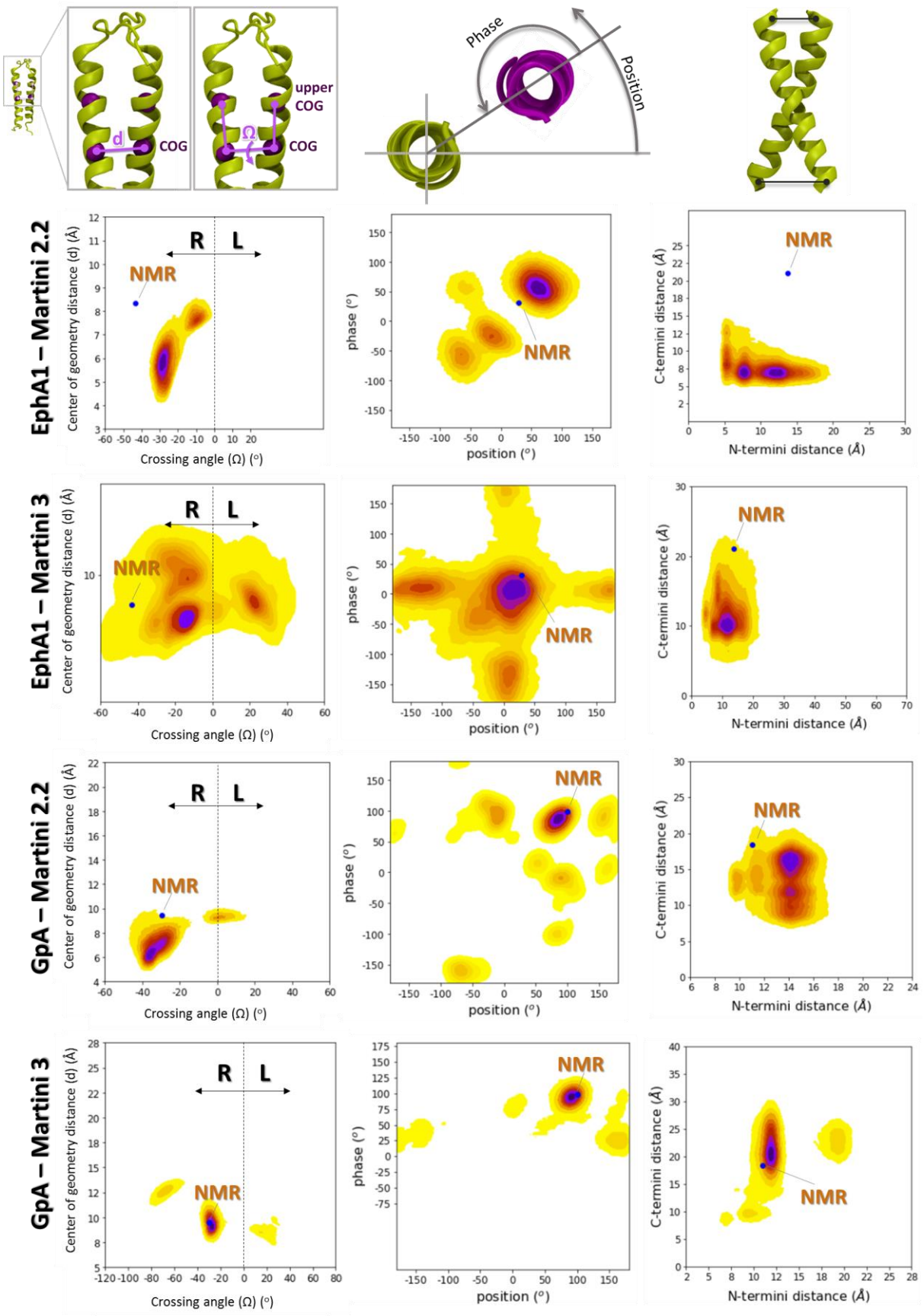


Figure A.16: Validation of CG simulation protocol in EphA1 and GpA TM dimers with Martini 2.2 and 3.

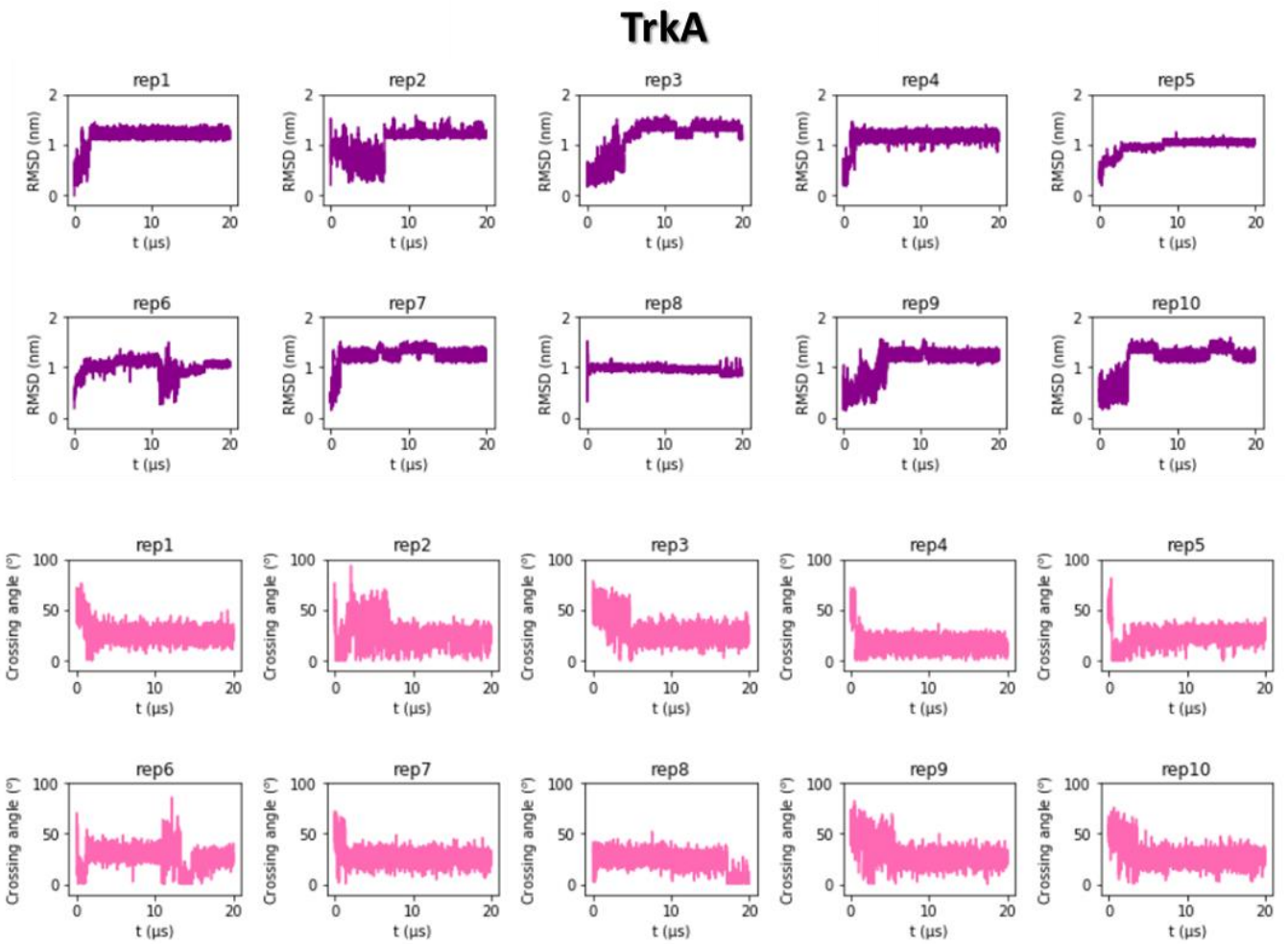


Figure A.17: Root mean squared deviation (RMSD) and absolute crossing angle calculations during the 10 CG replica simulations of TrkA TM dimer.

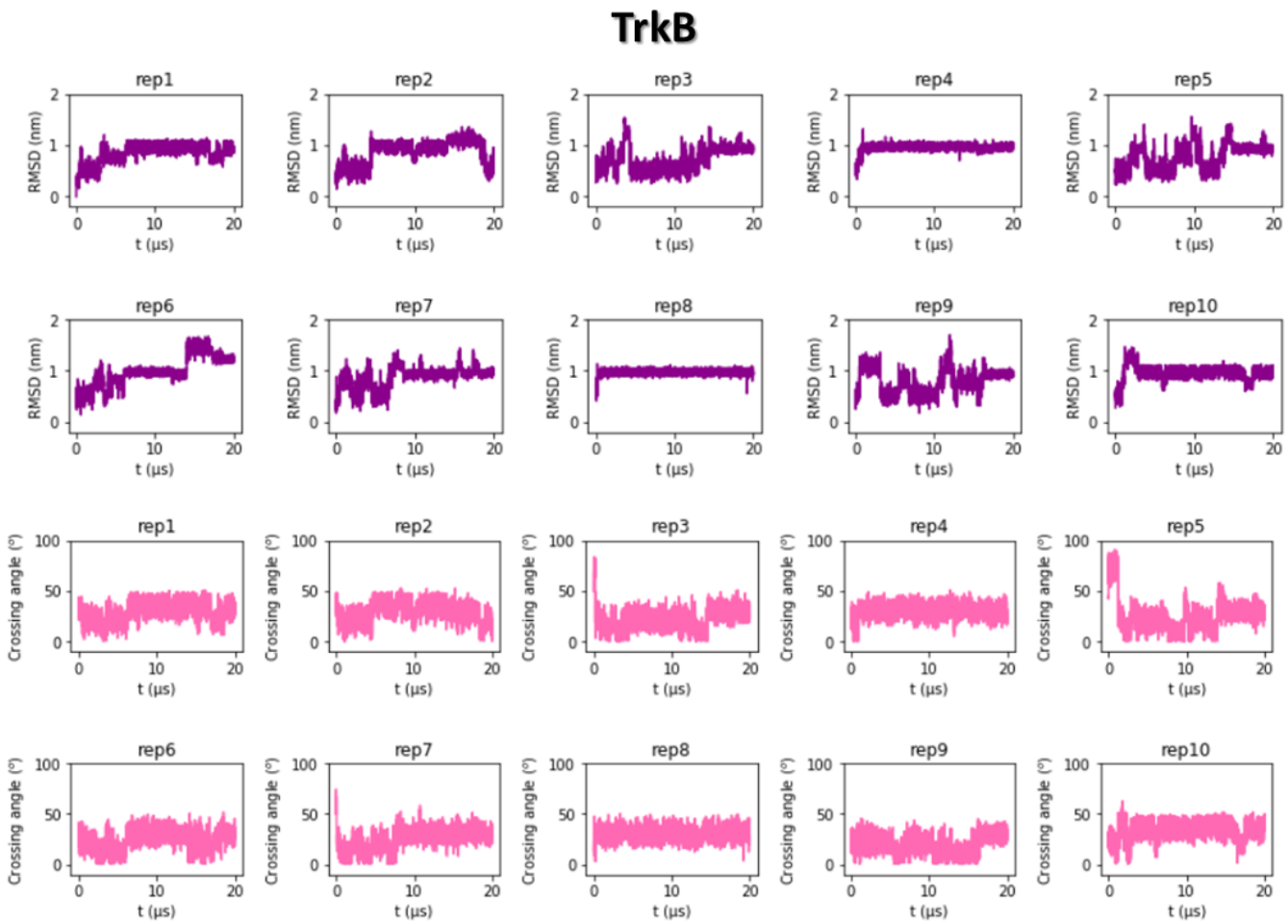
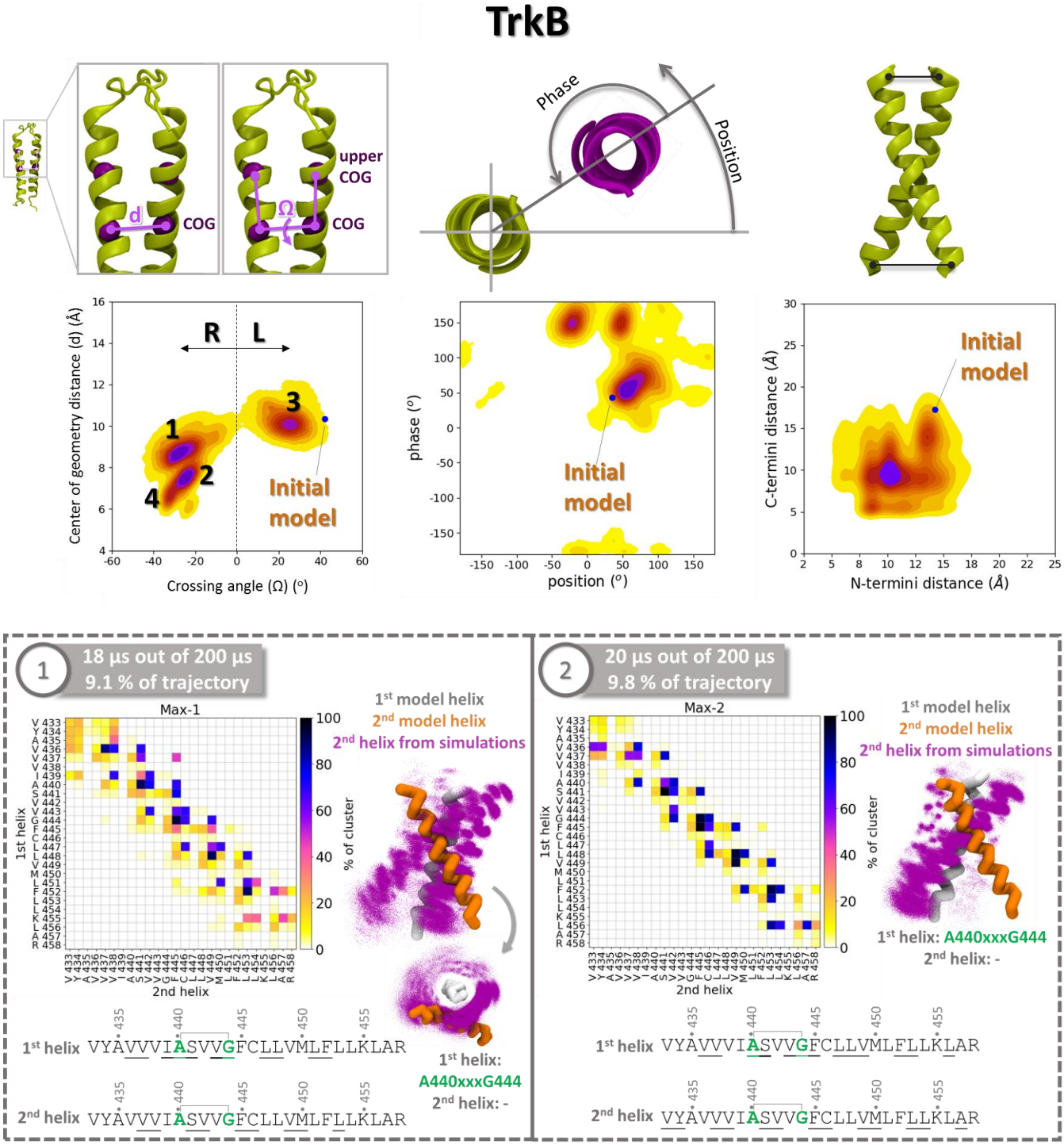


Figure A.18: Root mean squared deviation (RMSD) and absolute crossing angle calculations during the 10 CG replica simulations of TrkB TM dimer.


TrkB
d
COG
upper
COG
Ω
COG
Phase
Position
Center of geometry distance (d) (Å)
Crossing angle (Ω) (°)
R
L
1
2
3
4
Initial
model
phase (°)
position (°)
Initial
model
C-termini distance (Å)
N-termini distance (Å)
Initial
model
1
18 µs out of 200 µs
9.1 % of trajectory
Max-1
1st model helix
2nd model helix
2nd helix from simulations
1st helix
2nd helix
% of cluster
1st helix VYAVVVIASVVGFCLLVMLFLLKLAR
2nd helix VYAVVVIASVVGFCLLVMLFLLKLAR
1st helix:
A440xxxG444
2nd helix: -
2
20 µs out of 200 µs
9.8 % of trajectory
Max-2
1st model helix
2nd model helix
2nd helix from simulations
1st helix
2nd helix
% of cluster
1st helix: A440xxxG444
2nd helix: -
1st helix VYAVVVIASVVGFCLLVMLFLLKLAR
2nd helix VYAVVVIASVVGFCLLVMLFLLKLAR

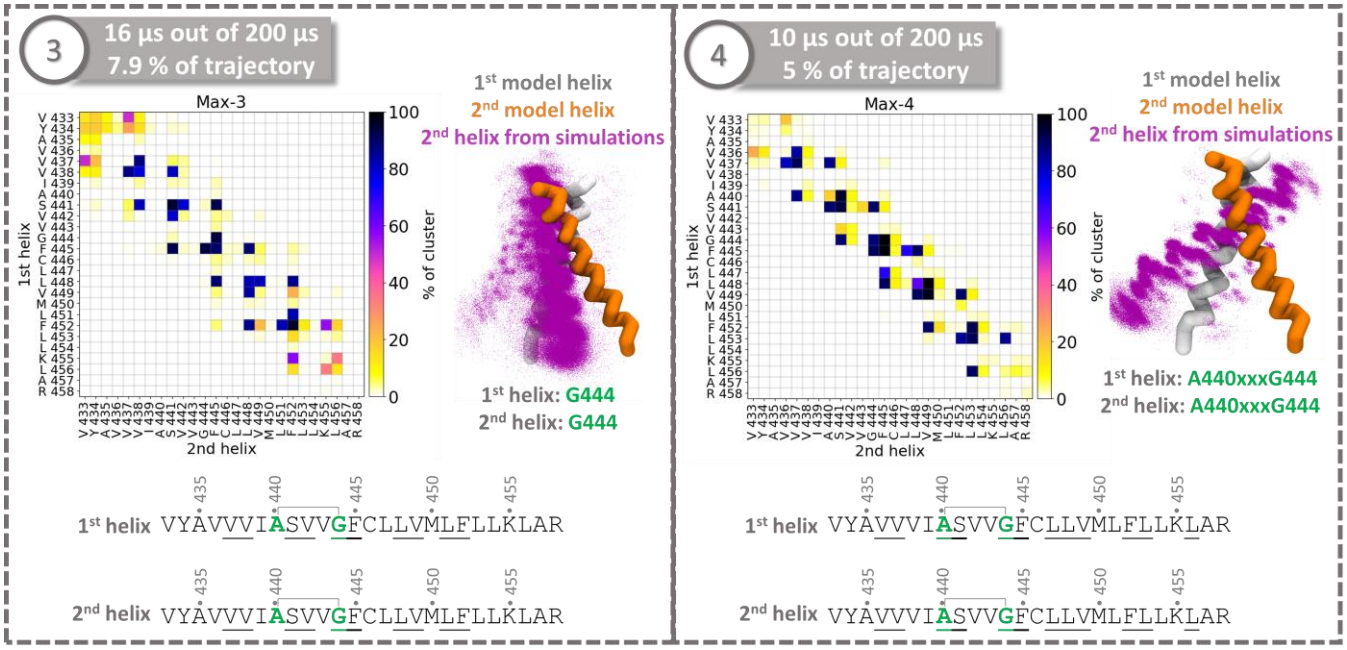


Figure A.19: Analysis of four TrkB arrangements from the CG simulations. The maxima are numbered in the d-Ω 2D heatmap and then analyzed in each subsequent box. In each box, I have calculated the percentage of the trajectory that was spent in this maximum, the contact map between the two helices and the residues that interact the most from each helix are underlined in the two sequences on the bottom. I have also visualized the arrangements by aligning all the frames to the 1st helix and then displaying the backbone of the 1st helix in white, the 2nd helix from the NMR structure of TrkA in orange, and the 2nd helix from the simulation frames in purple dots.

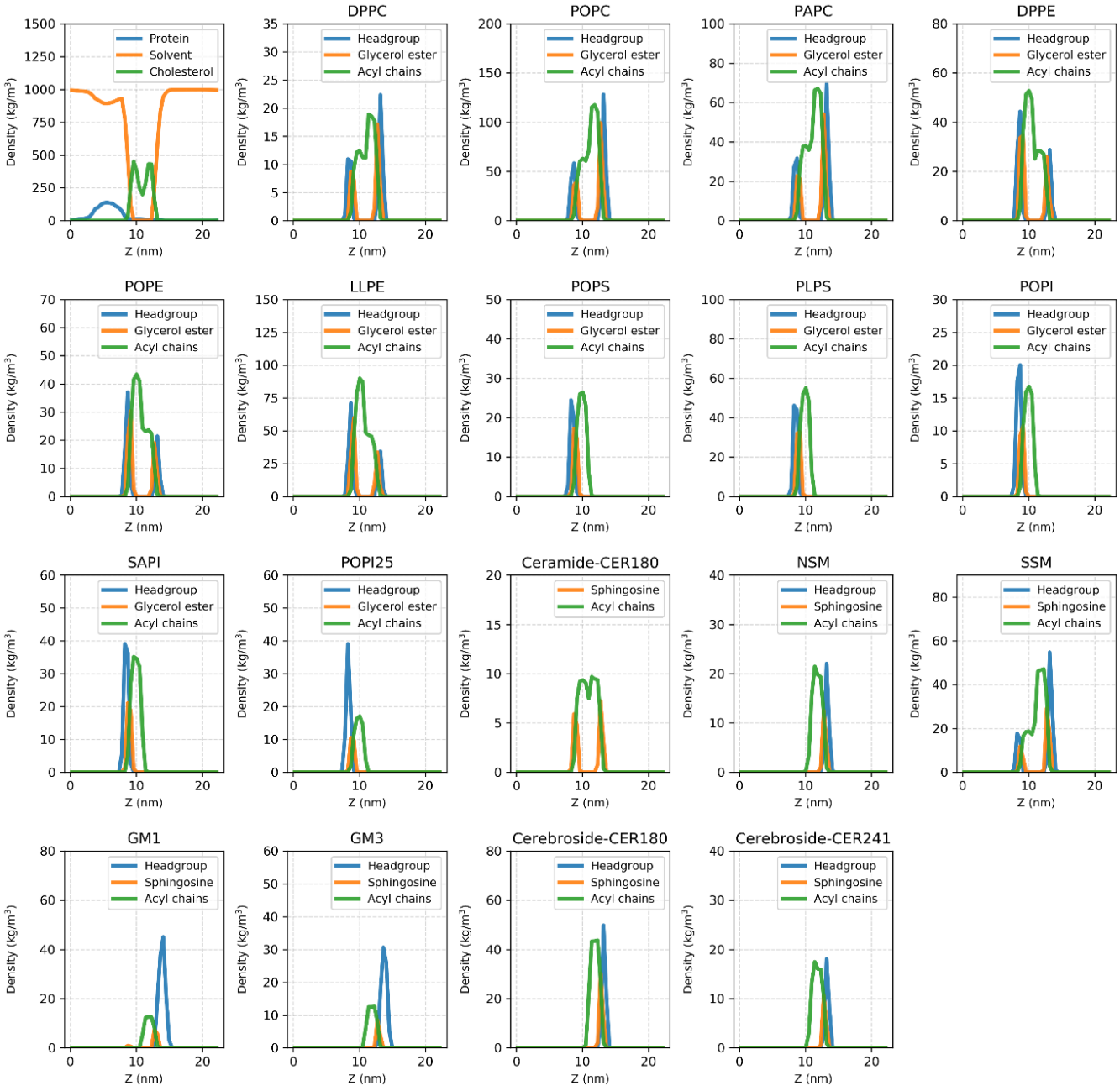


Figure A.20: Density profile along the z-axis (membrane normal) calculated for each lipid moiety (headgroup, glycerol ester or sphingosine backbone, acyl chains), as well as the protein, the solvent, and the cholesterol. The figure shows this analysis for one of the TM-IC replica simulations as an example.

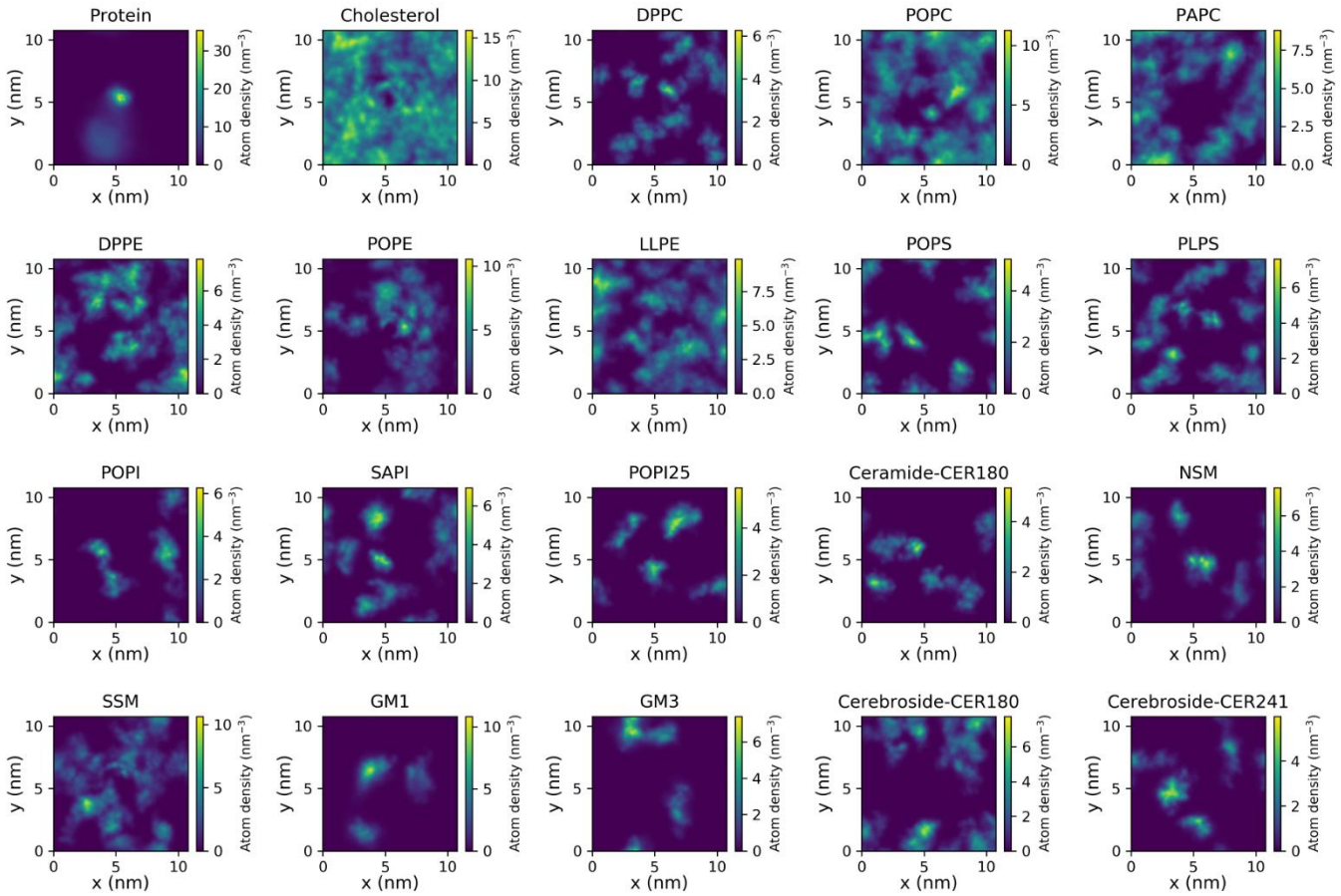


Figure A.21: Atom density in the membrane plane averaged over the whole extent of the membrane in the z-axis direction.

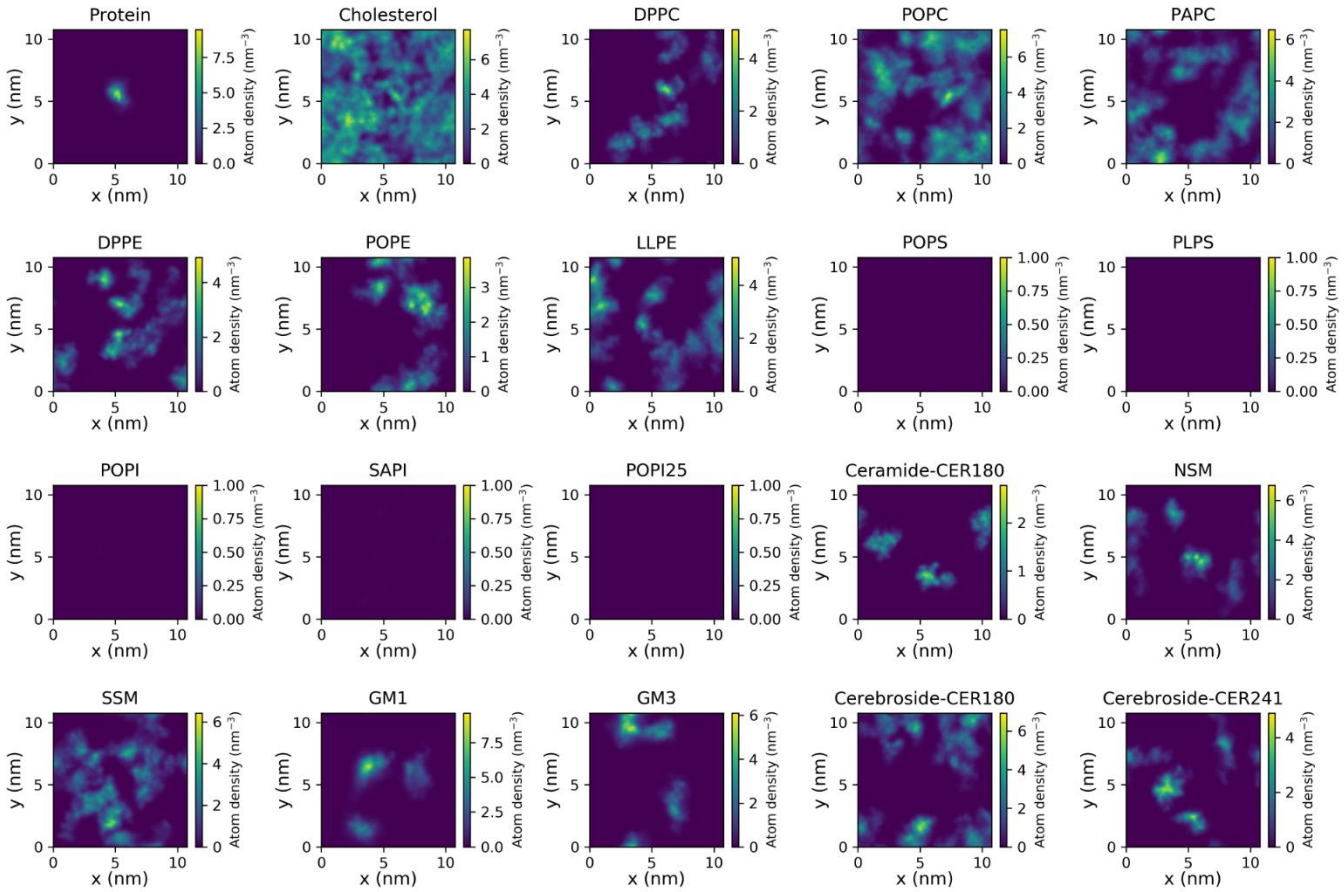


Figure A.22: Atom density in the membrane plane averaged over the outer leaflet of the membrane.

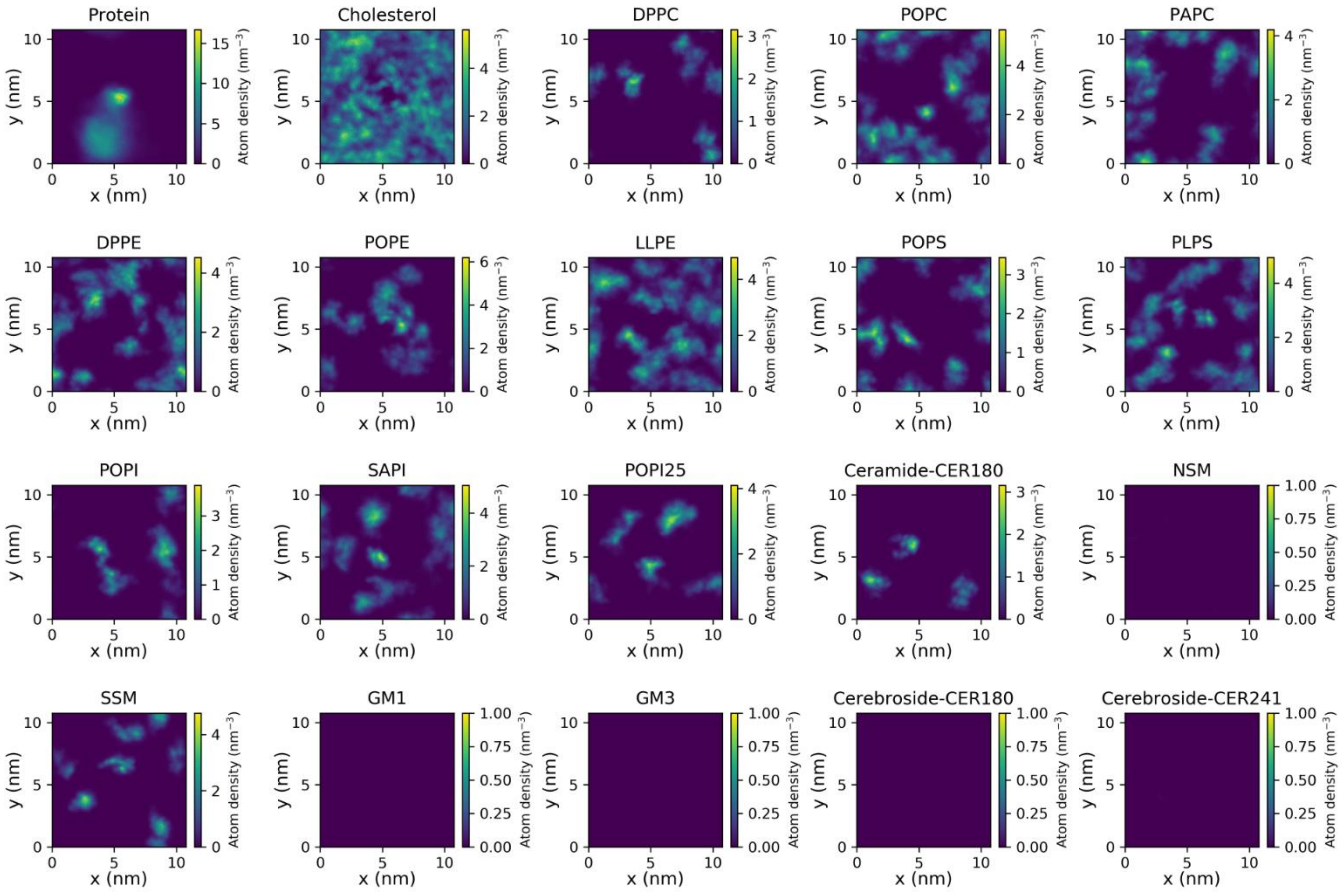


Figure A.23: Atom density in the membrane plane averaged over the inner leaflet of the membrane.

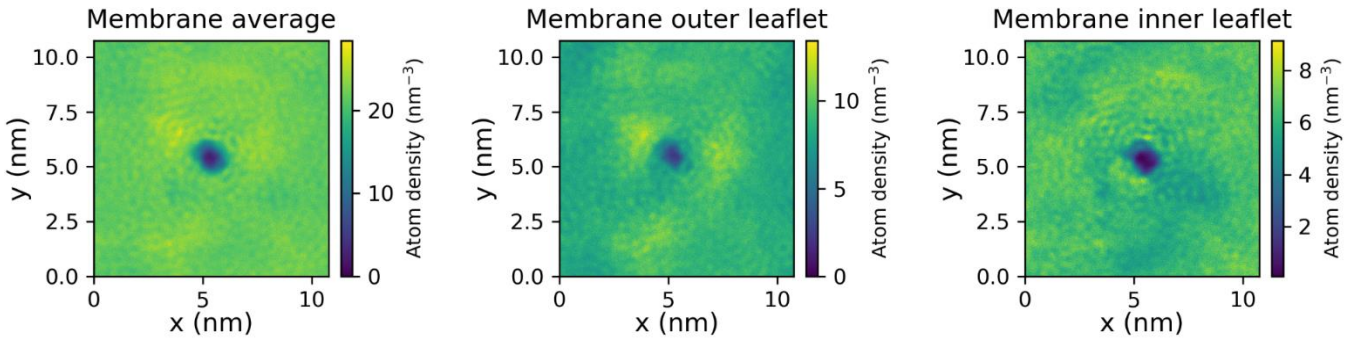


Figure A.24: Atom density of all lipids in the membrane.

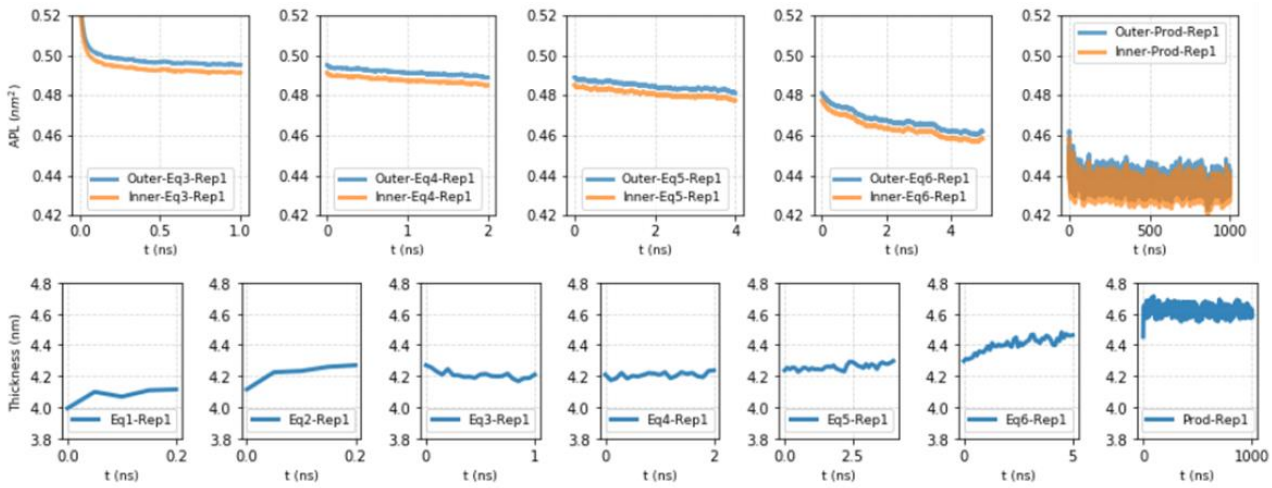


Figure A.25: Average area per lipid (APL) of the outer and inner leaflets and membrane thickness of the neuronal membrane for the different equilibration and production steps.

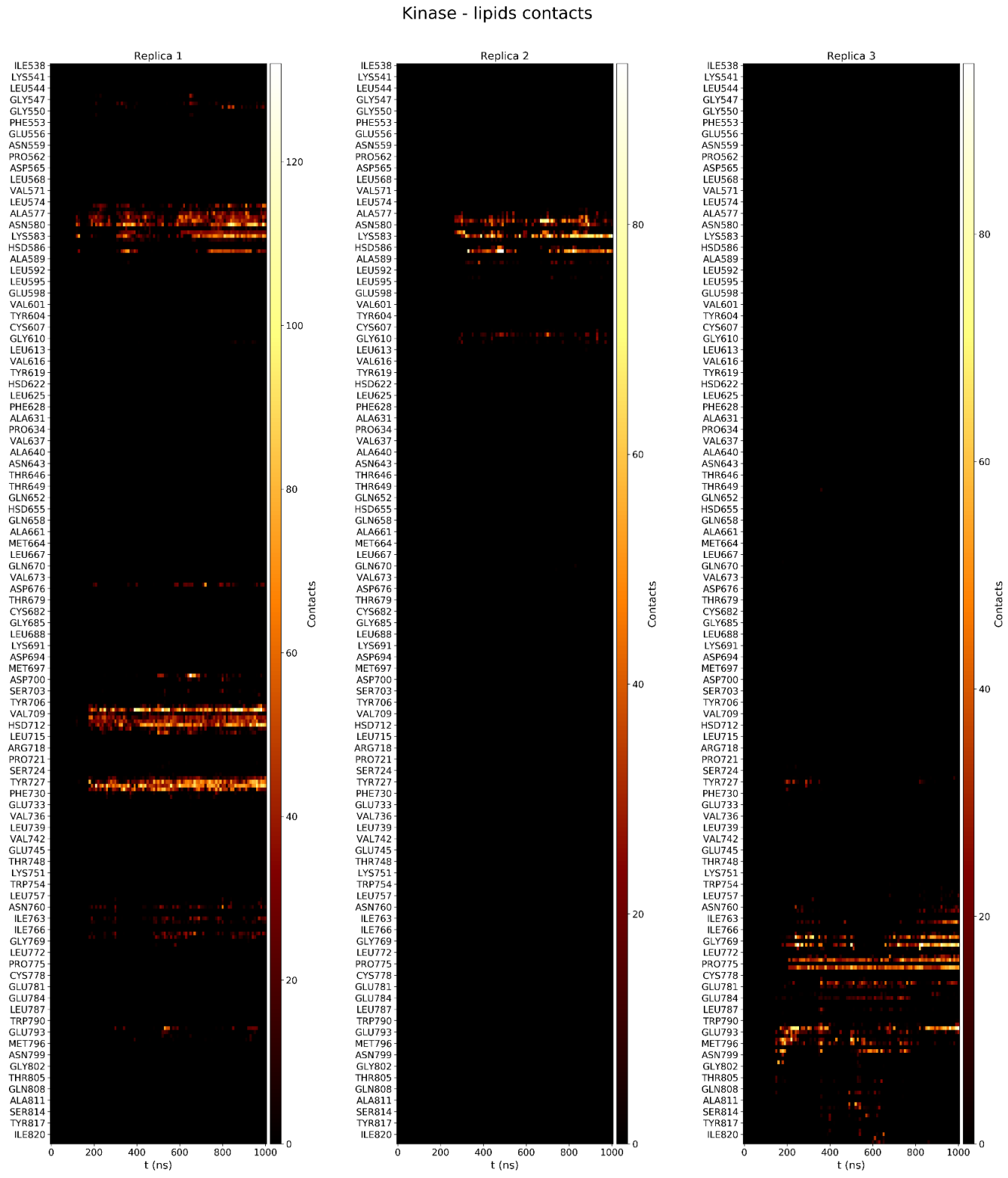


Figure A.26: Number of contacts between the TrkB kinase domain and the lipids of the neuronal membrane as function of time for the three replicas of the TM-IC system. The exact numbers of contacts are shown.

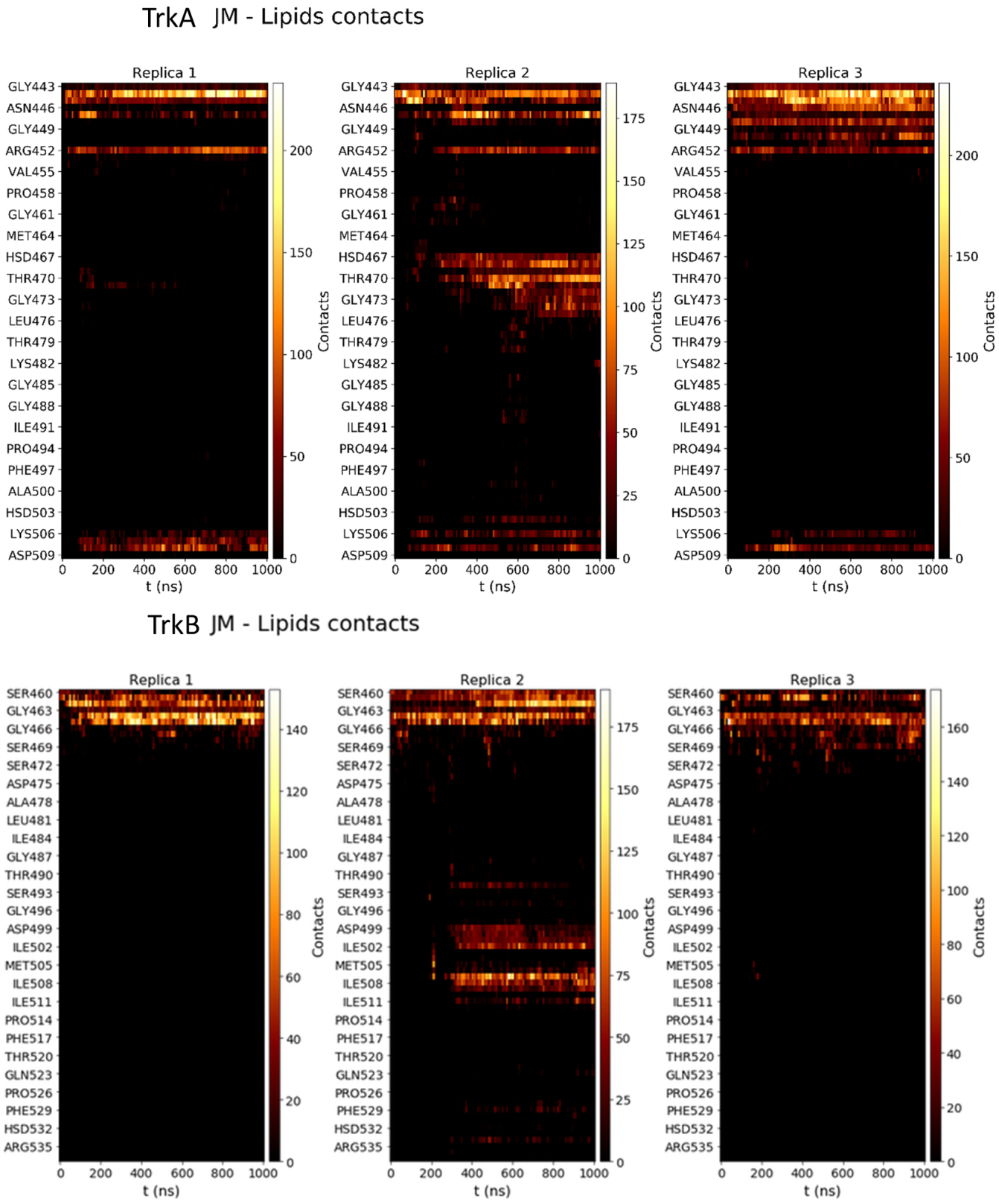
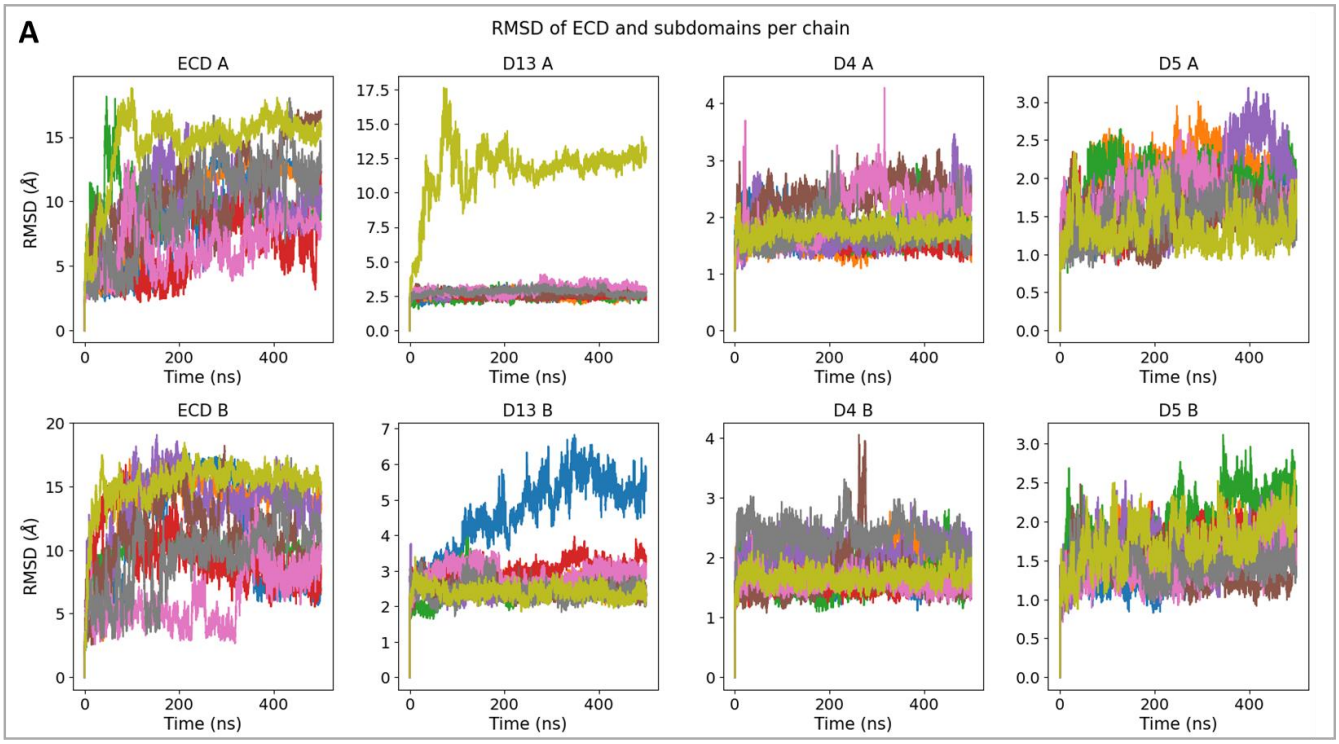
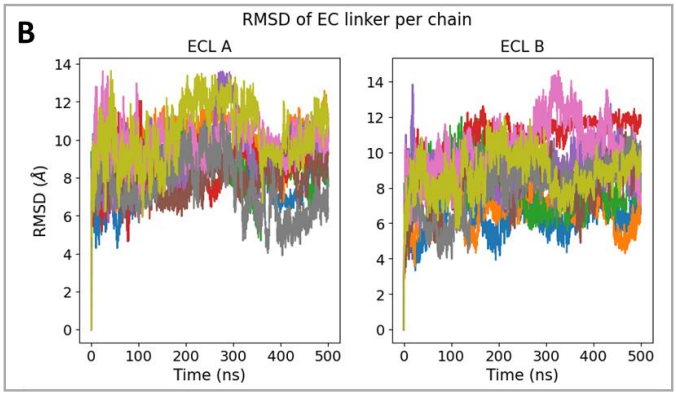
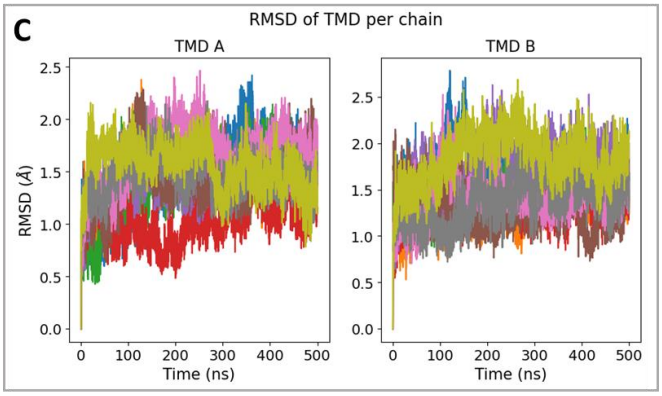
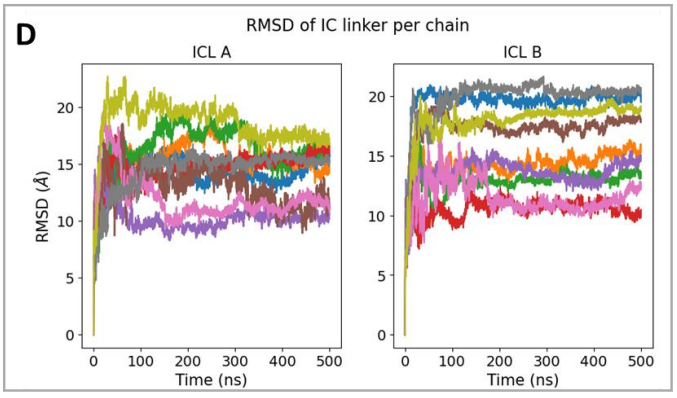


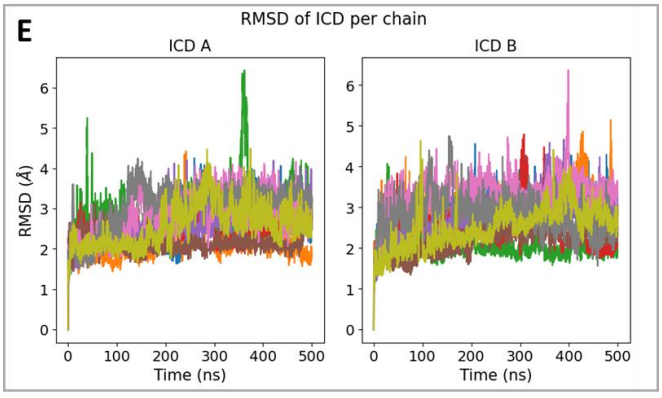
Figure A.27: Number of contacts between the JM domain of TrkA and TrkB receptors and the membrane. The heatmaps show the number of contacts over simulation time for each replica.


A
RMSD of ECD and subdomains per chain
ECD A
D13 A
D4 A
D5 A
ECD B
D13 B
D4 B
D5 B
RMSD (Å)
Time (ns)


B
RMSD of EC linker per chain
ECL A
ECL B
RMSD (Å)
Time (ns)


C
RMSD of TMD per chain
TMD A
TMD B
RMSD (Å)
Time (ns)


D
RMSD of IC linker per chain
ICL A
ICL B
RMSD (Å)
Time (ns)


E
RMSD of ICD per chain
ICD A
ICD B
RMSD (Å)
Time (ns)

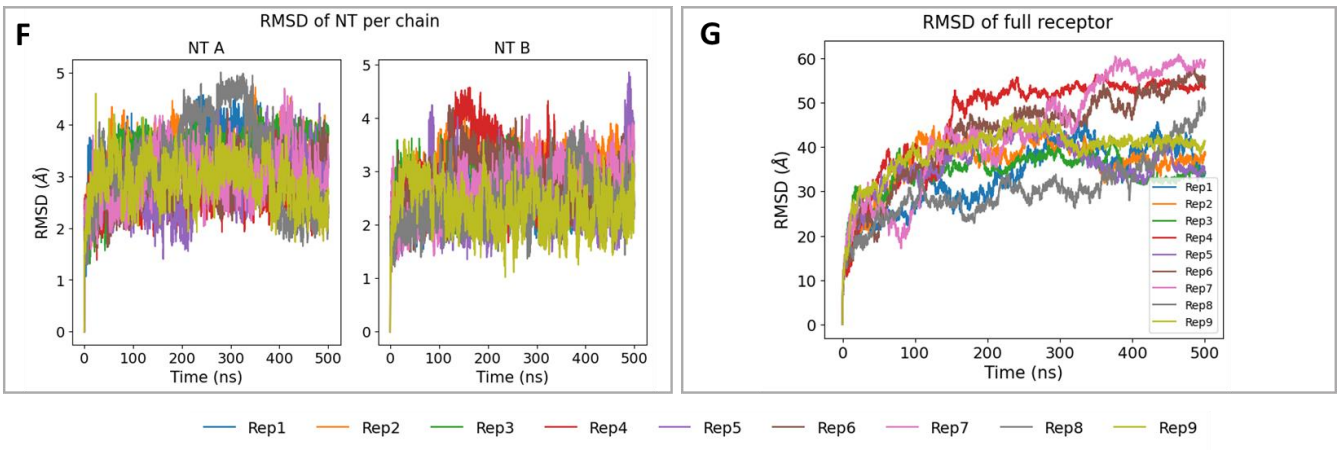
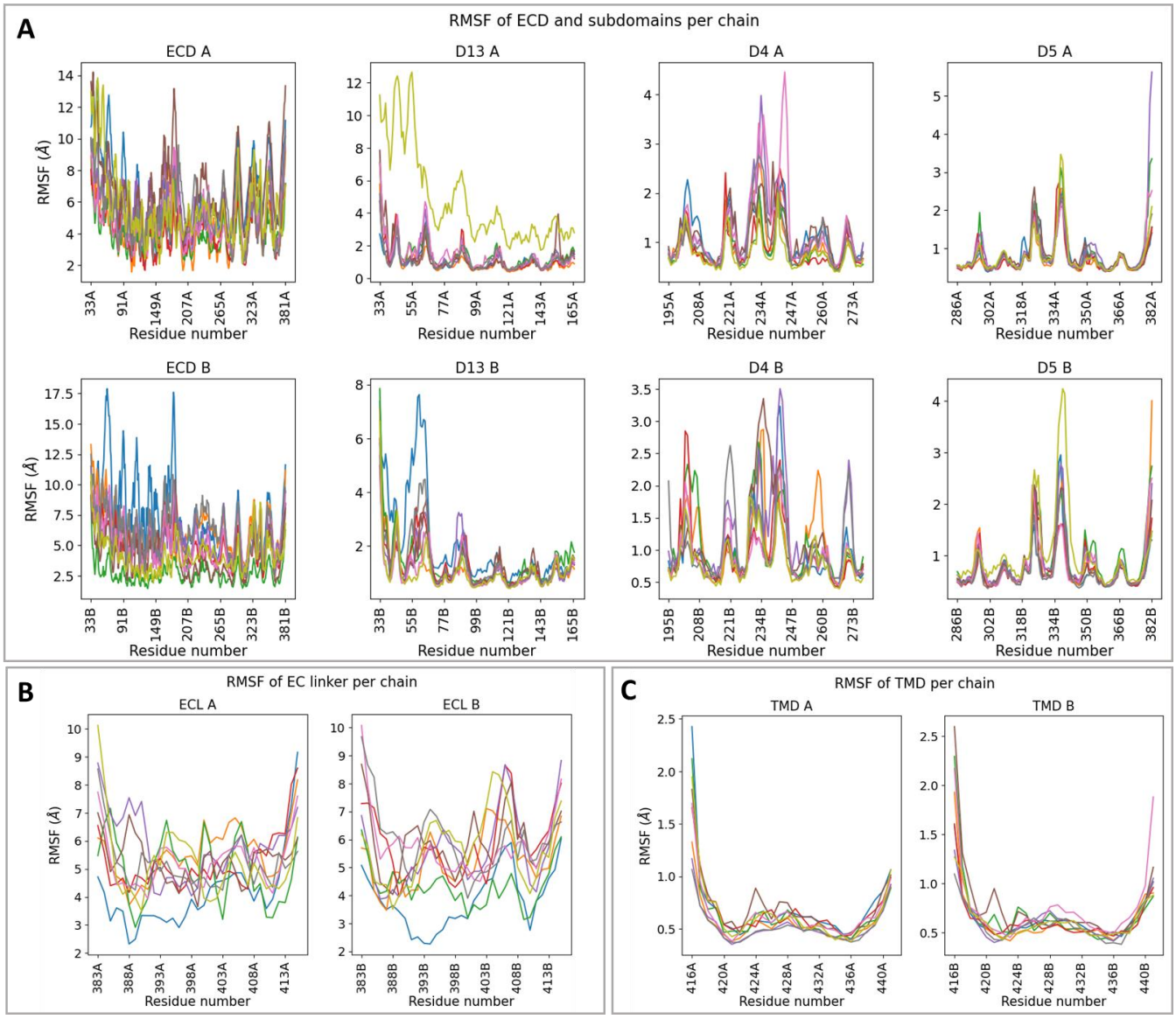


Figure A.28: RMSD values computed for the different domains from the full-length TrkA simulations.



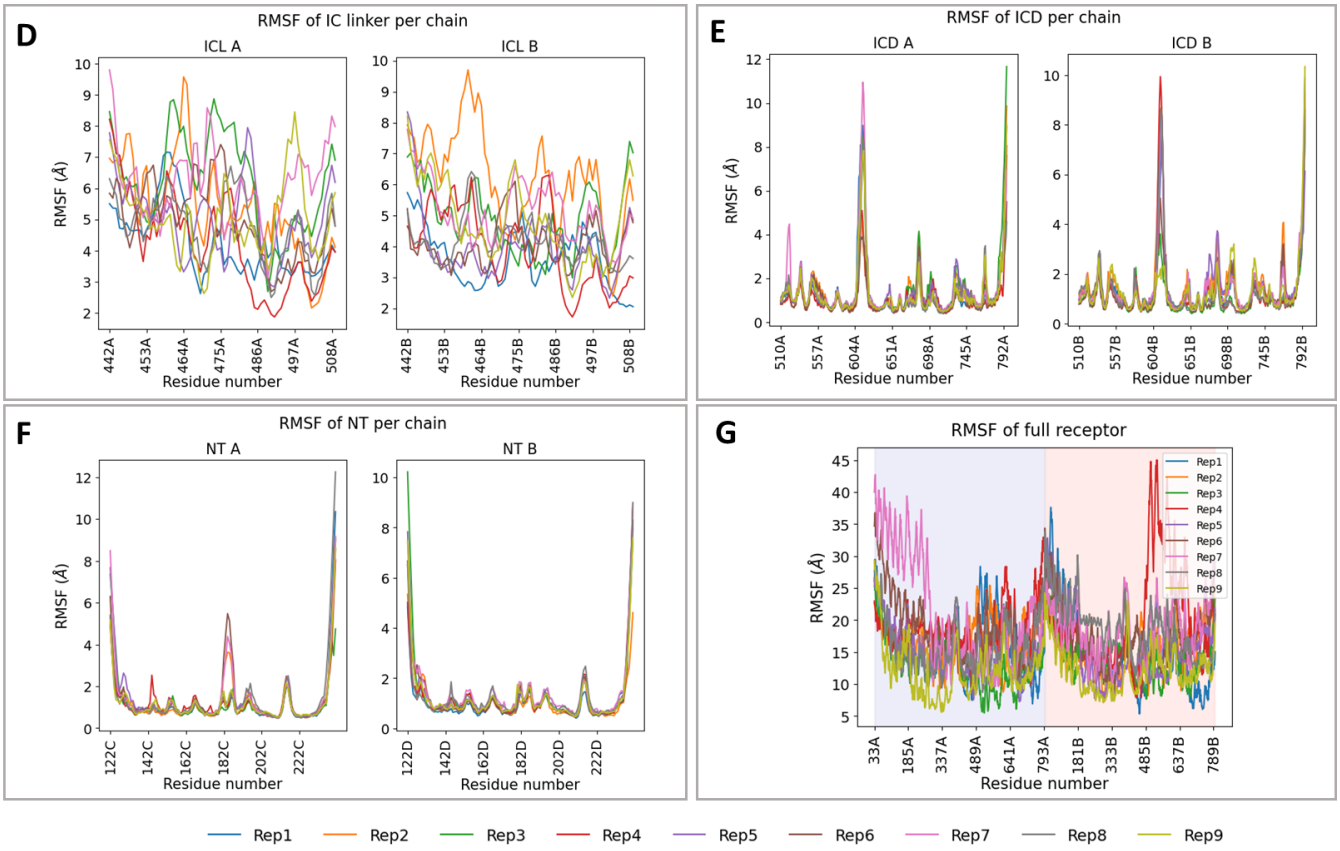
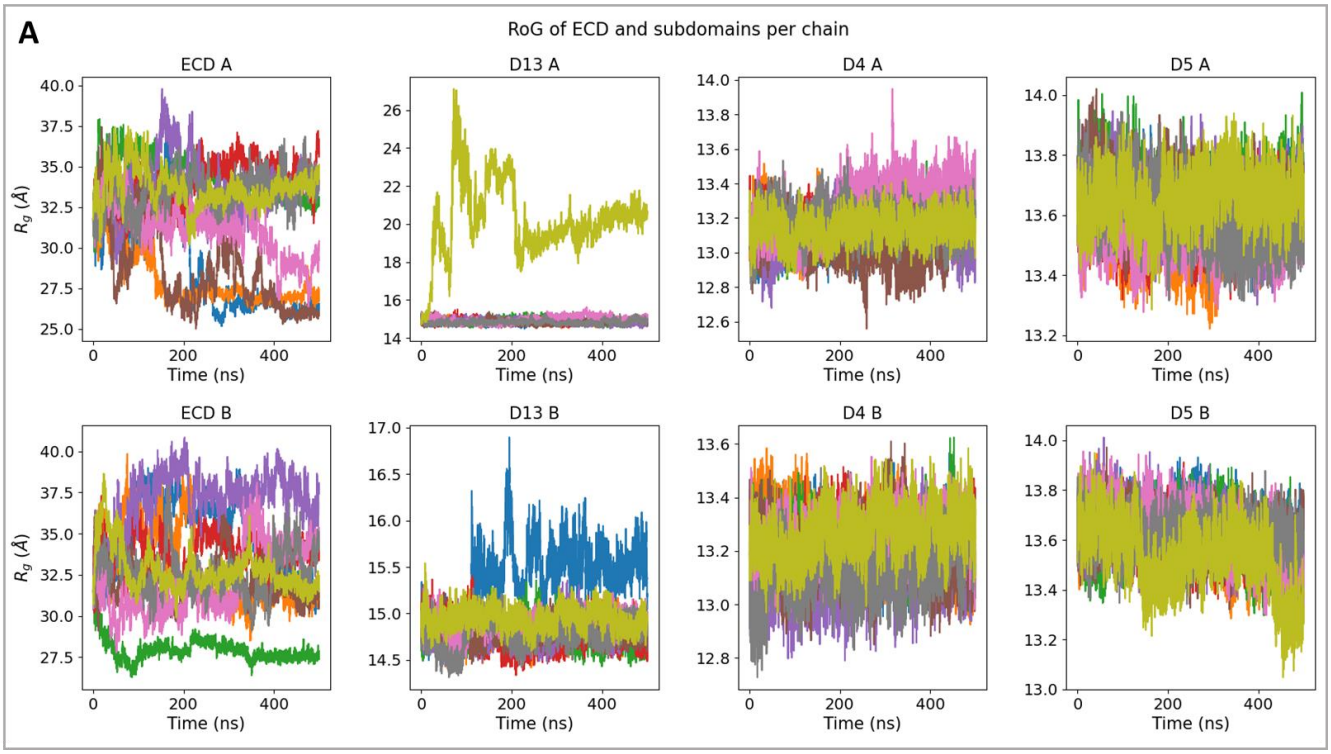


Figure A.29: RMSF calculations of the different domains from the full-length TrkA simulations.



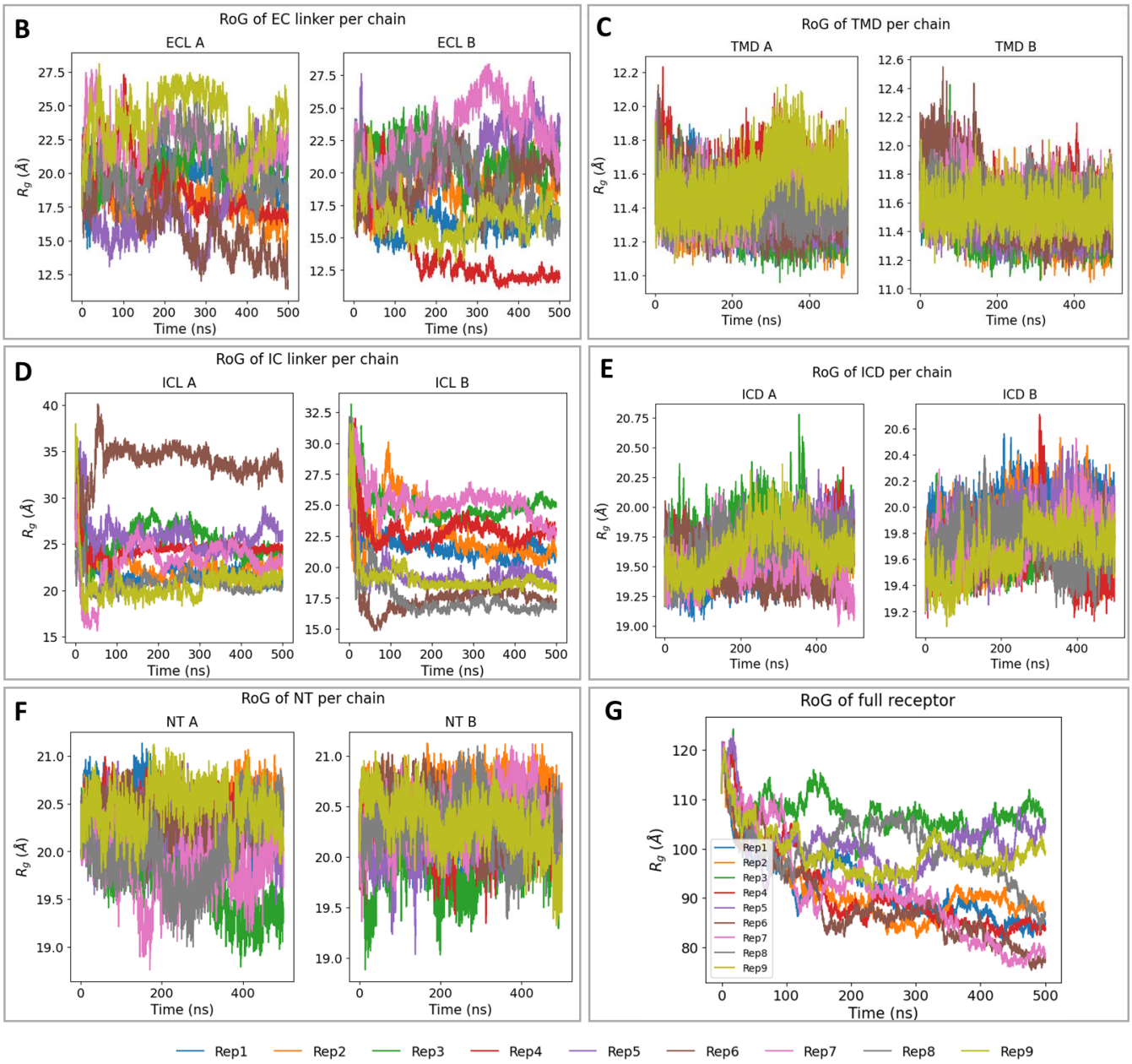


Figure A.30: Radius of gyration (R_g) calculations of the different domains from the full-length TrkA simulations.


A
RMSD of ECD and subdomains per chain
ECD A
D13 A
D4 A
D5 A
ECD B
D13 B
D4 B
D5 B
RMSD (Å)
Time (ns)
B
RMSD of EC linker per chain
ECL A
ECL B
C
RMSD of TMD per chain
TMD A
TMD B
D
RMSD of IC linker per chain
ICL A
ICL B
E
RMSD of ICD per chain
ICD A
ICD B

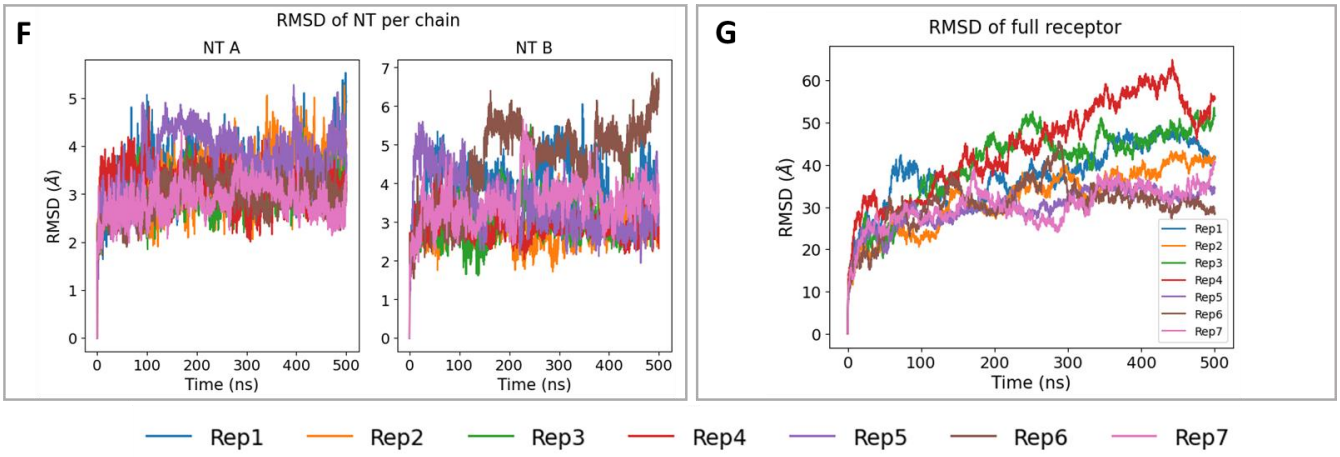
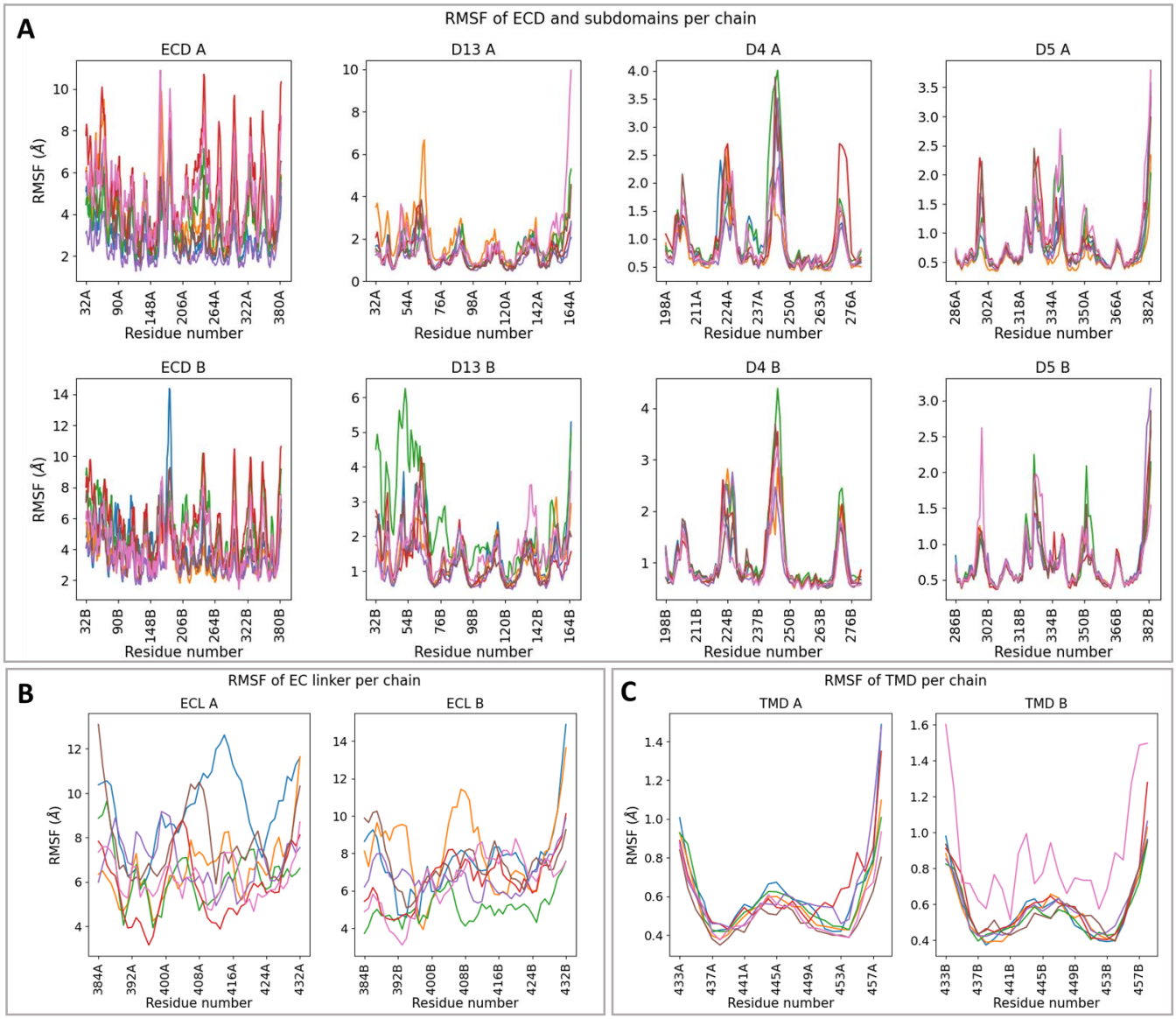


Figure A.31: RMSD calculations of the different domains from the full-length TrkB simulations.



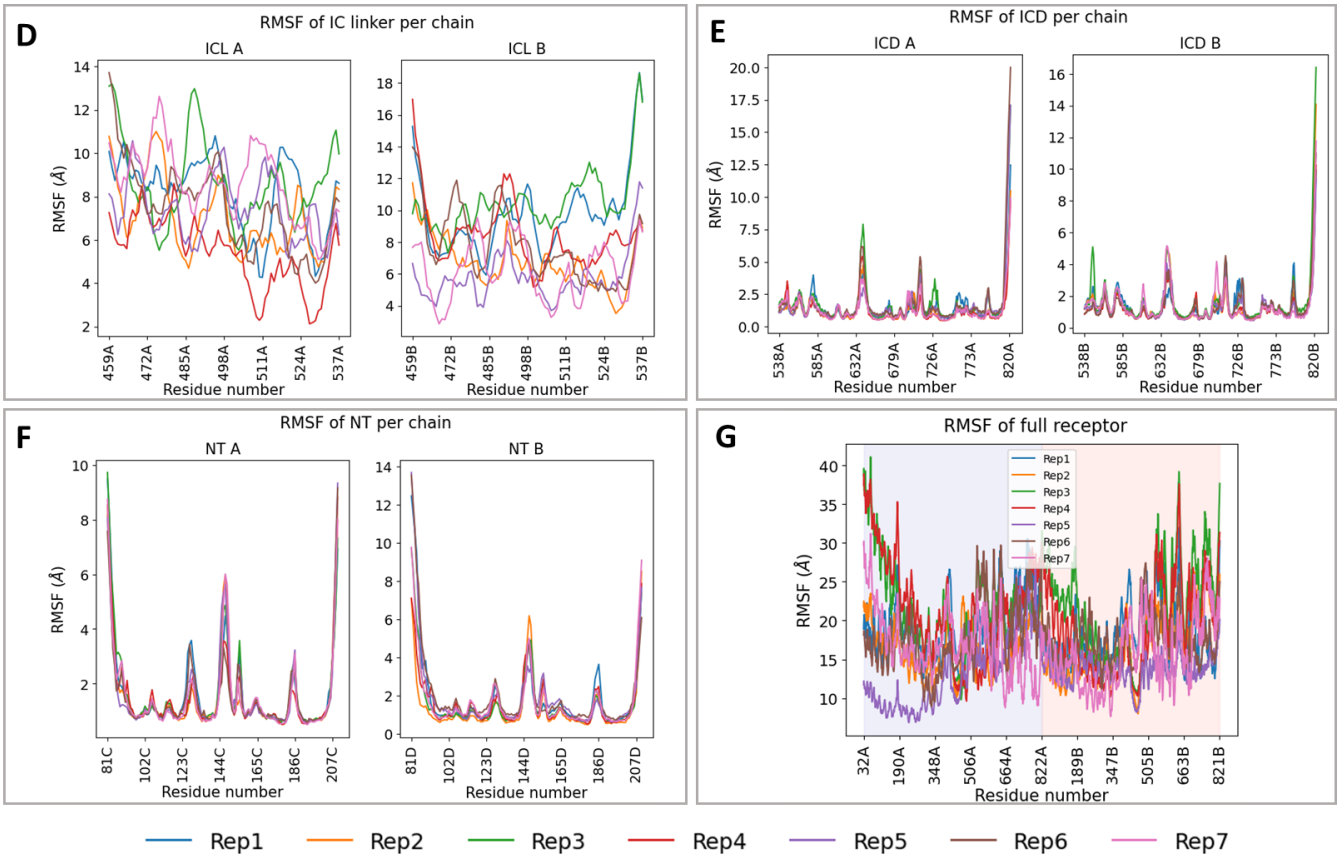


Figure A.32: RMSF calculations of the different domains from the full-length TrkB simulations.

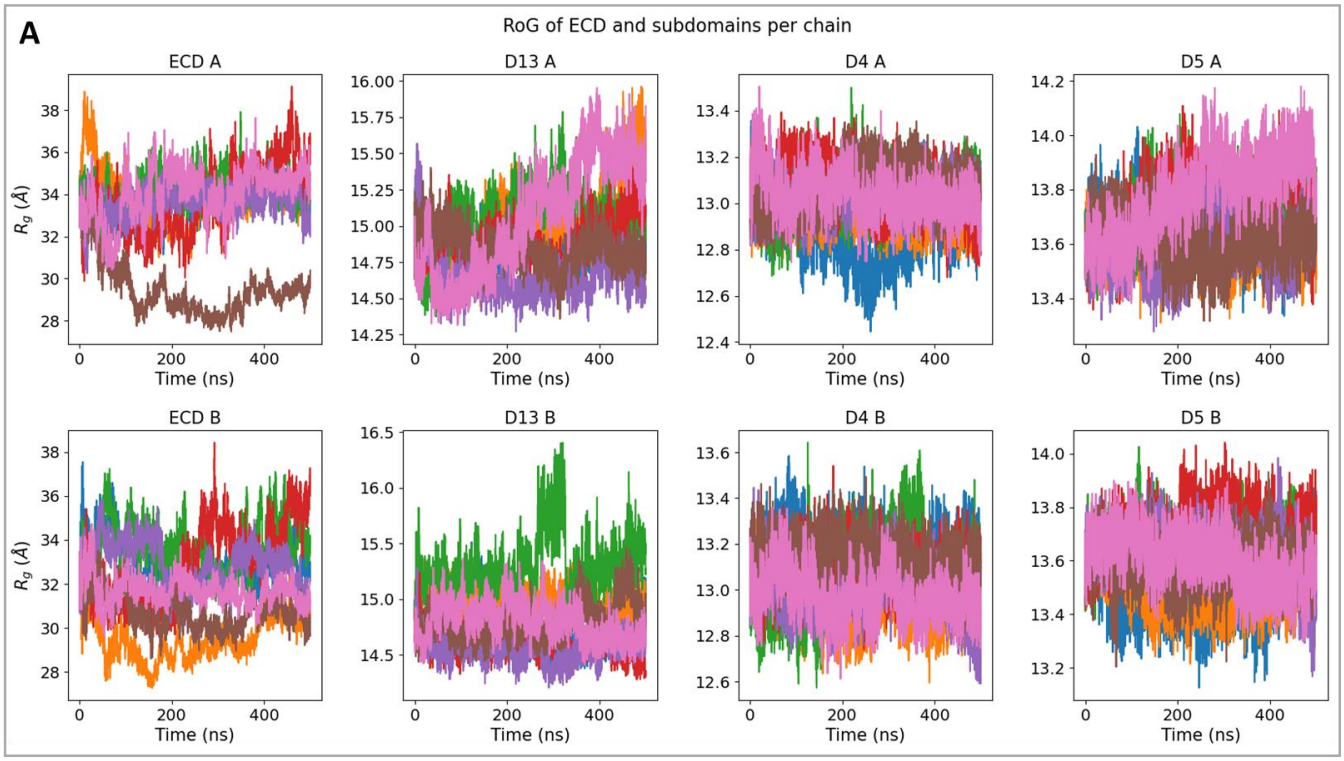
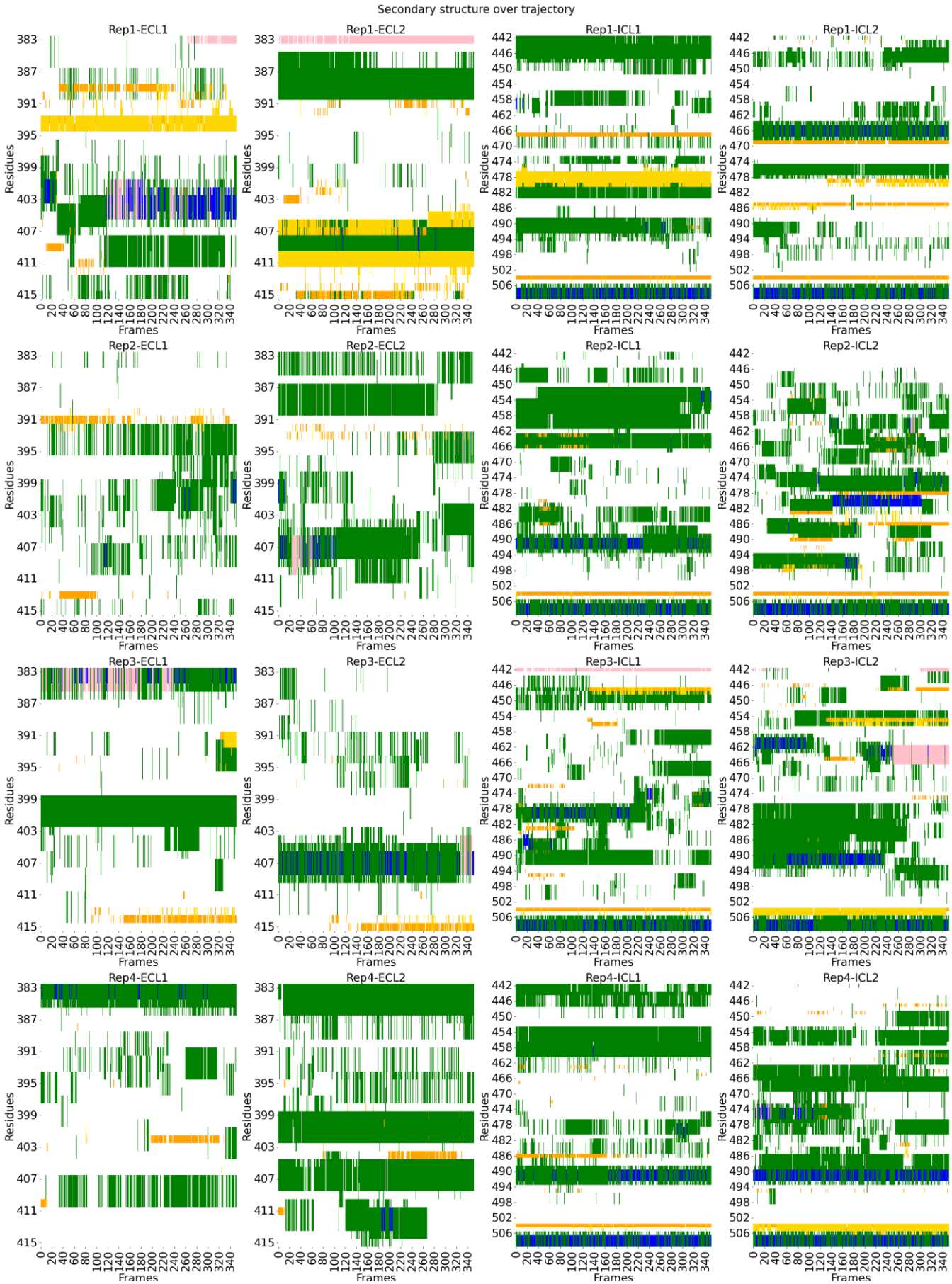
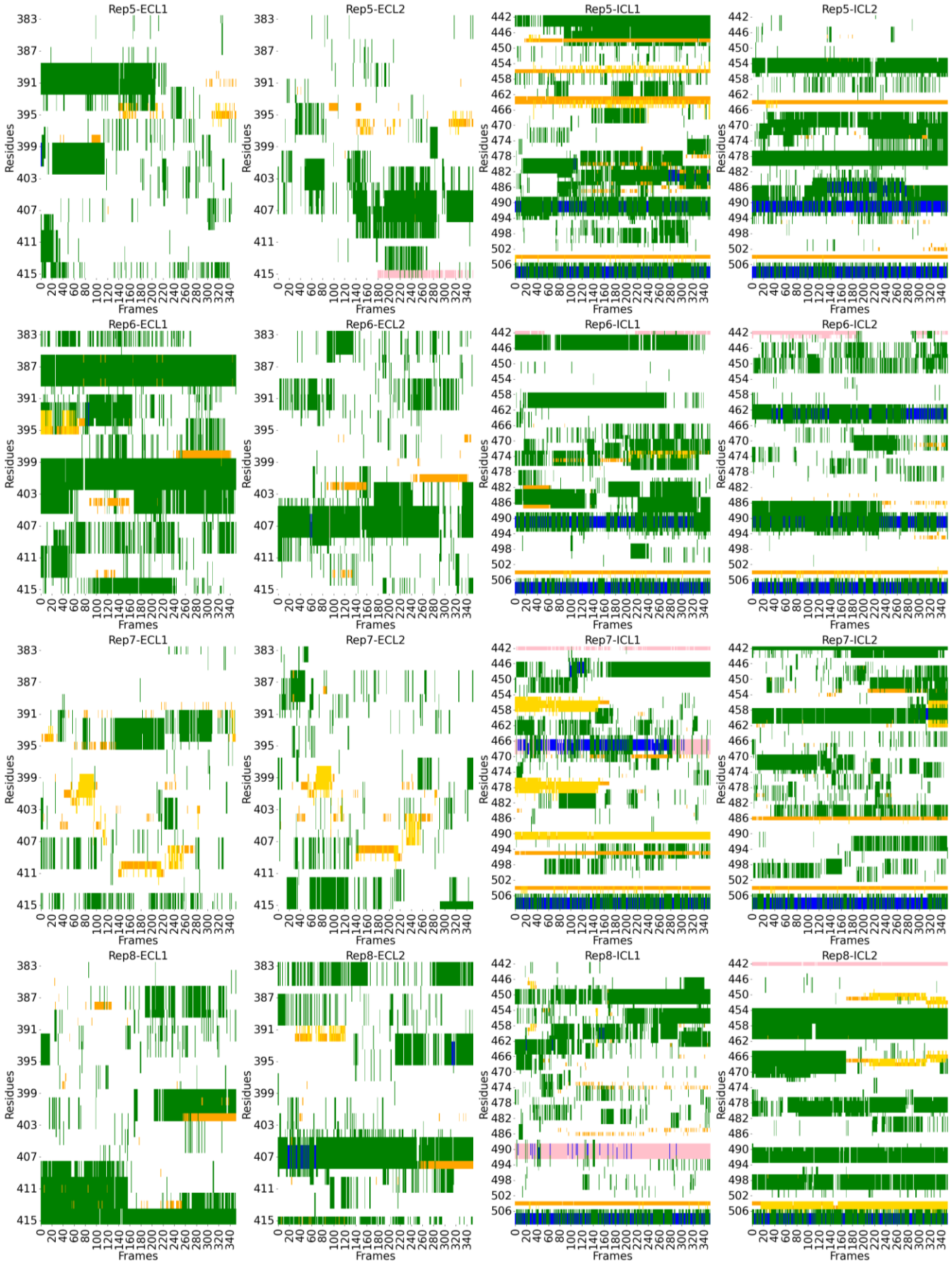




Figure A.33: Radius of gyration (R_g) calculations of the different domains from the full-length TrkB simulations.


Secondary structure over trajectory
Rep1-ECL1
Rep1-ECL2
Rep1-ICL1
Rep1-ICL2
Rep2-ECL1
Rep2-ECL2
Rep2-ICL1
Rep2-ICL2
Rep3-ECL1
Rep3-ECL2
Rep3-ICL1
Rep3-ICL2
Rep4-ECL1
Rep4-ECL2
Rep4-ICL1
Rep4-ICL2
Residues
Frames


Rep5-ECL1
Rep5-ECL2
Rep5-ICL1
Rep5-ICL2
Rep6-ECL1
Rep6-ECL2
Rep6-ICL1
Rep6-ICL2
Rep7-ECL1
Rep7-ECL2
Rep7-ICL1
Rep7-ICL2
Rep8-ECL1
Rep8-ECL2
Rep8-ICL1
Rep8-ICL2
Residues
Frames

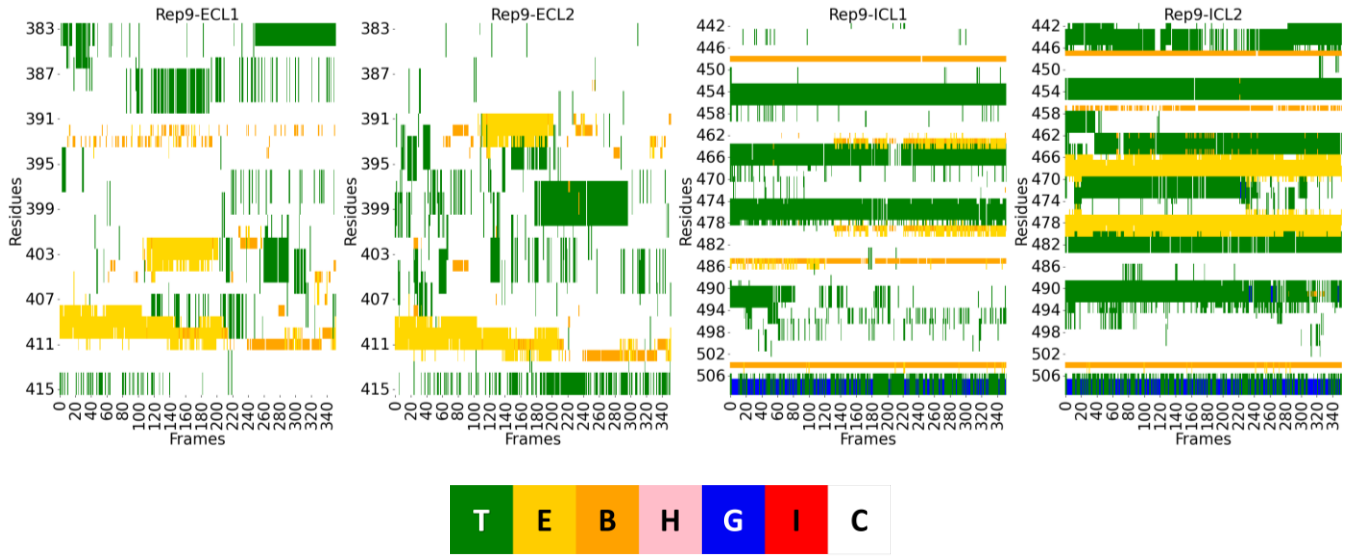
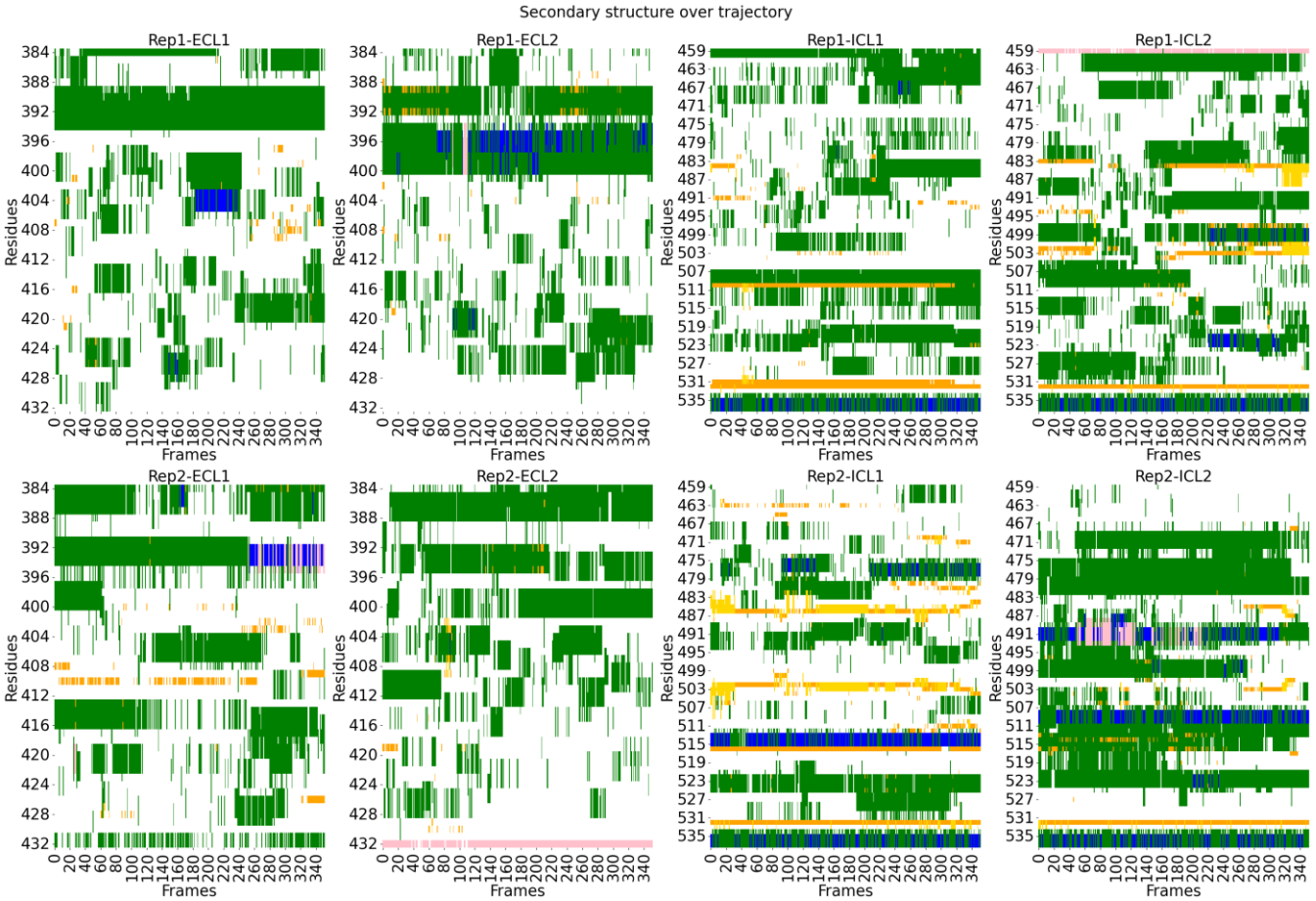
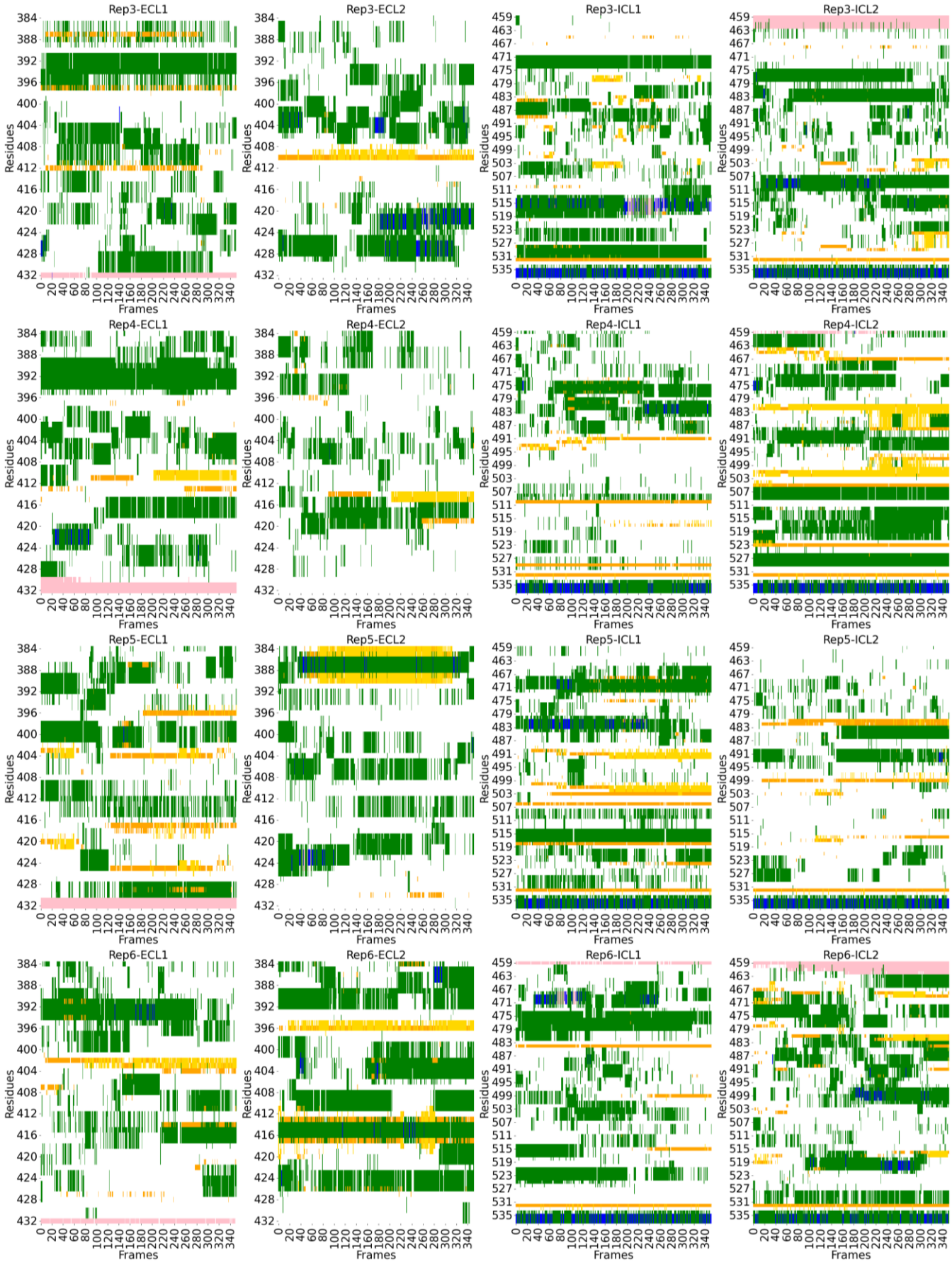


Figure A.34: Secondary structure calculation for the EC and IC linkers of TrkA receptor in the last 350 frames (last 350 ns) of the simulations. The letters in the legend correspond to: T: turn (a beta sheet component), E: extended configuration (the main component of beta sheets), B: isolated bridge, H: alpha helix, G: 3-10 helix, I: pi-helix, C: random coil (no structure).



9 Appendix


Rep3-ECL1
Rep3-ECL2
Rep3-ICL1
Rep3-ICL2
Rep4-ECL1
Rep4-ECL2
Rep4-ICL1
Rep4-ICL2
Rep5-ECL1
Rep5-ECL2
Rep5-ICL1
Rep5-ICL2
Rep6-ECL1
Rep6-ECL2
Rep6-ICL1
Rep6-ICL2
Residues
Frames

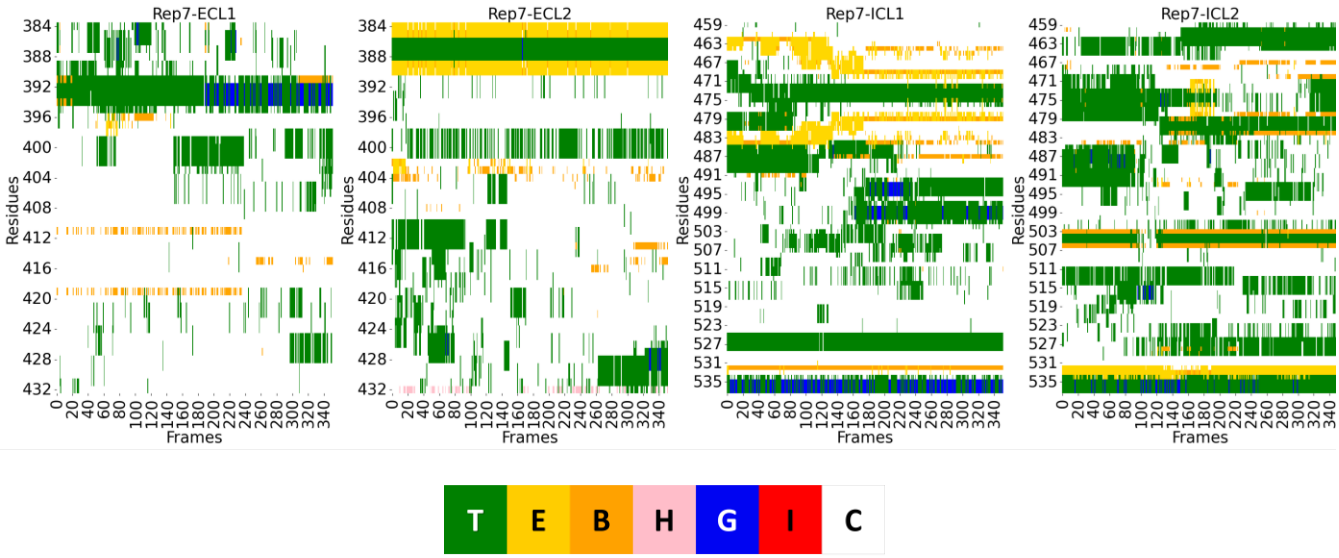
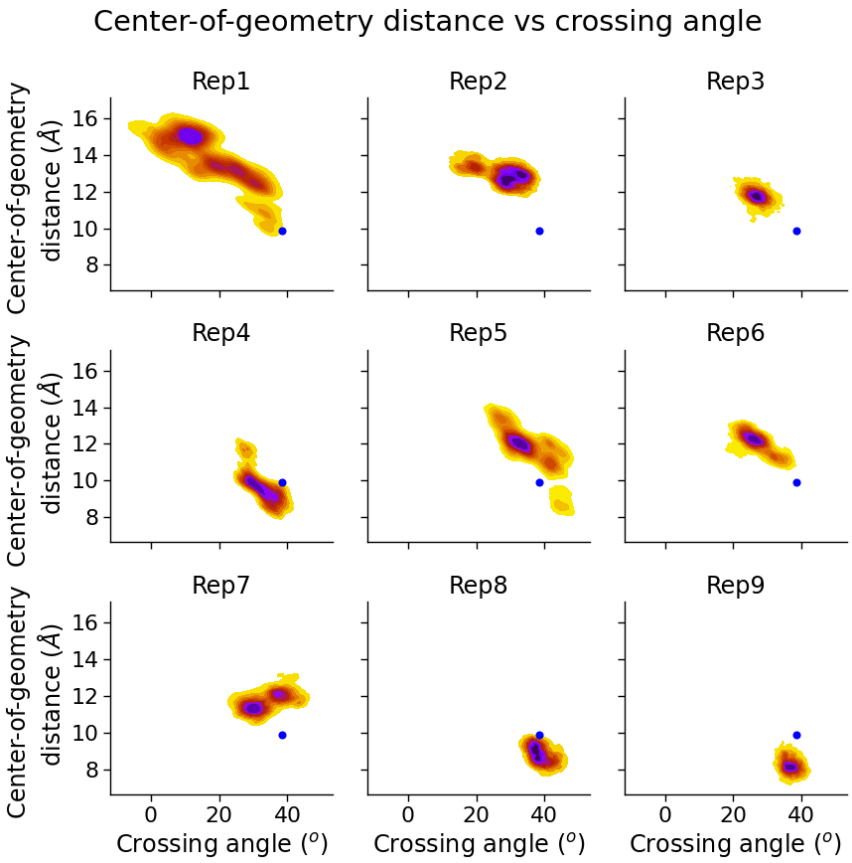
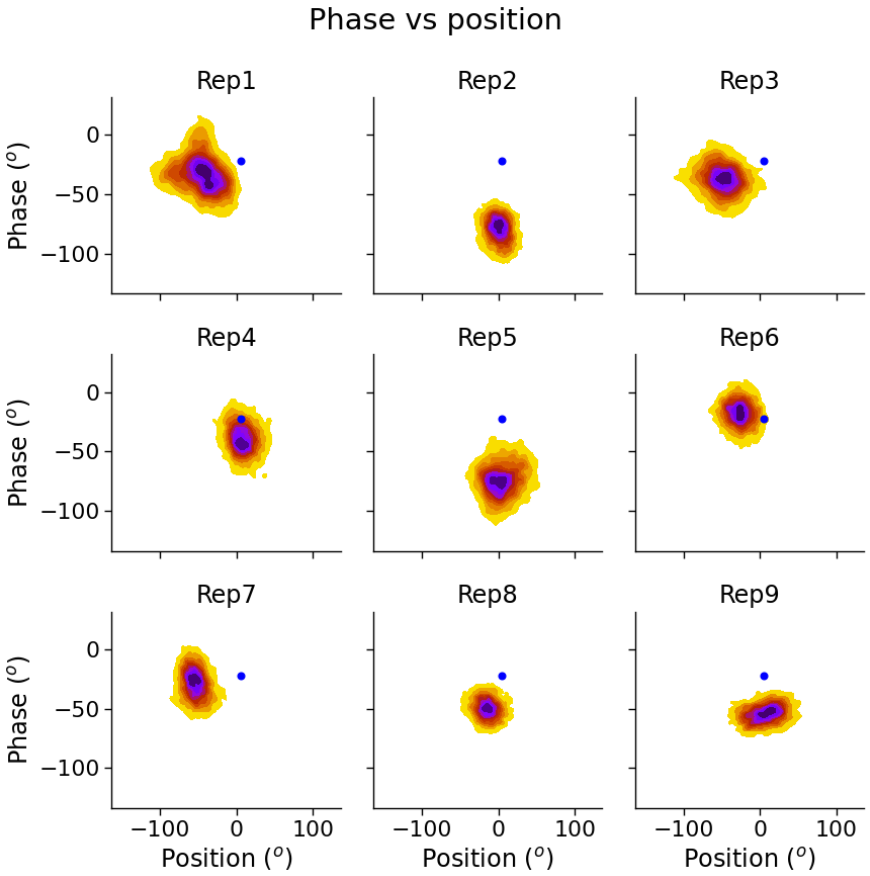


Figure A.35: Secondary structure calculation for the EC and IC linkers of TrkB receptor in the last 350 frames (last 350 ns) of the simulations. The letters in the legend correspond to: T: turn (a beta sheet component), E: extended configuration (the main component of beta sheets), B: isolated bridge, H: alpha helix, G: 3-10 helix, I: pi-helix, C: random coil (no structure).




Phase vs position
Rep1
Rep2
Rep3
Rep4
Rep5
Rep6
Rep7
Rep8
Rep9
Phase (°)
0
−50
−100
−100
0
100
Position (°)

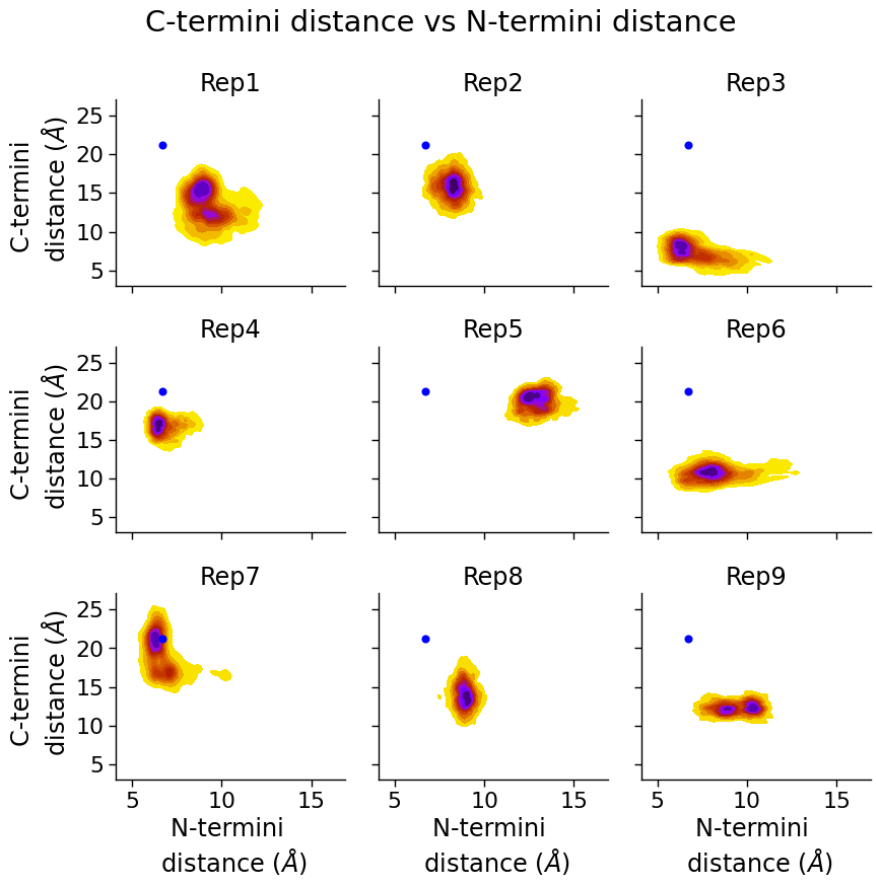
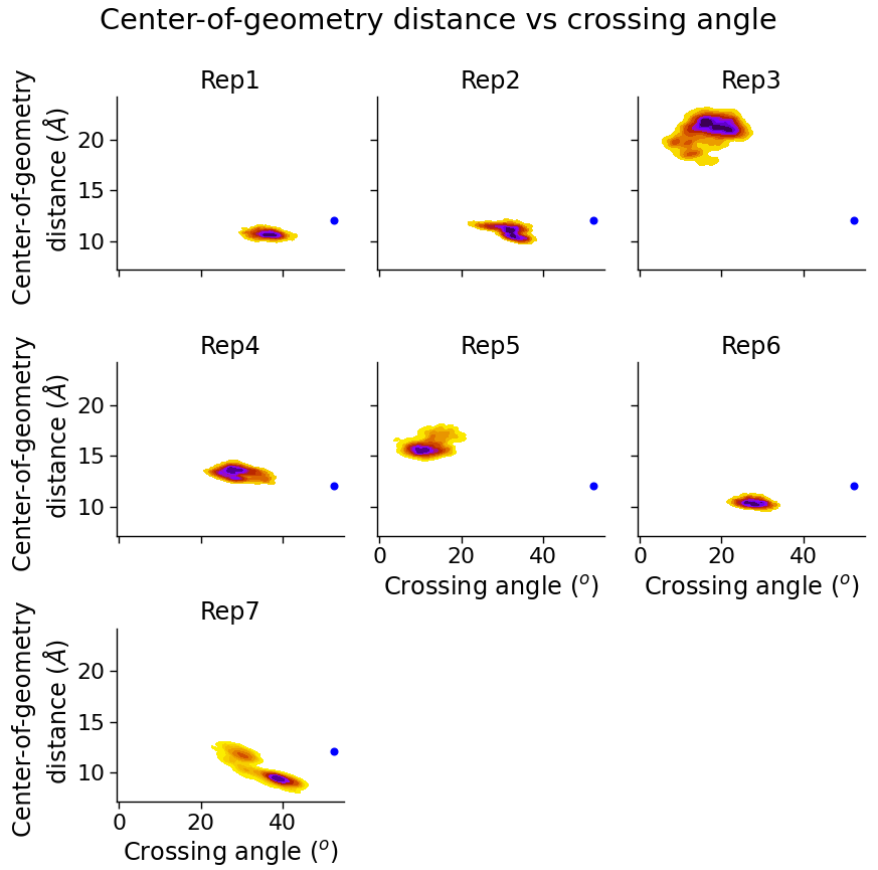
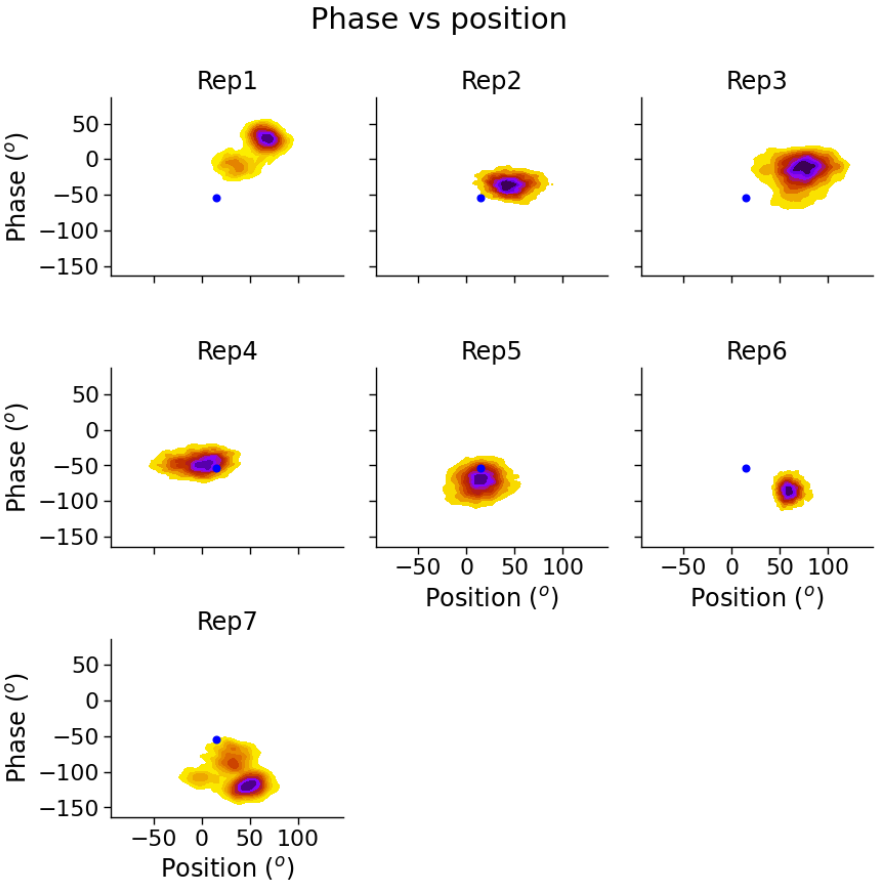


Figure A.36: Order parameters of TM helices in the several replicas of TrkA full-length systems. KDE plots for six different order parameters: interhelical distance (d), interhelical crossing angle (Ω), phase (rotation around 1st helix), position (rotation around 2nd helix), N- and C- termini distances. The NMR state is shown with a blue dot.


Center-of-geometry distance vs crossing angle
Rep1
Rep2
Rep3
Rep4
Rep5
Rep6
Rep7
Center-of-geometry distance (Å)
Crossing angle (°)
10
15
20
0
20
40


Phase vs position
Rep1
Rep2
Rep3
Rep4
Rep5
Rep6
Rep7
Phase (°)
Position (°)
50
0
−50
−100
−150
−50
0
50
100

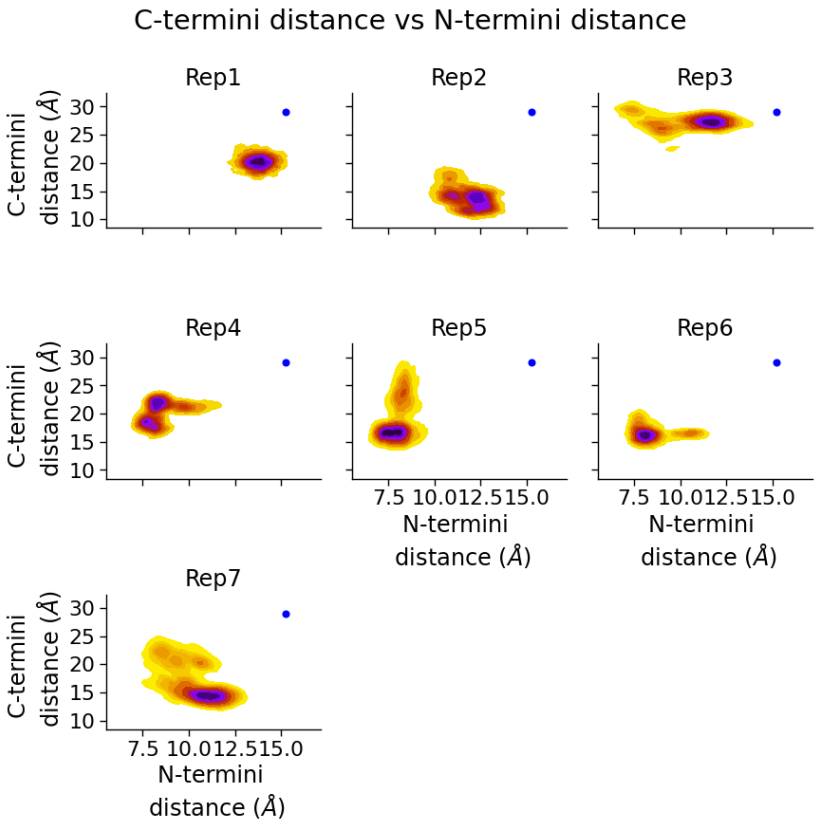


Figure A.37: Order parameters of TM helices in the several replicas of TrkB full-length systems. KDE plots for six different order parameters: interhelical distance (d), interhelical crossing angle (Ω), phase (rotation around 1st helix), position (rotation around 2nd helix), N- and C- termini distances. The NMR state is shown with a blue dot.

10 List of Tables

TABLE 1.1: NEUROTROPHINS AS LIGANDS FOR NEUROTROPHIN RECEPTORS. LIGANDS LISTED IN BOLD HAVE GREATER AFFINITY AND THEY ARE MORE IMPORTANT FOR RECEPTOR ACTIVATION *IN VIVO*. 13

TABLE 3.1: N-GLYCOSYLATION POSITIONS ON TRKA AND TRKB RECEPTORS. FOR TRKB, THE EXACT GLYCOSYLATION POSITIONS ARE KNOWN EXPERIMENTALLY.[44,48] FOR TRKA, THE GLYCOSYLATION POSITIONS THAT ARE KNOWN FROM THE PDB ID: 2IFG[28] CRYSTAL STRUCTURE ARE SHOWN WITH NORMAL FONT, WHILE THE ONES THAT I ADDED BASED ON COMPARISON WITH TRKB POSITIONS ARE SHOWN IN BOLD LETTERS. 59

TABLE 3.2: SUMMARY OF SYSTEMS THAT I SIMULATED WITH OR WITHOUT GLYCANS. 62

TABLE 4.1: BIOINFORMATICS WEBSERVERS AND DATABASES PREDICTING THE TM PART (UPPER TABLE) OR SECONDARY STRUCTURE TYPE (LOWER TABLE) OF THE TRKB SEQUENCE. THE TM SEQUENCE PREDICTED BY EACH SERVER IS HIGHLIGHTED WITH GREEN IN THE UPPER TABLE, WHILE THE HELICAL OR STRAND PREDICTED REGIONS ARE HIGHLIGHTED WITH GREEN AND ORANGE, RESPECTIVELY, IN THE LOWER TABLE. THE DASHED LINES DEFINE THE REGION THAT I MODELED AS HELICAL IN THE FINAL MODEL. 96

TABLE 4.2: SUMMARY OF THE SIMULATIONS THAT I PERFORMED. TEST SIMULATIONS ARE DENOTED WITH *. I PERFORMED SIMULATIONS FOR THE TRKA, TRKB TM DIMERS. IN SYSTEMS 3 AND 4, "I" IS THE IONIC STRENGTH AND "DPR" THE DETERGENT-TO-PROTEIN RATIO. IN SYSTEMS 7-10, I SPECIFY THE FORCE FIELD WITH MARTINI 2.2 (M2) AND MARTINI 3 (M3). THE SIMULATION TYPES THAT I PERFORMED ARE THE ALL-ATOM (AA), COARSE-GRAINED (CG) WITH MARTINI 2.2 AND COARSE-GRAINED METADYNAMICS (CG-METAD). 100

TABLE 5.1: COMPOSITION OF MODEL OF NEURONAL MEMBRANE BUILT FOR SIMULATING FULL-LENGTH NT RECEPTORS. THE LIPID TAIL TYPE IS PRESENTED IN THE FORM "[C ATOMS OF 1ST TAIL]:[NUMBER OF UNSATURATED BONDS IN THE 1ST TAIL]/ [C ATOMS OF 2ND TAIL]:[NUMBER OF UNSATURATED BONDS IN THE 2ND TAIL]". THE EXACT NUMBERS OF THE LIPIDS THAT WERE INCLUDED IN THE OUTER AND INNER LEAFLETS ARE PRESENTED IN THE 3RD AND 4TH COLUMNS. THE LIPIDS ARE PRESENTED IN ORDER FROM THE MOST TO THE LEAST ABUNDANT HEADGROUP SPECIES. THE LIPID NAME CODES ARE: 1,2-DIPALMITOYL-SN-GLYCERO-3-PHOSPHOCHOLINE (DPPC), 1-PALMITOYL-2-OLEOYL-SN-GLYCERO-3-PHOSPHOCHOLINE (POPC), 1-PALMITOYL-2-ARACHIDONOYL-SN- GLYCERO-3-PHOSPHOCHOLINE (PAPC), 1,2-DIPALMITOYL-SN-GLYCERO-3-PHOSPHOETHANOLAMINE (DPPE), 1-PALMITOYL-2-OLEOYL-SN-GLYCERO-3-PHOSPHOETHANOLAMINE (POPE), 2-LINOLEOYL-SN- GLYCERO-3-PHOSPHOETHANOLAMINE (LLPE), 1-PALMITOYL-2-OLEOYL-SN-GLYCERO-3-PHOSPHO-L-SERINE (POPS), 1-PALMITOYL-2-LINOLEOYL-SN-GLYCERO-3-PHOSPHO-L-SERINE (PLPS), 1-PALMITOYL-2-OLEOYL-SN-GLYCERO-3-PHOSPHOINOSITOL (POPI), 1-PALMITOYL-2-OLEOYL-SN-GLYCERO-3-PHOSPHOINOSITOL 4,5-BISPHOSPHATE (POPI25 OR PIP2), 1-STEAROYL-2-ARACHIDONOYL-SN-GLYCERO-3-PHOSPHOINOSITOL (SAPI), N-STEAROYL SPHINGOMYELIN (SSM), NEUTRAL SPHINGOMYELIN (NSM), CERAMIDE 18:0 (CER180), MONOSIALO-TETRAHEXOSYL-GANGLIOSIDE (GM1), MONOSIALO-DIHEXOSYL-GANGLIOSIDE (GM3), GALACTOSYL-CERAMIDE 18:0 (GALCER180), GALACTOSYL-CERAMIDE 24:1 (GALCER241). THE LAST TWO ARE CEREBROSIDES. 127

TABLE 5.2: EQUILIBRATION PROTOCOL OF TM-IC AND FULL-LENGTH SYSTEMS IN A MODEL NEURONAL MEMBRANE, WITH GRADUAL REMOVAL OF RESTRAINTS. 130

TABLE 5.3: SUMMARY OF SIMULATIONS TM-IC AND FULL-LENGTH HOMODIMERIC TRKA AND TRKB ATOMISTIC SYSTEMS IN A MODEL NEURONAL MEMBRANE. 131

TABLE 6.1: SMALL MOLECULE NEUROTROPHIN MIMETICS DISCOVERED BY THE EURONEUROTROPHIN CONSORTIUM. THE BIOLOGICAL TESTS (UNPUBLISHED EXCEPT FOR ENT-A011[330] AND ENT-A013E/Z[204], PROVIDED BY ALESSIA LATORRATA, NATIONAL HELLENIC RESEARCH

Foundation, NHRF, Greece) are the following for TrkA expressing cells: phosphorylation of TrkA (pTrkA), AKT (pAKT) and ERK1/2 (pERK1/2) and cell survival. Similar for TrkB expressing cells.....154
Table 6.2: Summary of simulated TrkA and TrkB EC systems with NTs and two compounds bound. 156
Table 6.3: Summary of simulated TrkA and TrkB TM systems with one or ten compound molecules. The inactive dimers correspond to the NMR structure of TrkA and to the model of TrkB based on the TrkA NMR structure. The active state of the dimers was modeled by manually rotating the helices. "Dist. Rest." refers to distance restraints. 157
Table 6.4: Summary of simulated systems for membrane permeability of the compounds. 158

11 List of Figures

FIGURE 1.1: CARTOON WITH MOLECULAR SURFACE MODEL OF THE NGF HOMODIMER X-RAY STRUCTURE (PDB ID: 1BET)[20] WITH THE POSITION OF THE LOOPS INDICATED.[25] 14

FIGURE 1.2: SCHEMATIC REPRESENTATION OF THE FULL-LENGTH TRK AND P75 RECEPTOR HOMODIMERS WITH NEUROTROPHINS (NTS) BOUND TO THE EXTRACELLULAR DOMAINS. TRK RECEPTORS CONSIST OF FIVE EXTRACELLULAR DOMAINS (D1–D5), AN A-HELICAL TRANSMEMBRANE (TM) DOMAIN, AND AN INTRACELLULAR SEGMENT CONTAINING A KINASE DOMAIN. THE P75 RECEPTOR HAS FOUR CYSTEINE-RICH DOMAINS (CRD1-CRD4) IN THE EXTRACELLULAR PART, AN A-HELICAL TRANSMEMBRANE DOMAIN, AS THE TRKS, AND AN INTRACELLULAR DEATH DOMAIN. BOTH TRK AND P75 RECEPTORS CAN ALSO EXIST IN PREFORMED INACTIVE DIMERS PRIOR TO NEUROTROPHIN BINDING. NEUROTROPHIN BINDING LEADS TO ACTIVATION OF THE RECEPTORS AND INITIATION OF SIGNALING CASCADES RELATED TO CELL SURVIVAL OR APOPTOSIS, DIFFERENTIATION AND SYNAPTIC PLASTICITY. 16

FIGURE 1.3: STRUCTURE OF OLIGOMANNOSE, COMPLEX AND HYBRID N-GLYCANS IN SYMBOL NOMENCLATURE FOR GLYCANS (SNFG) REPRESENTATION. THE TYPES OF SUGARS THAT CORRESPOND TO EACH SHAPE AND COLOR ARE EXPLAINED WITHIN THE LEGEND. THE NUMBERS IN EACH LINKAGE DENOTE THE POSITIONS OF THE TWO ATOMS FROM THE TWO SUGARS THAT PARTICIPATE IN THE GLYCOSIDIC BOND. THE LETTERS A AND B REFER TO THE CONFIGURATION (ALPHA OR BETA) OF THE ANOMERIC CARBON (C1) THAT FORMS THE GLYCOSIDIC BOND. THE GLYCAN CORE, WHICH IS COMMON IN ALL N-GLYCANS, IS SHOWN WITHIN A DASHED-LINE BOX. 18

FIGURE 1.4: FIRST CRYSTAL STRUCTURE OF THE NGF HOMODIMER (GREEN) IN COMPLEX WITH THE TRKA-D5 HOMODIMER (ORANGE) (PDB ID: 1WWW).[29] THE PROTEINS ARE SHOWN IN CARTOON AND MOLECULAR SURFACE REPRESENTATIONS. THE DASHED-CIRCLED REGIONS INDICATE THE *COMMON* AND *SPECIFIC PATCHES* OF INTERACTIONS BETWEEN NGF AND TRKA-D5. 21

FIGURE 1.5: CRYSTAL STRUCTURE OF THE NGF HOMODIMER (GREEN) IN COMPLEX WITH THE FULL EXTRACELLULAR TRKA HOMODIMER (ORANGE) (PDB ID: 2IFG).[28] THE PROTEINS ARE SHOWN IN CARTOON AND MOLECULAR SURFACE REPRESENTATIONS. THE DIFFERENT DOMAINS OF THE FULL EXTRACELLULAR TRKA ARE INDICATED IN THE FIGURE. N-GLYCANS THAT WERE SOLVED IN THE STRUCTURE ARE SHOWN IN 3D-SYMBOL NOMENCLATURE FOR GLYCANS (3D-SNFG). THE BLUE CUBES CORRESPOND TO N-ACETYL GLUCOSAMINE (GLCNAC) AND THE GREEN SPHERES TO MANNOSE (MAN). 22

FIGURE 1.6: CRYSTAL STRUCTURES OF THE NGF HOMODIMER (GREEN) IN COMPLEX WITH THE TRKA-D5 HOMODIMER (ORANGE) (PDB ID: 1WWW),[29] AND OF THE NT-4/5 HOMODIMER (PINK) IN COMPLEX WITH THE TRKB-D5 HOMODIMER (GREY) (PDB ID: 1HCF).[59] THE PROTEINS ARE SHOWN IN CARTOON AND MOLECULAR SURFACE REPRESENTATIONS. THE DASHED-CIRCLED REGIONS INDICATE THE *COMMON* AND *SPECIFIC PATCHES* OF INTERACTIONS BETWEEN NTS AND RECEPTORS. 23

FIGURE 1.7: CRYSTAL STRUCTURE OF THE INACTIVE TRKA KINASE DOMAIN IN CARTOON REPRESENTATION (PDB ID: 5H3Q).[70] THE KINASE DOMAIN STRUCTURE CAN BE DIVIDED IN TWO PARTS, THE N-LOBE (LIGHT PINK) AND THE C-LOBE (LIGHT GREEN). THE ATP BINDING SITE IS LOCATED IN A DEEP CLEFT AT THE INTERFACE OF THE TWO LOBES AND SITS BENEATH A HIGHLY CONSERVED LOOP (P LOOP IN MAGENTA) CONNECTING STRANDS B1 AND B2. THIS PHOSPHATE-BINDING LOOP, OR P LOOP, CONTAINS A CONSERVED GLYCINE-RICH SEQUENCE MOTIF (GXGXΦG), WHERE Φ IS USUALLY TYROSINE OR PHENYLALANINE. THE GLYCINE RESIDUES ALLOW THE P LOOP TO APPROACH THE PHOSPHATES OF ATP VERY CLOSELY AND TO COORDINATE THEM VIA BACKBONE INTERACTIONS. THE CONSERVED AROMATIC SIDE CHAIN CAPS THE SITE OF PHOSPHATE TRANSFER. THE GLYCINE RESIDUES MAKE THE P LOOP VERY FLEXIBLE IN THE ABSENCE OF ATP, A FACT THAT FACILITATES THE BINDING OF SMALL MOLECULE INHIBITORS. PHOSPHORYLATION OF THE ACTIVATION LOOP (GREEN) TYROSINES (Y676, Y680, AND Y681) STABILIZES IT IN AN OPEN AND EXTENDED CONFORMATION THAT IS PERMISSIVE FOR SUBSTRATE BINDING. IN THE UNPHOSPHORYLATED STATE, THE ACTIVATION

loop collapses into the active site, blocking the binding of the substrate, as shown here. Optimal phosphotransfer requires the precise spatial arrangement of several catalytic residues that are absolutely conserved among all known kinases. An important role is played by a conserved loop structure at the base of the active site, called the catalytic loop (dark green). A highly conserved to all kinases DFG motif is situated at the base of the activation loop. The structure of this motif, and in particular the conformation of Asp 668, is tightly coupled to phosphorylation of the activation loop. After the phosphorylation of the activation loop tyrosines, two more tyrosines are also phosphorylated by the other kinase monomer through trans-autophosphorylation. These are Y496 and Y791, which are located in the juxtamembrane (JM) domain and the C-terminal linker, respectively. 25

Figure 1.8: (A) Sequences of the transmembrane (TM) domains of TrkA and TrkB receptors. The helical part from the TrkA NMR structure (PDB ID: 2N90)[60] is indicated with the dashed lines, while the N- and C-terminal loop sequences in the solved NMR structure are also shown. Mutated residues in the NMR sequence compared to the human TrkA are colored blue. Residues participating in the GxxxG-like motifs are colored magenta, while the TrkA NMR interface LxxFAxxF motif is highlighted in yellow. (B) Cartoon representation of the TrkA NMR structure (PDB ID: 2N90), with the interface LxxFAxxF motif shown in yellow licorice representation. (C) Helical wheel representation of the TrkA TM helix. The arrow indicates the magnitude and direction of the hydrophobic moment (μ_H). Nonpolar and aromatic residues are yellow, glycines and alanines are grey, charged residues are blue, polar uncharged residues are purple, and asparagine is pink. The letters N and C in the projection indicate the N-and C-terminals, respectively. The active and inactive interfaces, as determined from mutagenesis studies,[60] are indicated with green and red lines, respectively. Rotation from the inactive to the active interface takes place upon NGF binding. The helical wheel was created with the HeliQuest webserver.[87] 29

Figure 1.9: Proposed mechanisms for creation of high-affinity NGF binding sites in the TrkA receptor by the p75 receptor. (A) Schematic representation of the NGF homodimer (yellow), the TrkA receptor monomer (purple) and the p75 receptor monomer (red). (B) Stabilized 1:2:1 hetero-complex model: p75 binds in a reversed way to NGF. TrkA activation is depicted by red phosphorylation sites in the intracellular domain. (C) Ligand-passing model presented in three steps. (D) Allosteric modulation model: p75 interacts directly with TrkA, inducing an allosteric conformational change in TrkA that allows better access of NGF to the TrkA binding site. (E) Inside-out activation model: TrkA exists in four states; low-affinity inactive form, a low-affinity active form, a high-affinity inactive, and high affinity active form. The last two are induced by binding of the intracellular segment of p75 to TrkA. The figure was created with BioRender.com. 32

Figure 2.1: Force field terms describing bonded and non-bonded interactions. The bonded terms describe (A) the bond stretching, and the fluctuations of (B) bond angles, (C) torsional angles, (D) improper torsions. The non-bonded terms describe (E) the electrostatic and (F) the van der Waals interactions. 47

Figure 3.1: Summary of modeled glycosylated and unglycosylated systems. The systems that I modeled to test glycosylation effects on the EC domains of TrkA and TrkB receptors were: the TrkA-EC homodimer with the NGF homodimer bound, the TrkA-EC monomer with the NGF homodimer bound, the TrkA-EC monomer, the TrkB-EC homodimer with the NT-4/5 homodimer bound, the TrkB-EC monomer with the NT-4/5 homodimer bound, the TrkB-EC monomer with and without glycans. All systems are shown in cartoon representation with the receptors in green and the NTs in blue. The systems with glycans present are shown with all the glycan conformations that were sampled during the simulations from all frames. The glycans are shown in yellow wire representation. The numbers of replica simulations that I ran for each system are also indicated. The last box at the bottom of the figure shows the type of the glycan, $Man_3GlcNAc_2$, which was attached to

all nine glycosylation positions of the receptors. This is in Symbol Nomenclature for Glycans (SNFG) representation with the blue boxes denoting the N-acetyl-glucosamine and the green circles the mannose monosaccharides. 60

Figure 3.2: Schematic representation of the solvent accessible surface area (SASA) calculations that I performed to estimate the contact area of the D5 domain with the NT. Specifically, I calculated the SASA of the D5 domain without and with the NT present and the difference of these two values gives the contact area of D5-NT. 64

Figure 3.3: Comparison of the several (A) TrkA and (B) TrkB systems with all the glycan conformations sampled during the simulations. The proteins are shown in cartoon green and blue representation, while glycans from the different frames are shown in yellow wire representation. I performed alignment to each subdomain before the visualization of its glycans.... 66

Figure 3.4: RMSD calculations of the glycosylated and non-glycosylated TrkA systems. (A), (B) and (C) show the RMSD for the TrkA-EC homodimer with the NGF homodimer bound, the TrkA-EC monomer with the NGF homodimer bound, the TrkA-EC monomer systems, respectively, with and without glycans. For each system the time-dependence of the RMSD of the protein backbone compared to the first frame (which corresponds to the crystal structure) is shown for each replica, as well as the distributions of RMSD values across all replicas and all frames of the production runs for the glycosylated and non-glycosylated systems (right plots in A, B and C). 68

Figure 3.5: RMSD calculations of the glycosylated and non-glycosylated TrkB systems. (A), (B) and (C) show the RMSD for the TrkB-EC homodimer with the NT-4/5 homodimer bound, the TrkB-EC monomer with the NT-4/5 homodimer bound, the TrkB-EC monomer systems, respectively, with and without glycans. For each system the time-dependence of the RMSD of the protein backbone compared to the first frame (which corresponds to the crystal structure) is shown for each replica, as well as the distributions of RMSD values across all replicas and all frames of the production runs for the glycosylated and non-glycosylated systems (right plots in A, B and C). 69

Figure 3.6: RMSD calculations for each domain of the glycosylated or not TrkA systems. (A), (B) and (C) show the internal RMSD for each domain the TrkA-EC homodimer with the NGF homodimer bound, the TrkA-EC monomer with the NGF homodimer bound, the TrkA-EC monomer systems, respectively, with and without glycans. For each system the time-dependent RMSD of the domain backbone compared to the first frame is shown for each replica. Before the calculation of the internal RMSD of each domain, I performed alignment of the respective domain at each simulation snapshot. Schematic representation of each system, with color coding of the domains is shown at the left. The first and second subunits are indicated with A and B letters in the legends. 71

Figure 3.7: RMSD calculations for each domain of the glycosylated or not TrkB systems. (A), (B) and (C) show the internal RMSD for each domain the TrkB-EC homodimer with the NT-4/5 homodimer bound, the TrkB-EC monomer with the NT-4/5 homodimer bound, the TrkB-EC monomer systems, respectively, with and without glycans. For each system the time-dependent RMSD of the domain backbone compared to the first frame is shown for each replica. Before the calculation of the internal RMSD of each domain, I performed alignment in the respective domain at each simulation snapshot. Schematic representation of each system, with color coding of the domains is shown at the left. The first and second monomers are indicated with A and B letters in the legends. 72

Figure 3.8: RMSF calculations of all the glycosylated or not TrkA and TrkB systems. (A), (B) and (C) show the average RMSF values from all replica simulations for each receptor and neurotrophin residue of the TrkA/B-EC homodimer with the NGF/ NT-4/5 homodimer bound, the TrkA/B-EC monomer with the NGF/ NT-4/5 homodimer bound, the TrkA/B-EC monomer systems, respectively, with and without glycans. The average RMSF values from all replica simulations are

shown for each residue in black and red lines for the glycosylated and non-glycosylated systems, respectively. The standard deviation for each residue is shown with shadowed areas. The residue positions of glycosylated asparagines are shown with vertical lines. Before the RMSF calculation, I aligned the systems to an average structure from the simulations. 74

Figure 3.9: Box plots of RMSF calculations of all glycans attached in asparagine residues in all TrkA and TrkB glycosylated systems. Before the RMSF calculation, I performed alignment of each EC segment to its average structure from the simulations. 75

Figure 3.10: Accessible surface area (ASA) of the (A) TrkA and (B) TrkB systems (including the ASA of the NTs when present) and the area shielded by glycans at multiple probe radii from 1.4 Å (water molecule) to 10 Å (protein-sized molecule). The values have been calculated and averaged across all replicas with glycans and are reported with standard deviation bars. The area shielded by the glycans is presented in yellow (rounded % values are reported), whereas the green line represents the accessible area of the protein in the absence of glycans. Highlighted in green is the area that remains accessible in the presence of glycans. The ASA for the TrkB dimer and monomer systems, which was calculated for probe radius 10 Å, is not shown because of incorrect calculation of the ASA in this high radius for the particular systems. 76

Figure 3.11: Violin plots with the contact area between the D5 domain of TrkA and TrkB receptors with the NTs NGF and NT4/5, respectively. The systems with the EC receptor dimer bound to NTs with and without glycans are shown in (A) and (B), while the systems with the EC receptor monomers bound to NTs with and without glycans are shown in (C) and (D). For the systems with the glycans (A) and (C), I calculated two contact areas: one between the D5 domain together with the glycans and the NT (yellow) and one between the D5 domain without the glycans and the NT (green). I calculated the contact area for each monomer separately and thus included two values (for the two monomers) from each frame. This comparison shows the participation of the glycans in the interaction of the D5 domain with the NT. For the systems without the glycans (B) and (D), only one contact area between the D5 domain and the NT could be calculated and thus there is only one distribution in these violin plots. 78

Figure 3.12: Radius of gyration of the receptor backbone of the glycosylated and not TrkA systems. The color coding of each domain is shown in the schematic representations of the systems at the left of the plots. The time-evolution of the radius of each receptor monomer is shown for each replica, as well as the distributions of the radius of gyration values across all replicas for the glycosylated and non-glycosylated systems (right plots). 80

Figure 3.13: Radius of gyration of the receptor backbone of the glycosylated and not TrkB systems. The color coding of each domain is shown in the schematic representations of the systems at the left of the plots. The time-evolution of the radius of each receptor monomer is shown for each replica, as well as the distributions of the radius of gyration values across all replicas for the glycosylated and non-glycosylated systems (right plots). 81

Figure 3.14: Conformations of the different non-glycosylated TrkA systems, which correspond to low radius of gyration values presented in Figure 3.12. All the conformations are taken from the last frame at 500 ns of the indicated replica simulations, except for the last one which is a rendering at 45 ns of the simulation. 82

Figure 3.15: Conformations of the different non-glycosylated TrkB systems, which correspond to low radius of gyration values presented in Figure 3.13. All the conformations are taken from the last frame of the indicated replica simulations. 83

Figure 3.16: Distance and angle analysis between the receptor monomer domains. (A, B, C) Calculated distances between the centers of geometry (COG) of the D13, D4 and D5 domains from the two different receptor monomers. (D, E, F) Calculated

ANGLES FORMED BETWEEN THE COG OF ONE OF THE SAME DOMAINS FROM THE FIRST RECEPTOR MONOMER, THE COG OF THE NT AND THE COG OF THE SAME DOMAIN FROM THE SECOND MONOMER. IN ALL CASES, THE DISTANCES AND ANGLES WERE CALCULATED EVERY 100 FRAMES AND THE FINAL DISTRIBUTIONS FROM THE GLYCOSYLATED AND NON-GLYCOSYLATED TRKA AND TRKB SYSTEMS ARE SHOWN IN VIOLIN PLOTS. THE QUARTILES OF THE DISTRIBUTIONS ARE INDICATED WITH DASHED LINES. 84

FIGURE 3.17: DISTANCE BETWEEN THE CENTERS OF GEOMETRY OF THE D13-D5 DOMAINS IN THE DIFFERENT TRKA AND TRKB SYSTEMS (A, B, C). IN ALL CASES, THE DISTANCES WERE CALCULATED EVERY 100 FRAMES AND THE FINAL DISTRIBUTIONS FROM THE GLYCOSYLATED AND NON-GLYCOSYLATED TRKA AND TRKB SYSTEMS ARE SHOWN IN VIOLIN PLOTS. THE QUARTILES OF THE DISTRIBUTIONS ARE INDICATED WITH DASHED LINES. 85

FIGURE 3.18: FIRST AND SECOND PRINCIPAL COMPONENTS OF THE D13 DOMAIN DURING THE SIMULATIONS OF THE GLYCOSYLATED AND NOT TRKA SYSTEMS. PCA ANALYSIS WAS PERFORMED TO THE D13 DOMAINS AFTER ALIGNMENT OF EACH EC SEGMENT TO THE D4 AND D5 DOMAINS. THE COLOR CODING OF EACH DOMAIN IS SHOWN IN THE SCHEMATIC REPRESENTATIONS OF THE SYSTEMS AT THE LEFT OF THE PLOTS. 86

FIGURE 3.19: FIRST AND SECOND PRINCIPAL COMPONENTS OF THE D13 DOMAIN DURING THE SIMULATIONS OF THE GLYCOSYLATED AND NOT TRKB SYSTEMS. PCA ANALYSIS WAS PERFORMED TO THE D13 DOMAINS AFTER ALIGNMENT OF EACH EC SEGMENT TO THE D4 AND D5 DOMAINS. THE COLOR CODING OF EACH DOMAIN IS SHOWN IN THE SCHEMATIC REPRESENTATIONS OF THE SYSTEMS AT THE LEFT OF THE PLOTS. 87

FIGURE 3.20: FIRST AND SECOND PRINCIPAL COMPONENTS OF THE LINKER BETWEEN THE D13-D4 DOMAINS PHI AND PSI DIHEDRALS DURING THE SIMULATIONS OF THE GLYCOSYLATED AND NOT TRKA SYSTEMS. DIHEDRAL PCA ANALYSIS WAS PERFORMED TO THE LINKER AFTER ALIGNMENT OF EACH EC SEGMENT TO THE D4 AND D5 DOMAINS. THE COLOR CODING OF EACH DOMAIN IS SHOWN IN THE SCHEMATIC REPRESENTATIONS OF THE SYSTEMS AT THE LEFT OF THE PLOTS. 88

FIGURE 3.21: FIRST AND SECOND PRINCIPAL COMPONENTS OF THE LINKER BETWEEN THE D13-D4 DOMAINS PHI AND PSI DIHEDRALS DURING THE SIMULATIONS OF THE GLYCOSYLATED AND NOT TRKA SYSTEMS. DIHEDRAL PCA ANALYSIS WAS PERFORMED TO THE LINKER AFTER ALIGNMENT OF EACH EC SEGMENT TO THE D4 AND D5 DOMAINS. THE COLOR CODING OF EACH DOMAIN IS SHOWN IN THE SCHEMATIC REPRESENTATIONS OF THE SYSTEMS AT THE LEFT OF THE PLOTS. 89

FIGURE 3.22: CONTACT ANALYSIS BETWEEN THE GLYCANS ATTACHED IN DIFFERENT ASPARAGINE RESIDUES AND THE PROTEIN RESIDUES FROM ALL GLYCOSYLATED TRKA AND TRKB SYSTEMS. A CONTACT IS DEFINED WHEN ANY HEAVY ATOM (NO HYDROGENS) FROM THE GLYCANS IS WITHIN 4.5 Å DISTANCE FROM ANY HEAVY ATOM OF A PROTEIN RESIDUE. THE GLYCANS ARE INDICATED WITH THE ASPARAGINE RESIDUES FROM EACH RECEPTOR MONOMER (WITH A AND B) IN THE X-AXIS, WHILE THE PROTEIN RESIDUES THAT INTERACT WITH THE GLYCANS ARE INDICATED AT THE TOP OF EACH BAR. THE HEIGHT OF THE BARS SHOWS THE PERCENTAGE OF THE TOTAL SIMULATION TIME IN WHICH THE CORRESPONDING CONTACTS ARE PRESENT. ONLY CONTACTS OF HIGHER THAN 10% OCCUPANCY ARE SHOWN. EACH SUBPLOT PRESENTS THE CONTACTS FOR EACH TRKA AND TRKB GLYCOSYLATED SYSTEM. 91

FIGURE 4.1: (A) SEQUENCES USED FOR THE SIMULATIONS OF THE TRANSMEMBRANE (TM) DOMAINS OF TRKA AND TRKB RECEPTORS. THE TM HELICAL PART FROM THE TRKA NMR STRUCTURE (PDB ID: 2N90)[60] IS INDICATED WITH THE DASHED LINES, WHILE THE N- AND C-TERMINAL LOOP SEQUENCES IN THE SOLVED NMR STRUCTURE ARE ALSO SHOWN. MUTATED RESIDUES IN THE NMR SEQUENCE COMPARED TO THE HUMAN TRKA ARE COLORED BLUE. RESIDUES PARTICIPATING IN THE GXXXG-LIKE MOTIFS ARE COLORED MAGENTA, WHILE THE TRKA NMR INTERFACE LXXFAXXF MOTIF IS HIGHLIGHTED IN YELLOW. (B) CARTOON REPRESENTATION OF THE TRKA NMR STRUCTURE (PDB ID: 2N90), WITH THE INTERFACE LXXFAXXF MOTIF SHOWN IN YELLOW LICORICE REPRESENTATION. (C) CARTOON REPRESENTATION OF THE TRKB-TM DIMER HOMOLOGY MODEL, WITH THE AXXXG MOTIF SHOWN IN MAGENTA LICORICE REPRESENTATION. 96

FIGURE 4.2: (A) EQUILIBRATED MICELLE FORMED BY DPC LIPIDS (CYAN) WITH THE TRKA TM HELICAL PROTEIN (PINK) IN THE MIDDLE IN CG REPRESENTATION. (B) SCHEMATIC REPRESENTATION OF THE COLLECTIVE VARIABLE (CV) USED IN THE CG-METAD SIMULATIONS. THE

distance between the centers of geometry (COG) of the backbone beads of the two helices was used as the biasing CV. In the figure, also the COGs of the N-terminal half of the helices are also shown, as they were later used for subsequent analysis. 101

Figure 4.3: Comparison of sampling in atomistic and coarse-grained simulations for the TrkA TM homodimer. I calculated the KDE plots for six different order parameters: interhelical distance (d), interhelical crossing angle (Ω), phase (rotation around 1ST helix), position (rotation around 2ND helix), N- and C- termini distances. The parameters are also visualized schematically above the plots. Right- and left-handed arrangements are indicated with (R and L). The NMR state is shown with a blue dot. In the coarse-grained simulations, I have visualized the dominant sampled state. 105

Figure 4.4: Comparison of sampling in atomistic and coarse-grained simulations for the TrkB TM homodimer. I calculated the KDE plots for six different order parameters: interhelical distance (d), interhelical crossing angle (Ω), phase (rotation around 1ST helix), position (rotation around 2ND helix), N- and C- termini distances. The parameters are also visualized schematically above the plots. Right- and left-handed arrangements are indicated with (R and L). The NMR state is shown with a blue dot. In the coarse-grained simulations, I have visualized the dominant sampled state. 106

Figure 4.5: TrkA TM NMR structure (PDB ID: 2N90) in CG representation level. The residues of the active and inactive interfaces are shown in green and red, respectively. The three GxxxG-like motifs are also indicated in the sequence. 108

Figure 4.6: Analysis of ten TrkA arrangements from the CG simulations in micelles. The maxima are numbered in the phase-position 2D heatmap and then analyzed in each subsequent box. In each box, I have calculated the percentage of the trajectory that was spent in this maximum, the contact map between the two helices and the residues that interact the most from each helix are underlined in the two sequences on the bottom. I have also visualized the arrangements by aligning all the frames to the 1ST helix and then displaying the backbone of the 1ST helix in white, the 2ND helix from the NMR structure in orange, and the 2ND helix from the simulation frames in purple dots. 112

Figure 4.7: Analysis of three TrkB arrangements from the CG simulations in micelles. The maxima are numbered in the phase-position 2D heatmap and then analyzed in each subsequent box. In each box, I have calculated the percentage of the trajectory that was spent in this maximum, the contact map between the two helices and the residues that interact the most from each helix are underlined in the two sequences on the bottom. I have also visualized the arrangements by aligning all the frames to the 1ST helix and then displaying the backbone of the 1ST helix in white, the 2ND helix from the NMR structure in orange, and the 2ND helix from the simulation frames in purple dots. 114

Figure 4.8: Convergence evaluation of CG-MetaD simulations on TrkA-TM dimer. (A) Time evolution of biasing CV (interhelical distance, d). (B) Time evolution of Gaussian height. (C) Time evolution of free energy along CV space. (D) Time evolution of free energy difference (ΔF) between bound state (d < 1.5 nm) and unbound state (d > 1.5 nm). (E) Free energy error as a function of simulation block sizes. 116

Figure 4.9: Time evolution of the six order parameters that describe the arrangements of the TrkA-TM helices during the CG-MetaD simulations. The values that correspond to the bound state (d < 1.5 nm) of the helices and those from all states (both bound and unbound) are plotted in the two rows of the figure. 117

Figure 4.10: Reweighting on TrkA CG-MetaD simulations. (A) Comparison of free energy profile on biasing CV, interhelical distance, calculated from the Gaussian metadynamics bias (MetaD bias) and from the reweighting procedure. (B) Free energy surface (FES) on crossing angle (Ω) – interhelical distance (d) space after reweighting. Isosurfaces are plotted every 3 kJ/mol. 118

11 List of Figures

FIGURE 4.11: FREE ENERGY SURFACES (FES) ON CROSSING ANGLE (Ω) – INTERHELICAL DISTANCE (D), PHASE-POSITION AND N- AND C-TERMINI DISTANCES AFTER REWEIGHTING. ISOSURFACES ARE PLOTTED EVERY 1 KJ/MOL. 119

FIGURE 4.12: CONVERGENCE EVALUATION OF CG-METAD SIMULATIONS ON TRKB-TM DIMER. (A) TIME EVOLUTION OF BIASING CV (INTERHELICAL DISTANCE, D). (B) TIME EVOLUTION OF GAUSSIAN HEIGHT. (C) TIME EVOLUTION OF FREE ENERGY ALONG CV SPACE. (D) TIME EVOLUTION OF FREE ENERGY DIFFERENCE (ΔF) BETWEEN BOUND STATE (D < 1.5 NM) AND UNBOUND STATE (D > 1.5 NM). (E) FREE ENERGY ERROR AS A FUNCTION OF SIMULATION BLOCK SIZES. 120

FIGURE 4.13: TIME EVOLUTION OF THE SIX ORDER PARAMETERS THAT DESCRIBE THE ARRANGEMENTS OF THE TRKB-TM HELICES DURING THE CG-METAD SIMULATIONS. THE VALUES THAT CORRESPOND TO THE BOUND STATE (D < 1.5 NM) OF THE HELICES AND THOSE FROM ALL STATES (BOTH BOUND AND UNBOUND) ARE PLOTTED IN THE TWO ROWS OF THE FIGURE 121

FIGURE 4.14: REWEIGHTING ON TRKB CG-METAD SIMULATIONS. (A) COMPARISON OF FREE ENERGY PROFILE ON BIASING CV, INTERHELICAL DISTANCE, CALCULATED FROM THE GAUSSIAN METADYNAMICS BIAS (METAD BIAS) AND FROM THE REWEIGHTING PROCEDURE. (B) FREE ENERGY SURFACE (FES) ON CROSSING ANGLE (Ω) – INTERHELICAL DISTANCE (D) SPACE AFTER REWEIGHTING. ISOSURFACES ARE PLOTTED EVERY 3 KJ/MOL. 122

FIGURE 4.15: FREE ENERGY SURFACES (FES) ON CROSSING ANGLE (Ω) – INTERHELICAL DISTANCE (D), PHASE-POSITION AND N- AND C-TERMINI DISTANCES AFTER REWEIGHTING. ISOSURFACES ARE PLOTTED EVERY 1 KJ/MOL. 123

FIGURE 5.1: PIE CHARTS WITH DISTRIBUTIONS OF HEADGROUPS AND NUMBERS OF UNSATURATED BONDS IN LIPID TAILS, FOR THE OUTER AND INNER LEAFLETS OF THE MODEL NEURONAL MEMBRANE. THE HEADGROUP TYPES ARE CHOLESTEROL, PHOSPHOCHOLINE (PC), PHOSPHOETHANOLAMINE (PE), SPHINGOMYELIN (SM), GLYCOLIPIDS (GLYCO) INCLUDING GM1, GM3, GALCER180, GALCER241, PHOSPHO-L-SERINE (PS), PHOSPHOINOSITOL (PI). THE RIGHT PLOTS INDICATE THE PERCENTAGES FOR DIFFERENT NUMBERS OF UNSATURATED BONDS IN EACH LIPID TAIL. 128

FIGURE 5.2: SYSTEMS WHOSE SIMULATION IS DESCRIBED IN THIS CHAPTER. (A) TM-IC TRKA/B MONOMER EMBEDDED IN NEURONAL MEMBRANE. (B) FULL-LENGTH TRKA/B DIMER BOUND TO NT (NGF AND NT-4/5) IN THE NEURONAL MEMBRANE. 132

FIGURE 5.3: TRKA KINASE – MEMBRANE INTERACTIONS DURING THE SIMULATIONS. UPPER PLOTS: CONTACT OCCUPANCY BETWEEN THE TRKA KINASE DOMAIN AND THE LIPIDS OF THE NEURONAL MEMBRANE AS FUNCTION OF TIME FOR THE THREE REPLICAS OF THE TM-IC SYSTEM. THE CONTACT OCCUPANCY DURING THE SIMULATIONS ARE NORMALIZED FROM 0 TO 1. LOWER FIGURE LEFT: SEQUENCE CONSERVATION OF THE THREE KINASE REGIONS THAT INTERACT WITH THE MEMBRANE. LOWER FIGURE RIGHT: TRKA KINASE DOMAIN WITH THE REGIONS THAT INTERACT WITH THE NEURONAL MEMBRANE COLORED IN MAGENTA. THE RESIDUE NUMBERS, AS WELL AS THE RESIDUE TYPES ARE INDICATED. 137

FIGURE 5.4: INTERACTING RESIDUES OF THE TRKA KINASE DOMAIN WITH THE MEMBRANE. UPPER FIGURE: ELECTROSTATIC ISOPOTENTIAL SURFACES AT +1 AND -1 CHARGE OF THE FIRST AND LAST FRAMES OF ONE OF THE TM-IC REPLICA SIMULATIONS. THE INCREASE IN THE POSITIVE CHARGE ON THE KINASE INTERACTION REGIONS AT THE END OF THE SIMULATION IS INDICATED WITH YELLOW BOXES. LOWER FIGURE: SIMULATION SNAPSHOT OF THE TRKA KINASE DOMAIN INTERACTING WITH THE NEURONAL MEMBRANE. INSETS SHOW INTERACTIONS BETWEEN THE POSITIVELY CHARGED RESIDUES OF THE KINASE AND THE NEGATIVELY CHARGED LIPIDS OF THE INNER LEAFLET. 138

FIGURE 5.5: NUMBER OF CONTACTS BETWEEN THE TM HELICES OF TRKA AND TRKB RECEPTORS AND CHOLESTEROL. A CONTACT IS DEFINED WHEN TWO ATOMS FROM A RESIDUE AND A CHOLESTEROL MOLECULE ARE WITHIN 4.5 Å. THE AVERAGE CONTACTS OVER THE VARIOUS REPLICAS AND THE STANDARD DEVIATION ARE PLOTTED IN THE FIRST BAR PLOTS. THE REST OF THE PLOTS ARE HEATMAPS OF THE NUMBER OF CONTACTS OVER SIMULATION TIME FOR EACH REPLICA. 139

FIGURE 5.6: SNAPSHOT FROM THE END OF REPLICA 9 OF TRKA FULL-LENGTH SYSTEM. THE RED CIRCLE INDICATES THE UNFOLDED REGION OF THE N-TERMINAL PART OF THE D13 DOMAIN. 140

FIGURE 5.7: NT-MEMBRANE DISTANCE AND EC DOMAIN-MEMBRANE NORMAL ANGLE CALCULATIONS FOR THE WHOLE LENGTH OF THE SIMULATIONS AND AFTER THE FIRST 150 NS OF THE FULL-LENGTH TRKA SYSTEMS. THE DISTRIBUTION DENSITIES OF THE DISTANCE AND ANGLE ARE ALSO PLOTTED PER REPLICA AND ALL TOGETHER. 142
FIGURE 5.8: NT-MEMBRANE DISTANCE AND EC DOMAIN-MEMBRANE NORMAL ANGLE CALCULATIONS FOR THE WHOLE LENGTH OF THE SIMULATIONS AND AFTER THE FIRST 150 NS OF THE FULL-LENGTH TRKB SYSTEMS. THE DISTRIBUTION DENSITIES OF THE DISTANCE AND ANGLE ARE ALSO PLOTTED PER REPLICA AND ALL TOGETHER. 143
FIGURE 5.9: HEATMAPS WITH THE PERCENTAGE OF INTER-HELIX CONTACTS OF THE TM DOMAIN OF TRKA DURING THE DIFFERENT REPLICA SIMULATIONS. THE RESIDUES THAT HAVE BEEN IDENTIFIED TO BE PART OF THE ACTIVE AND INACTIVE INTERFACE ARE INDICATED WITH GREEN AND RED COLUMNS AND ROWS, RESPECTIVELY. 144
FIGURE 5.10: HEATMAPS WITH THE PERCENTAGE OF INTER-HELIX CONTACTS OF THE TM DOMAIN OF TRKB DURING THE DIFFERENT REPLICA SIMULATIONS. THE AXXXG MOTIF, WHICH HAS BEEN SUGGESTED TO BE PART OF THE ACTIVE INTERFACE, IS INDICATED WITH GREEN COLUMNS AND ROWS. 145
FIGURE 5.11: NEURONAL MEMBRANE COMPOSITION PER LEAFLET. THE BARS SHOW THE RATIOS OF LIPID MOLECULES PER HEADGROUP TYPE. LIPID SPECIES AT HIGHER PERCENTAGE ARE MORE PROBABLE TO INTERACT WITH THE RECEPTORS. 146
FIGURE 5.12: BAR PLOTS WITH CONTACTS BETWEEN THE NEURONAL MEMBRANE LIPIDS AND THE DIFFERENT SEGMENTS OF TRKA RECEPTOR, AS WELL AS THE BOUND NT. THE RECEPTOR WAS SPLIT INTO: ECD: EXTRACELLULAR DOMAIN, ECL: EXTRACELLULAR LINKER, TMD: TRANSMEMBRANE DOMAIN, ICL: INTRACELLULAR LINKER, ICD: INTRACELLULAR DOMAIN. 149
FIGURE 5.13: BAR PLOTS WITH CONTACTS BETWEEN THE NEURONAL MEMBRANE LIPIDS AND THE DIFFERENT SEGMENTS OF TRKB RECEPTOR, AS WELL AS THE BOUND NT. THE RECEPTOR WAS SPLIT IN: ECD: EXTRACELLULAR DOMAIN, ECL: EXTRACELLULAR LINKER, TMD: TRANSMEMBRANE DOMAIN, ICL: INTRACELLULAR LINKER, ICD: INTRACELLULAR DOMAIN. 149
FIGURE 6.1: SUMMARY OF THE INITIAL CONFIGURATIONS OF THE SIMULATED SYSTEMS WITH THE NEUROTROPHIN MIMETICS. ONE FIGURE IS SHOWN FOR EACH SYSTEM, BUT THE SAME SYSTEMS WERE TESTED FOR BOTH TRKA AND TRKB RECEPTORS. (A) COMPOUNDS BOUND AT THE INTERFACE OF THE D5 DOMAIN OF TRKA AND TRKB RECEPTORS WITH THE NTS NGF AND NT-4/5, RESPECTIVELY. (B) ONE AND TEN COMPOUND MOLECULES PLACED CLOSE TO THE BILAYER FOR MEMBRANE PERMEABILITY SIMULATIONS SHOWN IN PERPENDICULAR VIEWS. (C) ONE OR TEN COMPOUNDS PLACED AT THE INTERFACE OR CLOSE TO THE TM HELICES OF TRKA AND TRKB RECEPTORS. THE P ATOMS OF THE POPC LIPIDS ARE SHOWN IN BLUE VDW REPRESENTATION. 161
FIGURE 6.2: DOCKING POSES OF TWO COMPOUNDS, ENT-A013E/Z AND ENT-A040, IN SITES 1A AND 1B AT THE TWO SYMMETRY-RELATED TRKA-D5 – NGF INTERFACES OF THE NEUROTROPHIN-RECEPTOR COMPLEX. THE CLOSE-UP VIEWS SHOW RESIDUES LINING THE TWO SITES. TRKA-D5 AND NGF ARE SHOWN IN GREEN AND BLUE NEW CARTOON REPRESENTATION, RESPECTIVELY, WITH SELECTED RESIDUES IN STICK REPRESENTATION COLORED BY ATOM TYPE. THE COMPOUNDS ARE SHOWN IN ORANGE STICK REPRESENTATION. THE FIGURE WAS MADE BY ALEXANDROS TSENGENES. 162
FIGURE 6.3: DOCKING POSES OF FOUR COMPOUNDS, ENT-A011, ENT-A044, ENT-A045 AND ENT-A061, IN SITES 1A AND 1B AT THE TWO SYMMETRY-RELATED TRKB-D5 – NT-4/5 INTERFACES OF THE NEUROTROPHIN-RECEPTOR COMPLEX. THE CLOSE-UP VIEWS SHOW RESIDUES LINING THE TWO SITES. TRKB-D5 AND NT-4/5 ARE SHOWN IN GREEN AND BLUE NEW CARTOON REPRESENTATION, RESPECTIVELY, WITH SELECTED RESIDUES IN STICK REPRESENTATION COLORED BY ATOM TYPE. THE COMPOUNDS ARE SHOWN IN ORANGE STICK REPRESENTATION. 163
FIGURE 6.4: MEMBRANE PERMEABILITY SIMULATIONS FOR ONE AND TEN MOLECULES PLACED CLOSE TO A MEMBRANE. IN THE FIGURE, AN EXAMPLE WITH COMPOUND ENT-A011 IS SHOWN. SPECIFICALLY, SIMULATION SNAPSHOTS ARE SHOWN WITH COMPOUND ENT-A011 ENTERING AT

11 List of Figures

CONSECUTIVE SIMULATION TIMES A POPC MEMBRANE AS A SINGLE MOLECULE (LEFT) AND AS AN AGGREGATE OF TEN MOLECULES (RIGHT). THE P ATOMS OF THE POPC LIPIDS ARE SHOWN IN BLUE VDW REPRESENTATION. 164

12 References

1. Sofroniew, M. V., Howe, C. L. & Mobley, W. C. Nerve growth factor signaling, neuroprotection, and neural repair. *Annu. Rev. Neurosci.* **24**, 1217–1281 (2001).
2. Vogt Weisenhorn, D. M., Roback, J., Young, A. N. & Wainer, B. H. Cellular Aspects of Trophic Actions in the Nervous System. in 177–265 (1999). doi:10.1016/S0074-7696(08)61388-1
3. Poon, V. Y., Choi, S. & Park, M. Growth factors in synaptic function. *Front. Synaptic Neurosci.* **5**, (2013).
4. Park, H. & Poo, M. Neurotrophin regulation of neural circuit development and function. *Nat. Rev. Neurosci.* **14**, 7–23 (2013).
5. Butte, M. J. Neurotrophic factor structures reveal clues to evolution, binding, specificity, and receptor activation. *Cell. Mol. Life Sci.* **58**, 1003–1013 (2001).
6. Levi-Montalcini, R. & Hamburger, V. A diffusible agent of mouse sarcoma, producing hyperplasia of sympathetic ganglia and hyperneurotization of viscera in the chick embryo. *J. Exp. Zool.* **123**, 233–287 (1953).
7. LEVI-MONTALCINI, R. & HAMBURGER, V. Selective growth stimulating effects of mouse sarcoma on the sensory and sympathetic nervous system of the chick embryo. *J. Exp. Zool.* **116**, 321–61 (1951).
8. Huang, E. J. & Reichardt, L. F. Trk Receptors: Roles in Neuronal Signal Transduction. *Annu. Rev. Biochem.* **72**, 609–642 (2003).
9. Bruno, M. A. & Cuello, A. C. Activity-dependent release of precursor nerve growth factor, conversion to mature nerve growth factor, and its degradation by a protease cascade. *Proc. Natl. Acad. Sci.* **103**, 6735–6740 (2006).
10. Lee, R., Kermani, P., Teng, K. K. & Hempstead, B. L. Regulation of Cell Survival by Secreted Proneurotrophins. *Science (80-.).* **294**, 1945–1948 (2001).
11. Nykjaer, A. *et al.* Sortilin is essential for proNGF-induced neuronal cell death. *Nature* **427**, 843–848 (2004).
12. Glerup, S. *et al.* SorCS2 Regulates Dopaminergic Wiring and Is Processed into an Apoptotic Two-Chain Receptor in Peripheral Glia. *Neuron* **82**, 1074–1087 (2014).
13. Dicou, E. Multiple biological activities for two peptides derived from the nerve growth factor precursor. *Biochem. Biophys. Res. Commun.* **347**, 833–837 (2006).
14. Roux, P. Neurotrophin signaling through the p75 neurotrophin receptor. *Prog. Neurobiol.* **67**, 203–233 (2002).
15. Chao, M. V. Neurotrophins and their receptors: A convergence point for many signalling pathways. *Nat. Rev. Neurosci.* **4**, 299–309 (2003).
16. Kaplan, D. R., Hempstead, B. L., Martin-Zanca, D., Chao, M. V. & Parada, L. F. The trk Proto-Oncogene Product: a Signal Transducing Receptor for Nerve Growth Factor. *Science (80-.).* **252**, 554–558 (1991).
17. Rodríguez-Tébar, A., Dechant, G., Götz, R. & Barde, Y. A. Binding of neurotrophin-3 to its neuronal receptors and interactions with nerve growth factor and brain-derived neurotrophic factor. *EMBO J.* **11**, 917–922 (1992).
18. Landreth, G. E. & Shooter, E. M. Nerve growth factor receptors on PC12 cells: ligand-induced conversion from low- to high-affinity states. *Proc. Natl. Acad. Sci.* **77**, 4751–4755 (1980).
19. Bothwell, M. Functional interactions of neurotrophins and neurotrophin receptors. *Annu. Rev. Neurosci.* **18**, 223–253 (1995).

20. McDonald, N. Q. *et al.* New protein fold revealed by a 2.3-Å resolution crystal structure of nerve growth factor. *Nature* **354**, 411–414 (1991).

21. Fandl, J. P. *et al.* Characterization and crystallization of recombinant human neurotrophin-4. *J. Biol. Chem.* **269**, 755–9 (1994).

22. Butte, M. J., Hwang, P. K., Mobley, W. C. & Fletterick, R. J. Crystal Structure of Neurotrophin-3 Homodimer Shows Distinct Regions Are Used To Bind Its Receptors ,. *Biochemistry* **37**, 16846–16852 (1998).

23. Robinson, R. C. *et al.* The structures of the neurotrophin 4 homodimer and the brain-derived neurotrophic factor/neurotrophin 4 heterodimer reveal a common Trk-binding site. *Protein Sci.* **8**, 2589–2597 (2008).

24. Robinson, R. C., Radziejewski, C., Stuart, D. I. & Jones, E. Y. Structure of the Brain-Derived Neurotrophic Factor/Neurotrophin 3 Heterodimer. *Biochemistry* **34**, 4139–4146 (1995).

25. Paoletti, F. *et al.* Conformational Rigidity within Plasticity Promotes Differential Target Recognition of Nerve Growth Factor. *Front. Mol. Biosci.* **3**, (2016).

26. Conroy, J. N. & Coulson, E. J. High-affinity TrkA and p75 neurotrophin receptor complexes: A twisted affair. *J. Biol. Chem.* **298**, 101568 (2022).

27. Ivanisevic, L., Zheng, W., Woo, S. B., Neet, K. E. & Saragovi, H. U. TrkA Receptor "Hot Spots" for Binding of NT-3 as a Heterologous Ligand. *J. Biol. Chem.* **282**, 16754–16763 (2007).

28. Wehrman, T. *et al.* Structural and Mechanistic Insights into Nerve Growth Factor Interactions with the TrkA and p75 Receptors. *Neuron* **53**, 25–38 (2007).

29. Wiesmann, C., Ultsch, M. H., Bass, S. H. & de Vos, A. M. Crystal structure of nerve growth factor in complex with the ligand-binding domain of the TrkA receptor. *Nature* **401**, 184–188 (1999).

30. Ultsch, M. H. *et al.* Crystal structures of the neurotrophin-binding domain of TrkA, TrkB and TrkC. *J. Mol. Biol.* **290**, 149–159 (1999).

31. Windisch, J. M., Marksteiner, R. & Schneider, R. Nerve Growth Factor Binding Site on TrkA Mapped to a Single 24-Amino Acid Leucine-rich Motif. *J. Biol. Chem.* **270**, 28133–28138 (1995).

32. Windisch, J. M., Marksteiner, R., Lang, M. E., Auer, B. & Schneider, R. Brain-derived neurotrophic factor, neurotrophin-3, and neurotrophin-4 bind to a single leucine-rich motif of TrkB. *Biochemistry* **34**, 11256–11263 (1995).

33. Kaplan, D. R. & Miller, F. D. Signal transduction by the neutrophin receptors. *Curr. Opin. Cell Biol.* **9**, 213–221 (1997).

34. Cabrera, N., Díaz-Rodríguez, E., Becker, E., Martín-Zanca, D. & Pandiella, A. TrkA receptor ectodomain cleavage generates a tyrosine-phosphorylated cell-associated fragment. *J. Cell Biol.* **132**, 427–436 (1996).

35. Arevalo, J. C. *et al.* TrkA Immunoglobulin-Like Ligand Binding Domains Inhibit Spontaneous Activation of the Receptor. *Mol. Cell. Biol.* **20**, 5908–5916 (2000).

36. Strohmaier, C., Carter, B. D., Urfer, R., Barde, Y. A. & Dechant, G. A splice variant of the neurotrophin receptor trkB with increased specificity for brain-derived neurotrophic factor. *EMBO J.* **15**, 3332–7 (1996).

37. Clary, D. O. & Reichardt, L. F. An alternatively spliced form of the nerve growth factor receptor TrkA confers an enhanced response to neurotrophin 3. *Proc. Natl. Acad. Sci.* **91**, 11133–11137 (1994).

38. Baldwin, A. N., Bitler, C. M., Welcher, A. A. & Shooter, E. M. Studies on the structure and binding properties of the cysteine-rich domain of rat low affinity nerve growth factor receptor (p75NGFR). *J. Biol. Chem.* **267**, 8352–8359 (1992).

39. Nykjaer, A., Willnow, T. E. & Petersen, C. M. p75NTR--live or let die. *Curr. Opin. Neurobiol.* **15**, 49–57 (2005).

40. Bibel, M. Biochemical and functional interactions between the neurotrophin receptors trk and p75NTR. *EMBO J.* **18**, 616–622 (1999).

41. Makkerh, J. P. S. *et al.* p75 neurotrophin receptor reduces ligand-induced Trk receptor ubiquitination and delays Trk receptor internalization and degradation. *EMBO Rep.* **6**, 936–941 (2005).

42. Varki A, Cummings RD, Esko JD, Stanley P, Hart GW, Aebi M, Mohnen D, Kinoshita T, Packer NH, Prestegard JH, Schnaar RL, Seeberger PH, E. *Essentials of Glycobiology*. (Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press, 2022). doi:10.1101/9781621824213

43. Settineri, C. A. *et al.* Characterization of the Glycosylation on Recombinant Human Low-Affinity Nerve Growth Factor Receptor. in *Techniques in Protein Chemistry III* 295–301 (Elsevier, 1992). doi:10.1016/B978-0-12-058756-8.50035-3

44. Haniu, M. *et al.* Interactions between Brain-derived Neurotrophic Factor and the TRKB Receptor. *J. Biol. Chem.* **272**, 25296–25303 (1997).

45. Yeaman, C. *et al.* The O-glycosylated Stalk Domain Is Required for Apical Sorting of Neurotrophin Receptors in Polarized MDCK Cells. *J. Cell Biol.* **139**, 929–940 (1997).

46. Monlauzeur, L., Breuza, L. & Le Bivic, A. Putative O-Glycosylation Sites and a Membrane Anchor Are Necessary for Apical Delivery of the Human Neurotrophin Receptor in Caco-2 Cells. *J. Biol. Chem.* **273**, 30263–30270 (1998).

47. Breuza, L., Garcia, M., Delgrossi, M.-H. & Bivic, A. Le. Role of the Membrane-Proximal O-glycosylation Site in Sorting of the Human Receptor for Neurotrophins to the Apical Membrane of MDCK Cells. *Exp. Cell Res.* **273**, 178–186 (2002).

48. Haniu, M. *et al.* Extracellular Domain of Neurotrophin Receptor trkB: Disulfide Structure, N-Glycosylation Sites, and Ligand Binding. *Arch. Biochem. Biophys.* **322**, 256–264 (1995).

49. Baldwin, A. N. & Shooter, E. M. Zone Mapping of the Binding Domain of the Rat Low Affinity Nerve Growth Factor Receptor by the Introduction of Novel N-Glycosylation Sites. *J. Biol. Chem.* **270**, 4594–4602 (1995).

50. Watson, F. L., Porcionatto, M. A., Bhattacharyya, A., Stiles, C. D. & Segal, R. A. TrkA glycosylation regulates receptor localization and activity. *J. Neurobiol.* **39**, 323–336 (1999).

51. Grob, P. M., Ross, A. H., Koprowski, H. & Bothwell, M. Characterization of the human melanoma nerve growth factor receptor. *J. Biol. Chem.* **260**, 8044–9 (1985).

52. Trabjerg, E., Kartberg, F., Christensen, S. & Rand, K. D. Conformational characterization of nerve growth factor-β reveals that its regulatory pro-part domain stabilizes three loop regions in its mature part. *J. Biol. Chem.* **292**, 16665–16676 (2017).

53. Benicky, J., Sanda, M., Brnakova Kennedy, Z. & Goldman, R. N-Glycosylation is required for secretion of the precursor to brain-derived neurotrophic factor (proBDNF) carrying sulfated LacdiNAc structures. *J. Biol. Chem.* **294**, 16816–16830 (2019).

54. Jullien, J., Guili, V., Reichardt, L. F. & Rudkin, B. B. Molecular Kinetics of Nerve Growth Factor Receptor Trafficking and Activation. *J. Biol. Chem.* **277**, 38700–38708 (2002).

55. Woronowicz, A. *et al.* Dependence of neurotrophic factor activation of Trk tyrosine kinase receptors on cellular sialidase. *Glycobiology* **17**, 10–24 (2007).

56. Chapman, B. S., Eckart, M. R., Kaufman, S. E. & Lapointe, G. R. O-Linked Oligosaccharide on the 75-kDa Neurotrophin Receptor. *J. Neurochem.* **66**, 1707–1716 (2002).

57. Settanni, G., Cattaneo, A. & Carloni, P. Molecular Dynamics Simulations of the NGF-TrkA Domain 5 Complex and Comparison with Biological Data. *Biophys. J.* **84**, 2282–2292 (2003).

58. Zaccaro, M. C., Ivanisevic, L., Perez, P., Meakin, S. O. & Saragovi, H. U. p75 Co-receptors Regulate Ligand-dependent and Ligand-independent Trk Receptor Activation, in Part by Altering Trk Docking Subdomains. *J. Biol. Chem.* **276**, 31023–31029 (2001).

59. Banfield, M. J. *et al.* Specificity in Trk Receptor:Neurotrophin Interactions: The Crystal Structure of Trkb-D5 in Complex with Neurotrophin-4/5. *Structure* **9**, 1191–1199 (2001).

60. Franco, M. L. *et al.* Structural basis of the transmembrane domain dimerization and rotation in the activation mechanism of the TRKA receptor by nerve growth factor. *J. Biol. Chem.* **295**, 275–286 (2020).

61. Shen, J. *et al.* Extracellular Juxtamembrane Motif Critical for TrkB Preformed Dimer and Activation. *Cells* **8**, 932 (2019).

62. Shen, J. & Maruyama, I. N. Brain-derived neurotrophic factor receptor TrkB exists as a preformed dimer in living cells. *J. Mol. Signal.* **7**, 2 (2012).

63. Shen, J. & Maruyama, I. N. Nerve growth factor receptor TrkA exists as a preformed, yet inactive, dimer in living cells. *FEBS Lett.* **585**, 295–299 (2011).

64. Hubbard, S. R. & Till, J. H. Protein Tyrosine Kinase Structure and Function. *Annu. Rev. Biochem.* **69**, 373–398 (2000).

65. Jiang, G. & Hunter, T. Receptor signaling: When dimerization is not enough. *Curr. Biol.* **9**, R568–R571 (1999).

66. Huse, M. & Kuriyan, J. The Conformational Plasticity of Protein Kinases. *Cell* **109**, 275–282 (2002).

67. Cunningham, M. E. A function-structure model for NGF-activated TRK. *EMBO J.* **17**, 7282–7293 (1998).

68. Mischel, P. S. *et al.* Nerve Growth Factor Signals via Preexisting TrkA Receptor Oligomers. *Biophys. J.* **83**, 968–976 (2002).

69. Obermeier, A. *et al.* Neuronal differentiation signals are controlled by nerve growth factor receptor/Trk binding sites for SHC and PLC gamma. *EMBO J.* **13**, 1585–1590 (1994).

70. Furuya, N. *et al.* The juxtamembrane region of TrkA kinase is critical for inhibitor selectivity. *Bioorg. Med. Chem. Lett.* **27**, 1233–1236 (2017).

71. Wybenga-Groot, L. E. *et al.* Structural Basis for Autoinhibition of the EphB2 Receptor Tyrosine Kinase by the Unphosphorylated Juxtamembrane Region. *Cell* **106**, 745–757 (2001).

72. Bertrand, T. *et al.* The Crystal Structures of TrkA and TrkB Suggest Key Regions for Achieving Selective Inhibition. *J. Mol. Biol.* **423**, 439–453 (2012).

73. Hanks, S. K., Quinn, A. M. & Hunter, T. The Protein Kinase Family: Conserved Features and Deduced Phylogeny of the Catalytic Domains. *Science (80-.).* **241**, 42–52 (1988).

74. Ahmed, F. & Hristova, K. Dimerization of the Trk receptors in the plasma membrane: effects of their cognate ligands. *Biochem. J.* **475**, 3669–3685 (2018).

75. Lemmon, M. A. & Schlessinger, J. Cell Signaling by Receptor Tyrosine Kinases. *Cell* **141**, 1117–1134 (2010).

76. Davies, A. M., Lee, K.-F. & Jaenisch, R. p75-Deficient trigeminal sensory neurons have an altered response to NGF but not to other neurotrophins. *Neuron* **11**, 565–574 (1993).

77. Mahadeo, D., Kaplan, L., Chao, M. V & Hempstead, B. L. High affinity nerve growth factor binding displays a faster rate of association than p140trk binding. Implications for multi-subunit polypeptide receptors. *J. Biol. Chem.* **269**, 6884–91 (1994).

78. Segal, R. A. *et al.* Differential Utilization of Trk Autophosphorylation Sites. *J. Biol. Chem.* **271**, 20175–20181 (1996).

79. Obermeier, A. *et al.* Tyrosine 785 is a major determinant of Trk-substrate interaction. *EMBO J.* **12**, 933–941 (1993).

80. Obermeier, A. *et al.* Identification of Trk binding sites for SHC and phosphatidylinositol 3'-kinase and formation of a multimeric signaling complex. *J. Biol. Chem.* **268**, 22963–6 (1993).

81. Bell, C. A. *et al.* Rotational Coupling of the Transmembrane and Kinase Domains of the Neu Receptor

Tyrosine Kinase. *Mol. Biol. Cell* **11**, 3589–3599 (2000).

82. Brown, R. J. *et al.* Model for growth hormone receptor activation based on subunit rotation within a receptor dimer. *Nat. Struct. Mol. Biol.* **12**, 814–821 (2005).

83. Maruyama, I. N. Activation of transmembrane cell-surface receptors via a common mechanism? The "rotation model". *BioEssays* **37**, 959–967 (2015).

84. Moriki, T., Maruyama, H. & Maruyama, I. N. Activation of preformed EGF receptor dimers by ligand-induced rotation of the transmembrane domain. *J. Mol. Biol.* **311**, 1011–1026 (2001).

85. Maruyama, I. Mechanisms of Activation of Receptor Tyrosine Kinases: Monomers or Dimers. *Cells* **3**, 304–330 (2014).

86. Park, Y. *et al.* Single-Molecule Rotation for EGFR Conformational Dynamics in Live Cells. *J. Am. Chem. Soc.* **140**, 15161–15165 (2018).

87. Gautier, R., Douguet, D., Antonny, B. & Drin, G. HELIQUEST: a web server to screen sequences with specific -helical properties. *Bioinformatics* **24**, 2101–2102 (2008).

88. Russ, W. P. & Engelman, D. M. The GxxxG motif: A framework for transmembrane helix-helix association. *J. Mol. Biol.* **296**, 911–919 (2000).

89. Walters, R. F. S. & DeGrado, W. F. Helix-packing motifs in membrane proteins. *Proc. Natl. Acad. Sci.* **103**, 13658–13663 (2006).

90. Moore, D. T., Berger, B. W. & DeGrado, W. F. Protein-Protein Interactions in the Membrane: Sequence, Structural, and Biological Motifs. *Structure* **16**, 991–1001 (2008).

91. Casarotto, P. C. *et al.* Antidepressant drugs act by directly binding to TRKB neurotrophin receptors. *Cell* **184**, 1299-1313.e19 (2021).

92. Matusica, D. & Coulson, E. J. Local versus long-range neurotrophin receptor signalling: Endosomes are not just carriers for axonal transport. *Semin. Cell Dev. Biol.* **31**, 57–63 (2014).

93. Lad, S. P., Peterson, D. A., Bradshaw, R. A. & Neet, K. E. Individual and Combined Effects of TrkA and p75NTR Nerve Growth Factor Receptors. *J. Biol. Chem.* **278**, 24808–24817 (2003).

94. Teng, K. K. & Hempstead, B. L. Neurotrophins and their receptors: signaling trios in complex biological systems. *Cell. Mol. Life Sci.* **61**, 35–48 (2004).

95. Chao, M. V & Hempstead, B. L. p75 and Trk: a two-receptor system. *Trends Neurosci.* **18**, 321–6 (1995).

96. He, X. & Garcia, K. C. Structure of Nerve Growth Factor Complexed with the Shared Neurotrophin Receptor p75. *Science (80-.).* **304**, 870–875 (2004).

97. Wiesmann, C. & de Vos, A. M. Nerve growth factor: structure and function. *Cell. Mol. Life Sci.* **58**, 748–759 (2001).

98. Barker, P. A. High Affinity Not in the Vicinity? *Neuron* **53**, 1–4 (2007).

99. Hempstead, B. L., Patil, N., Thiel, B. & Chao, M. V. Deletion of cytoplasmic sequences of the nerve growth factor receptor leads to loss of high affinity ligand binding. *J. Biol. Chem.* **265**, 9595–9598 (1990).

100. Esposito, D. *et al.* The Cytoplasmic and Transmembrane Domains of the p75 and Trk A Receptors Regulate High Affinity Binding to Nerve Growth Factor. *J. Biol. Chem.* **276**, 32687–32695 (2001).

101. Hempstead, B. L., Martin-Zanca, D., Kaplan, D. R., Parada, L. F. & Chao, M. V. High-affinity NGF binding requires coexpression of the trk proto-oncogene and the low-affinity NGF receptor. *Nature* **350**, 678–683 (1991).

102. Rydén, M., Hempstead, B. & Ibáñez, C. F. Differential Modulation of Neuron Survival during Development by Nerve Growth Factor Binding to the p75 Neurotrophin Receptor. *J. Biol. Chem.* **272**, 16322–16328

(1997).

103. Horton, A. *et al.* NGF Binding to p75 Enhances the Sensitivity of Sensory and Sympathetic Neurons to NGF at Different Stages of Development. *Mol. Cell. Neurosci.* **10**, 162–172 (1997).

104. Matusica, D. *et al.* An Intracellular Domain Fragment of the p75 Neurotrophin Receptor (p75NTR) Enhances Tropomyosin Receptor Kinase A (TrkA) Receptor Function. *J. Biol. Chem.* **288**, 11144–11154 (2013).

105. Berg, M. M., Sternberg, D. W., Hempstead, B. L. & Chao, M. V. The low-affinity p75 nerve growth factor (NGF) receptor mediates NGF-induced tyrosine phosphorylation. *Proc. Natl. Acad. Sci.* **88**, 7106–7110 (1991).

106. Benedetti, M., Levi, A. & Chao, M. V. Differential expression of nerve growth factor receptors leads to altered binding affinity and neurotrophin responsiveness. *Proc. Natl. Acad. Sci.* **90**, 7859–7863 (1993).

107. Egawa, J., Pearn, M. L., Lemkuil, B. P., Patel, P. M. & Head, B. P. Membrane lipid rafts and neurobiology: age-related changes in membrane lipids and loss of neuronal function. *J. Physiol.* **594**, 4565–4579 (2016).

108. Lorent, J. H. & Levental, I. Structural determinants of protein partitioning into ordered membrane domains and lipid rafts. *Chem. Phys. Lipids* **192**, 23–32 (2015).

109. Sezgin, E., Levental, I., Mayor, S. & Eggeling, C. The mystery of membrane organization: composition, regulation and roles of lipid rafts. *Nat. Rev. Mol. Cell Biol.* **18**, 361–374 (2017).

110. Levental, I., Levental, K. R. & Heberle, F. A. Lipid Rafts: Controversies Resolved, Mysteries Remain. *Trends Cell Biol.* **30**, 341–353 (2020).

111. Moll, T. *et al.* Membrane lipid raft homeostasis is directly linked to neurodegeneration. *Essays Biochem.* **65**, 999–1011 (2021).

112. Limpert, A. S., Karlo, J. C. & Landreth, G. E. Nerve Growth Factor Stimulates the Concentration of TrkA within Lipid Rafts and Extracellular Signal-Regulated Kinase Activation through c-Cbl-Associated Protein. *Mol. Cell. Biol.* **27**, 5686–5698 (2007).

113. Spencer, A. *et al.* Nerve Growth Factor Signaling from Membrane Microdomains to the Nucleus: Differential Regulation by Caveolins. *Int. J. Mol. Sci.* **18**, 693 (2017).

114. Suzuki, S. *et al.* BDNF-induced recruitment of TrkB receptor into neuronal lipid rafts. *J. Cell Biol.* **167**, 1205–1215 (2004).

115. Sharma, D. *et al.* Hypoxia-mediated alteration in cholesterol oxidation and raft dynamics regulates BDNF signalling and neurodegeneration in hippocampus. *J. Neurochem.* **148**, 238–251 (2019).

116. Higuchi, H. PKA phosphorylates the p75 receptor and regulates its localization to lipid rafts. *EMBO J.* **22**, 1790–1800 (2003).

117. Gil, C., Cubí, R. & Aguilera, J. Shedding of the p75 NTR neurotrophin receptor is modulated by lipid rafts. *FEBS Lett.* **581**, 1851–1858 (2007).

118. Zonta, B. & Minichiello, L. Synaptic membrane rafts: traffic lights for local neurotrophin signaling? *Front. Synaptic Neurosci.* **5**, (2013).

119. Duchemin, A.-M., Ren, Q., Mo, L., Neff, N. H. & Hadjiconstantinou, M. GM1 ganglioside induces phosphorylation and activation of Trk and Erk in brain. *J. Neurochem.* **81**, 696–707 (2002).

120. Kimura, M., Hidari, K. I.-P. J., Suzuki, T., Miyamoto, D. & Suzuki, Y. Engagement of endogenous ganglioside GM1a induces tyrosine phosphorylation involved in neuron-like differentiation of PC12 cells. *Glycobiology* **11**, 335–343 (2001).

121. Rabin, S. J. & Mocchetti, I. GM1 Ganglioside Activates the High-Affinity Nerve Growth Factor Receptor trkA. *J. Neurochem.* **65**, 347–354 (2002).

122. Mutoh, T., Tokuda, A., Miyadai, T., Hamaguchi, M. & Fujiki, N. Ganglioside GM1 binds to the Trk protein

and regulates receptor function. *Proc. Natl. Acad. Sci.* **92**, 5087–5091 (1995).

123. Mutoh, T., Hamano, T., Tokuda, A. & Kuriyama, M. Unglycosylated Trk protein does not co-localize nor associate with ganglioside GM1 in stable clone of PC12 cells overexpressing Trk (PCtrk cells). *Glycoconj. J.* **17**, 233–7

124. Mutoh, T. *et al.* Stable transfection of GM1 synthase gene into GM1-deficient NG108-15 cells, CR-72 cells, rescues the responsiveness of Trk-neurotrophin receptor to its ligand, NGF. *Neurochem. Res.* **27**, 801–6 (2002).

125. Chiricozzi, E. *et al.* Role of the GM1 ganglioside oligosaccharide portion in the TrkA-dependent neurite sprouting in neuroblastoma cells. *J. Neurochem.* **143**, 645–659 (2017).

126. Bachis, A., Rabin, S. J., Fiacco, M. & Mocchetti, I. Gangliosides prevent excitotoxicity through activation of TrkB receptor. *Neurotox. Res.* **4**, 225–234 (2002).

127. Rabin, S. J., Bachis, A. & Mocchetti, I. Gangliosides Activate Trk Receptors by Inducing the Release of Neurotrophins. *J. Biol. Chem.* **277**, 49466–49472 (2002).

128. MacPhee, I. & Barker, P. A. Extended Ceramide Exposure Activates the trkA Receptor by Increasing Receptor Homodimer Formation. *J. Neurochem.* **72**, 1423–1430 (2001).

129. Muragaki, Y. *et al.* Expression of trk receptors in the developing and adult human central and peripheral nervous system. *J. Comp. Neurol.* **356**, 387–397 (1995).

130. Wetts, R. & Vaughn, J. E. Peripheral and Central Target Requirements for Survival of Embryonic Rat Dorsal Root Ganglion Neurons in Slice Cultures. *J. Neurosci.* **18**, 6905–6913 (1998).

131. Longo, F. M. & Massa, S. M. Small-molecule modulation of neurotrophin receptors: a strategy for the treatment of neurological disease. *Nat. Rev. Drug Discov.* **12**, 507–525 (2013).

132. Vetter, M. L., Martin-Zanca, D., Parada, L. F., Bishop, J. M. & Kaplan, D. R. Nerve growth factor rapidly stimulates tyrosine phosphorylation of phospholipase C-gamma 1 by a kinase activity associated with the product of the trk protooncogene. *Proc. Natl. Acad. Sci.* **88**, 5650–5654 (1991).

133. Arévalo, J. C. & Wu, S. H. Neurotrophin signaling: many exciting surprises! *Cell. Mol. Life Sci.* **63**, 1523–1537 (2006).

134. Kaplan, D. R. & Miller, F. D. Neurotrophin signal transduction in the nervous system. *Curr. Opin. Neurobiol.* **10**, 381–391 (2000).

135. Xing, J., Kornhauser, J. M., Xia, Z., Thiele, E. A. & Greenberg, M. E. Nerve Growth Factor Activates Extracellular Signal-Regulated Kinase and p38 Mitogen-Activated Protein Kinase Pathways To Stimulate CREB Serine 133 Phosphorylation. *Mol. Cell. Biol.* **18**, 1946–1955 (1998).

136. Arévalo, J. C., Yano, H., Teng, K. K. & Chao, M. V. A unique pathway for sustained neurotrophin signaling through an ankyrin-rich membrane-spanning protein. *EMBO J.* **23**, 2358–2368 (2004).

137. Yamada, M. *et al.* Insulin Receptor Substrate (IRS)-1 and IRS-2 Are Tyrosine-phosphorylated and Associated with Phosphatidylinositol 3-Kinase in Response to Brain-derived Neurotrophic Factor in Cultured Cerebral Cortical Neurons. *J. Biol. Chem.* **272**, 30334–30339 (1997).

138. Vanhaesebroeck, B. & Alessi, D. R. The PI3K-PDK1 connection: more than just a road to PKB. *Biochem. J.* **346 Pt 3**, 561–76 (2000).

139. Bellacosa, A., Testa, Staal, S. & Tsichlis, P. A retroviral oncogene, akt, encoding a serine-threonine kinase containing an SH2-like region. *Science (80-.).* **254**, 274–277 (1991).

140. Kane, L. P., Shapiro, V. S., Stokoe, D. & Weiss, A. Induction of NF-κB by the Akt/PKB kinase. *Curr. Biol.* **9**, 601-S1 (1999).

141. Biggs, W. H., Meisenhelder, J., Hunter, T., Cavenee, W. K. & Arden, K. C. Protein kinase B/Akt-mediated

phosphorylation promotes nuclear exclusion of the winged helix transcription factor FKHR1. *Proc. Natl. Acad. Sci.* **96**, 7421–7426 (1999).

142. Chuang, H. *et al.* Bradykinin and nerve growth factor release the capsaicin receptor from PtdIns(4,5)P2-mediated inhibition. *Nature* **411**, 957–962 (2001).

143. Minichiello, L. *et al.* Essential Role for TrkB Receptors in Hippocampus-Mediated Learning. *Neuron* **24**, 401–414 (1999).

144. Lewin, G. R. & Barde, Y.-A. Physiology of the Neurotrophins. *Annu. Rev. Neurosci.* **19**, 289–317 (1996).

145. Bibel, M. & Barde, Y.-A. Neurotrophins: key regulators of cell fate and cell shape in the vertebrate nervous system. *Genes Dev.* **14**, 2919–2937 (2000).

146. Hempstead, B. Dissecting the Diverse Actions of Pro- and Mature Neurotrophins. *Curr. Alzheimer Res.* **3**, 19–24 (2006).

147. Korte, M. *et al.* Hippocampal long-term potentiation is impaired in mice lacking brain-derived neurotrophic factor. *Proc. Natl. Acad. Sci.* **92**, 8856–8860 (1995).

148. Chen, K. S. *et al.* Disruption of a Single Allele of the Nerve Growth Factor Gene Results in Atrophy of Basal Forebrain Cholinergic Neurons and Memory Deficits. *J. Neurosci.* **17**, 7288–7296 (1997).

149. Capsoni, S. *et al.* Alzheimer-like neurodegeneration in aged antinerve growth factor transgenic mice. *Proc. Natl. Acad. Sci.* **97**, 6826–6831 (2000).

150. Calissano, P., Matrone, C. & Amadoro, G. Nerve growth factor as a paradigm of neurotrophins related to Alzheimer's disease. *Dev. Neurobiol.* NA-NA (2010). doi:10.1002/dneu.20759

151. Venero, J. L., Knüsel, B., Beck, K. D. & Hefti, F. Expression of neurotrophin and trk receptor genes in adult rats with fimbria transections: Effect of intraventricular nerve growth factor and brain-derived neurotrophic factor administration. *Neuroscience* **59**, 797–815 (1994).

152. Cooper, J. D., Lindholm, D. & Sofroniew, M. V. Reduced transport of [125I]nerve growth factor by cholinergic neurons and down-regulated trka expression in the medial septum of aged rats. *Neuroscience* **62**, 625–629 (1994).

153. Hefti, F. Nerve growth factor promotes survival of septal cholinergic neurons after fimbrial transections. *J. Neurosci.* **6**, 2155–2162 (1986).

154. Finch, C. E. Neuron atrophy during aging: programmed or sporadic? *Trends Neurosci.* **16**, 104–110 (1993).

155. Fischer, W., Chen, K. S., Gage, F. H. & Björklund, A. Progressive decline in spatial learning and integrity of forebrain cholinergic neurons in rats during aging. *Neurobiol. Aging* **13**, 9–23 (1992).

156. Fischer, W. *et al.* Amelioration of cholinergic neuron atrophy and spatial memory impairment in aged rats by nerve growth factor. *Nature* **329**, 65–68 (1987).

157. Hock, C., Heese, K., Hulette, C., Rosenberg, C. & Otten, U. Region-Specific Neurotrophin Imbalances in Alzheimer Disease. *Arch. Neurol.* **57**, 846 (2000).

158. de Lacalle, S., Cooper, J. ., Svendsen, C. ., Dunnett, S. . & Sofroniew, M. . Reduced retrograde labelling with fluorescent tracer accompanies neuronal atrophy of basal forebrain cholinergic neurons in aged rats. *Neuroscience* **75**, 19–27 (1996).

159. Scott, S., Mufson, E., Weingartner, J., Skau, K. & Crutcher, K. Nerve growth factor in Alzheimer's disease: increased levels throughout the brain coupled with declines in nucleus basalis. *J. Neurosci.* **15**, 6213–6221 (1995).

160. Sofroniew, M. V. & Mobley, W. C. On the possibility of positive-feedback in trophic interactions between afferent and target neurons. *Semin. Neurosci.* **5**, 309–312 (1993).

161. Murer, M. . *et al.* An immunohistochemical study of the distribution of brain-derived neurotrophic factor

in the adult human brain, with particular reference to Alzheimer's disease. *Neuroscience* **88**, 1015–1032 (1999).

162. Narisawa-Saito, M., Wakabayashi, K., Tsuji, S., Takahashi, H. & Nawa, H. Regional specificity of alterations in NGF, BDNF and NT-3 levels in Alzheimer's disease. *Neuroreport* **7**, 2925–2928 (1996).

163. Durany, N. *et al.* Brain-derived neurotrophic factor and neurotrophin-3 levels in Alzheimer's disease brains. *Int. J. Dev. Neurosci.* **18**, 807–13 (2000).

164. Nagahara, A. H. *et al.* Neuroprotective effects of brain-derived neurotrophic factor in rodent and primate models of Alzheimer's disease. *Nat. Med.* **15**, 331–337 (2009).

165. Blurton-Jones, M. *et al.* Neural stem cells improve cognition via BDNF in a transgenic model of Alzheimer disease. *Proc. Natl. Acad. Sci.* **106**, 13594–13599 (2009).

166. Mariga, A., Zavadil, J., Ginsberg, S. D. & Chao, M. V. Withdrawal of BDNF from hippocampal cultures leads to changes in genes involved in synaptic function. *Dev. Neurobiol.* **75**, 173–192 (2015).

167. Berchtold, N. C. *et al.* Synaptic genes are extensively downregulated across multiple brain regions in normal human aging and Alzheimer's disease. *Neurobiol. Aging* **34**, 1653–1661 (2013).

168. Murer, M. ., Yan, Q. & Raisman-Vozari, R. Brain-derived neurotrophic factor in the control human brain, and in Alzheimer's disease and Parkinson's disease. *Prog. Neurobiol.* **63**, 71–124 (2001).

169. Parain, K. *et al.* Reduced expression of brain-derived neurotrophic factor protein in Parkinson's disease substantia nigra. *Neuroreport* **10**, 557–561 (1999).

170. Fenner, M. E., Achim, C. L. & Fenner, B. M. Expression of full-length and truncated trkB in human striatum and substantia nigra neurons: implications for Parkinson's disease. *J. Mol. Histol.* **45**, 349–361 (2014).

171. Jha, S. K., Jha, N. K., Kar, R., Ambasta, R. K. & Kumar, P. p38 MAPK and PI3K/AKT Signalling Cascades inParkinson's Disease. *Int. J. Mol. Cell. Med.* **4**, 67–86 (2015).

172. Mitsumoto, H. *et al.* Arrest of Motor Neuron Disease in wobbler Mice Cotreated with CNTF and BDNF. *Science (80-.).* **265**, 1107–1110 (1994).

173. Korkmaz, O. T. *et al.* 7,8-Dihydroxyflavone improves motor performance and enhances lower motor neuronal survival in a mouse model of amyotrophic lateral sclerosis. *Neurosci. Lett.* **566**, 286–291 (2014).

174. Todd, D. *et al.* A Monoclonal Antibody TrkB Receptor Agonist as a Potential Therapeutic for Huntington's Disease. *PLoS One* **9**, e87923 (2014).

175. Fulgenzi, G. *et al.* BDNF modulates heart contraction force and long-term homeostasis through truncated TrkB.T1 receptor activation. *J. Cell Biol.* **210**, 1003–1012 (2015).

176. Yanpallewar, S. U., Barrick, C. A., Buckley, H., Becker, J. & Tessarollo, L. Deletion of the BDNF Truncated Receptor TrkB.T1 Delays Disease Onset in a Mouse Model of Amyotrophic Lateral Sclerosis. *PLoS One* **7**, e39946 (2012).

177. Nguyen, K. Q., Rymar, V. V. & Sadikot, A. F. Impaired TrkB Signaling Underlies Reduced BDNF-Mediated Trophic Support of Striatal Neurons in the R6/2 Mouse Model of Huntington's Disease. *Front. Cell. Neurosci.* **10**, (2016).

178. Altar, C. A. *et al.* Anterograde transport of brain-derived neurotrophic factor and its role in the brain. *Nature* **389**, 856–860 (1997).

179. Gauthier, L. R. *et al.* Huntingtin Controls Neurotrophic Support and Survival of Neurons by Enhancing BDNF Vesicular Transport along Microtubules. *Cell* **118**, 127–138 (2004).

180. Zuccato, C. *et al.* Loss of Huntingtin-Mediated BDNF Gene Transcription in Huntington's Disease. *Science (80-.).* **293**, 493–498 (2001).

181. ZUCCATO, C. & CATTANEO, E. Role of brain-derived neurotrophic factor in Huntington's disease. *Prog.*

Neurobiol. **81**, 294–330 (2007).

182. Zuccato, C. *et al.* Systematic Assessment of BDNF and Its Receptor Levels in Human Cortices Affected by Huntington's Disease. *Brain Pathol.* **18**, 225–238 (2007).

183. Ochs, G. *et al.* A phase I/II trial of recombinant methionyl human brain derived neurotrophic factor administered by intrathecal infusion to patients with amyotrophic lateral sclerosis. *Amyotroph. Lateral Scler. Other Mot. Neuron Disord.* **1**, 201–206 (2000).

184. Dey, N. D. *et al.* Genetically engineered mesenchymal stem cells reduce behavioral deficits in the YAC 128 mouse model of Huntington's disease. *Behav. Brain Res.* **214**, 193–200 (2010).

185. Cao, L. *et al.* Molecular therapy of obesity and diabetes by a physiological autoregulatory approach. *Nat. Med.* **15**, 447–454 (2009).

186. Weissmiller, A. M. & Wu, C. Current advances in using neurotrophic factors to treat neurodegenerative disorders. *Transl. Neurodegener.* **1**, 14 (2012).

187. Schlachetzki, J. C. M., Pizzo, D. P., Morrissette, D. A. & Winkler, J. Intracerebroventricular Administration of Nerve Growth Factor Induces Gliogenesis in Sensory Ganglia, Dorsal Root, and within the Dorsal Root Entry Zone. *Biomed Res. Int.* **2014**, 1–9 (2014).

188. Birch, A. M. & Kelly, Á. M. Chronic intracerebroventricular infusion of nerve growth factor improves recognition memory in the rat. *Neuropharmacology* **75**, 255–261 (2013).

189. Apfel, S. C. Nerve growth factor for the treatment of diabetic neuropathy: What went wrong, what went right, and what does the future hold? in 393–413 (2002). doi:10.1016/S0074-7742(02)50083-0

190. Apfel, S. C. *et al.* Efficacy and Safety of Recombinant Human Nerve Growth Factor in Patients With Diabetic Polyneuropathy. *JAMA* **284**, 2215 (2000).

191. Dyck, P. J. *et al.* Intradermal recombinant human nerve growth factor induces pressure allodynia and lowered heat-pain threshold in humans. *Neurology* **48**, 501–505 (1997).

192. Eriksdotter Jönhagen, M. *et al.* Intracerebroventricular Infusion of Nerve Growth Factor in Three Patients with Alzheimer's Disease. *Dement. Geriatr. Cogn. Disord.* **9**, 246–257 (1998).

193. Isaacson, L. G., Saffran, B. N. & Crutcher, K. A. Intracerebral NGF infusion induces hyperinnervation of cerebral blood vessels. *Neurobiol. Aging* **11**, 51–55 (1990).

194. Williams, L. R. Hypophagia is induced by intracerebroventricular administration of nerve growth factor. *Exp. Neurol.* **113**, 31–37 (1991).

195. Winkler, J. *et al.* Reversible schwann cell hyperplasia and sprouting of sensory and sympathetic neurites after intraventricular administration of nerve growth factor. *Ann. Neurol.* **41**, 82–93 (1997).

196. Jang, S.-W. *et al.* Gambogic amide, a selective agonist for TrkA receptor that possesses robust neurotrophic activity, prevents neuronal cell death. *Proc. Natl. Acad. Sci.* **104**, 16329–16334 (2007).

197. Scarpi, D. *et al.* Low molecular weight, non-peptidic agonists of TrkA receptor with NGF-mimetic activity. *Cell Death Dis.* **3**, e339–e339 (2012).

198. Harada, K., Kubo, M. & Fukuyama, Y. Chemistry and Neurotrophic Activities of (–)-Talaumidin and Its Derivatives. *Front. Chem.* **8**, (2020).

199. Bruno, M. A. Long-Lasting Rescue of Age-Associated Deficits in Cognition and the CNS Cholinergic Phenotype by a Partial Agonist Peptidomimetic Ligand of TrkA. *J. Neurosci.* **24**, 8009–8018 (2004).

200. Sisti, F. M., dos Santos, N. A. G., do Amaral, L. & dos Santos, A. C. The Neurotrophic-Like Effect of Carvacrol: Perspective for Axonal and Synaptic Regeneration. *Neurotox. Res.* **39**, 886–896 (2021).

201. Calogeropoulou, T. *et al.* Novel Dehydroepiandrosterone Derivatives with Antiapoptotic, Neuroprotective Activity. *J. Med. Chem.* **52**, 6569–6587 (2009).

202. Pediaditakis, I. *et al.* Selective and differential interactions of BNN27, a novel C17-spiroepoxy steroid derivative, with TrkA receptors, regulating neuronal survival and differentiation. *Neuropharmacology* **111**, 266–282 (2016).

203. Yilmaz, C. *et al.* ENT-A010, a Novel Steroid Derivative, Displays Neuroprotective Functions and Modulates Microglial Responses. *Biomolecules* **12**, 424 (2022).

204. Rogdakis, T. *et al.* Development and Biological Characterization of a Novel Selective TrkA Agonist with Neuroprotective Properties against Amyloid Toxicity. *Biomedicines* **10**, 614 (2022).

205. O'Leary, P. D. & Hughes, R. A. Design of Potent Peptide Mimetics of Brain-derived Neurotrophic Factor. *J. Biol. Chem.* **278**, 25738–25744 (2003).

206. Gudasheva, T. A., Logvinov, I. O., Antipova, T. A. & Seredenin, S. B. Brain-derived neurotrophic factor loop 4 dipeptide mimetic GSB-106 activates TrkB, Erk, and Akt and promotes neuronal survival in vitro. *Dokl. Biochem. Biophys.* **451**, 212–214 (2013).

207. Zainullina, L. F., Vakhitova, Y. V., Lusta, A. Y., Gudasheva, T. A. & Seredenin, S. B. Dimeric mimetic of BDNF loop 4 promotes survival of serum-deprived cell through TrkB-dependent apoptosis suppression. *Sci. Rep.* **11**, 7781 (2021).

208. Massa, S. M. *et al.* Small molecule BDNF mimetics activate TrkB signaling and prevent neuronal degeneration in rodents. *J. Clin. Invest.* **120**, 1774–1785 (2010).

209. English, A. W., Liu, K., Nicolini, J. M., Mulligan, A. M. & Ye, K. Small-molecule trkB agonists promote axon regeneration in cut peripheral nerves. *Proc. Natl. Acad. Sci.* **110**, 16217–16222 (2013).

210. Jang, S.-W. *et al.* A selective TrkB agonist with potent neurotrophic activities by 7,8-dihydroxyflavone. *Proc. Natl. Acad. Sci.* **107**, 2687–2692 (2010).

211. Jiang, M. *et al.* Small-molecule TrkB receptor agonists improve motor function and extend survival in a mouse model of Huntington's disease. *Hum. Mol. Genet.* **22**, 2462–2470 (2013).

212. Botsakis, K. *et al.* BNN-20, a synthetic microneurotrophin, strongly protects dopaminergic neurons in the "weaver" mouse, a genetic model of dopamine-denervation, acting through the TrkB neurotrophin receptor. *Neuropharmacology* **121**, 140–157 (2017).

213. Massa, S. M. Small, Nonpeptide p75NTR Ligands Induce Survival Signaling and Inhibit proNGF-Induced Death. *J. Neurosci.* **26**, 5288–5300 (2006).

214. Bai, Y. *et al.* Chronic and Acute Models of Retinal Neurodegeneration TrkA Activity Are Neuroprotective whereas p75NTR Activity Is Neurotoxic through a Paracrine Mechanism. *J. Biol. Chem.* **285**, 39392–39400 (2010).

215. Delbary-Gossart, S. *et al.* A novel inhibitor of p75-neurotrophin receptor improves functional outcomes in two models of traumatic brain injury. *Brain* **139**, 1762–1782 (2016).

216. Pediaditakis, I. *et al.* BNN27, a 17-Spiroepoxy Steroid Derivative, Interacts With and Activates p75 Neurotrophin Receptor, Rescuing Cerebellar Granule Neurons from Apoptosis. *Front. Pharmacol.* **7**, (2016).

217. Boltaev, U. *et al.* Multiplex quantitative assays indicate a need for reevaluating reported small-molecule TrkB agonists. *Sci. Signal.* **10**, (2017).

218. Berrera, M., Cattaneo, A. & Carloni, P. Molecular Simulation of the Binding of Nerve Growth Factor Peptide Mimics to the Receptor Tyrosine Kinase A. *Biophys. J.* **91**, 2063–2071 (2006).

219. Wang, Z. *et al.* Coaction of Electrostatic and Hydrophobic Interactions: Dynamic Constraints on Disordered TrkA Juxtamembrane Domain. *J. Phys. Chem. B* **123**, 10709–10717 (2019).

220. Wu, X., Li, Q., Wan, S. & Zhang, J. Molecular dynamics simulation and free energy calculation studies of the binding mechanism of allosteric inhibitors with TrkA kinase. *J. Biomol. Struct. Dyn.* **39**, 202–208 (2021).

221. Paoletti, F. *et al.* Endogenous modulators of neurotrophin signaling: Landscape of the transient ATP-NGF interactions. *Comput. Struct. Biotechnol. J.* **19**, 2938–2949 (2021).

222. Franco, M. L. *et al.* Interaction between the transmembrane domains of neurotrophin receptors p75 and TrkA mediates their reciprocal activation. *J. Biol. Chem.* **297**, 100926 (2021).

223. Cannarozzo, C. *et al.* Cholesterol-recognition motifs in the transmembrane domain of the tyrosine kinase receptor family: The case of TRKB. *Eur. J. Neurosci.* **53**, 3311–3322 (2021).

224. Resende-Lara, P. T., Perahia, D., Scott, A. L. & Braz, A. S. K. Unveiling functional motions based on point mutations in biased signaling systems: A normal mode study on nerve growth factor bound to TrkA. *PLoS One* **15**, e0231542 (2020).

225. Pietropaolo, A. *et al.* Binding of Zn(II) to Tropomyosin Receptor Kinase A in Complex with Its Cognate Nerve Growth Factor: Insights from Molecular Simulation and in Vitro Essays. *ACS Chem. Neurosci.* **9**, 1095–1103 (2018).

226. EuroNeurotrophin - a European training network for the discovery of neurotrophins small molecule mimetics as potential therapeutic agents for neurodegeneration and neuroinflammation. Available at: https://www.euroneurotrophin.eu/.

227. Schlick, T. *Molecular Modeling and Simulation: An Interdisciplinary Guide*. **21**, (Springer New York, 2010).

228. Hinchliffe, A. *Molecular Modelling for Beginners, 2nd Edition*. (2008).

229. Leach, A. R. *Molecular Modelling: Principles and Applications*. (Longman, 1996).

230. Cramer, C. J. *Essentials of Computational Chemistry: Theories and Models, 2nd Edition*. (2004).

231. Jensen, J. H. *Molecular Modeling Basics*. (CRC Press, 2010). doi:10.1201/9781420075274

232. Marrink, S. J., Risselada, H. J., Yefimov, S., Tieleman, D. P. & de Vries, A. H. The MARTINI Force Field: Coarse Grained Model for Biomolecular Simulations. *J. Phys. Chem. B* **111**, 7812–7824 (2007).

233. Monticelli, L. *et al.* The MARTINI Coarse-Grained Force Field: Extension to Proteins. *J. Chem. Theory Comput.* **4**, 819–834 (2008).

234. Marrink, S. J. *et al.* Computational Modeling of Realistic Cell Membranes. *Chem. Rev.* **119**, 6184–6226 (2019).

235. Wassenaar, T. A. *et al.* High-Throughput Simulations of Dimer and Trimer Assembly of Membrane Proteins. The DAFT Approach. *J. Chem. Theory Comput.* **11**, 2278–2291 (2015).

236. Majumder, A., Kwon, S. & Straub, J. E. On Computing Equilibrium Binding Constants for Protein–Protein Association in Membranes. *J. Chem. Theory Comput.* **18**, 3961–3971 (2022).

237. Dominguez, L., Foster, L., Meredith, S. C., Straub, J. E. & Thirumalai, D. Structural Heterogeneity in Transmembrane Amyloid Precursor Protein Homodimer Is a Consequence of Environmental Selection. *J. Am. Chem. Soc.* **136**, 9619–9626 (2014).

238. Lelimousin, M., Limongelli, V. & Sansom, M. S. P. Conformational Changes in the Epidermal Growth Factor Receptor: Role of the Transmembrane Domain Investigated by Coarse-Grained MetaDynamics Free Energy Calculations. *J. Am. Chem. Soc.* **138**, 10611–10622 (2016).

239. Bond, P. J. & Sansom, M. S. P. Insertion and Assembly of Membrane Proteins via Simulation. *J. Am. Chem. Soc.* **128**, 2697–2704 (2006).

240. Souza, P. C. T. *et al.* Martini 3: a general purpose force field for coarse-grained molecular dynamics. *Nat. Methods* **18**, 382–388 (2021).

241. Laio, A. & Parrinello, M. Escaping free-energy minima. *Proc. Natl. Acad. Sci.* **99**, 12562–12566 (2002).

242. Barducci, A., Bussi, G. & Parrinello, M. Well-Tempered Metadynamics: A Smoothly Converging and Tunable

Free-Energy Method. *Phys. Rev. Lett.* **100**, 020603 (2008).

243. Hénin, J., Lelièvre, T., Shirts, M. R., Valsson, O. & Delemotte, L. Enhanced Sampling Methods for Molecular Dynamics Simulations [Article v1.0]. *Living J. Comput. Mol. Sci.* **4**, (2022).

244. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).

245. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).

246. Sali, A. & Blundell, T. L. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **234**, 779–815 (1993).

247. Friesner, R. A. *et al.* Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *J. Med. Chem.* **47**, 1739–1749 (2004).

248. Halgren, T. A. *et al.* Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. *J. Med. Chem.* **47**, 1750–9 (2004).

249. Friesner, R. A. *et al.* Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein−Ligand Complexes. *J. Med. Chem.* **49**, 6177–6196 (2006).

250. Schrödinger Release 2020-2: Glide, Schrödinger, LLC, New York, NY, 2020.

251. Jorgensen, W. L., Maxwell, D. S. & Tirado-Rives, J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.* **118**, 11225–11236 (1996).

252. Lu, C. *et al.* OPLS4: Improving Force Field Accuracy on Challenging Regimes of Chemical Space. *J. Chem. Theory Comput.* **17**, 4291–4300 (2021).

253. Davis, S. J. & Crispin, M. Solutions to the Glycosylation Problem for Low- and High-Throughput Structural Glycoproteomics. in *Functional and Structural Proteomics of Glycoproteins* 127–158 (Springer Netherlands, 2010). doi:10.1007/978-90-481-9355-4_6

254. Kozak, S. *et al.* Homogeneously N -glycosylated proteins derived from the GlycoDelete HEK293 cell line enable diffraction-quality crystallogenesis. *Acta Crystallogr. Sect. D Struct. Biol.* **76**, 1244–1255 (2020).

255. Casalino, L. *et al.* Beyond Shielding: The Roles of Glycans in the SARS-CoV-2 Spike Protein. *ACS Cent. Sci.* **6**, 1722–1734 (2020).

256. Harbison, A. M. *et al.* Fine-tuning the spike: role of the nature and topology of the glycan shield in the structure and dynamics of the SARS-CoV-2 S. *Chem. Sci.* **13**, 386–395 (2022).

257. Vuorio, J. *et al.* N-Glycosylation can selectively block or foster different receptor–ligand binding modes. *Sci. Rep.* **11**, 5239 (2021).

258. Mehdipour, A. R. & Hummer, G. Dual nature of human ACE2 glycosylation in binding to SARS-CoV-2 spike. *Proc. Natl. Acad. Sci.* **118**, (2021).

259. Dobson, C. L. *et al.* Engineering the surface properties of a human monoclonal antibody prevents self-association and rapid clearance in vivo. *Sci. Rep.* **6**, 38644 (2016).

260. Banfield, M. J. *et al.* Specificity in Trk Receptor:Neurotrophin Interactions. *Structure* **9**, 1191–1199 (2001).

261. Park, S.-J. *et al.* CHARMM-GUI Glycan Modeler for modeling and simulation of carbohydrates and glycoconjugates. *Glycobiology* **29**, 320–331 (2019).

262. Abraham, M. J. *et al.* GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1–2**, 19–25 (2015).

263. Huang, J. *et al.* CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat. Methods* **14**, 71–73 (2017).

264. Nosé, S. A molecular dynamics method for simulations in the canonical ensemble. *Mol. Phys.* **52**, 255–268 (1984).

265. Hoover, W. G. Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A* **31**, 1695–1697 (1985).

266. Parrinello, M. & Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **52**, 7182–7190 (1981).

267. Nosé, S. & Klein, M. L. Constant pressure molecular dynamics for molecular systems. *Mol. Phys.* **50**, 1055–1076 (1983).

268. Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: An N ·log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **98**, 10089–10092 (1993).

269. Essmann, U. *et al.* A smooth particle mesh Ewald method. *J. Chem. Phys.* **103**, 8577–8593 (1995).

270. Hess, B., Bekker, H., Berendsen, H. J. C. & Fraaije, J. G. E. M. LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* **18**, 1463–1472 (1997).

271. Humphrey, W., Dalke, A. & Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38 (1996).

272. Gowers, R. *et al.* MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations. in 98–105 (2016). doi:10.25080/Majora-629e541a-00e

273. Michaud-Agrawal, N., Denning, E. J., Woolf, T. B. & Beckstein, O. MDAnalysis: A toolkit for the analysis of molecular dynamics simulations. *J. Comput. Chem.* **32**, 2319–2327 (2011).

274. Hunter, J. D. Matplotlib: A 2D Graphics Environment. *Comput. Sci. Eng.* **9**, 90–95 (2007).

275. Waskom, M. Seaborn: statistical data visualization. *J. Open Source Softw.* **6**, 3021 (2021).

276. Pedregosa, F. ; Varoquaux, G. ; Gramfort, A. ; Michel, V., ; Thirion, B. ; Grisel, O. ; Blondel, M. and Prettenhofer, P., ; Weiss, R. ; Dubourg, V. ; Vanderplas, J. ; Passos, A. ;, Cournapeau, D. and Brucher, M. ; & Perrot, M. and Duchesnay, E. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).

277. Harris, C. R. *et al.* Array programming with NumPy. *Nature* **585**, 357–362 (2020).

278. The pandas development team. pandas-dev/pandas: Pandas. (2020). doi:10.5281/zenodo.3509134

279. McKinney, W. Data Structures for Statistical Computing in Python. in *Proceedings of the 9th Python in Science Conference* (eds. van der Walt, S. & Millman, J.) 56–61 (2010). doi:10.25080/Majora-92bf1922-00a

280. Forsyth, P. A. *et al.* p75 Neurotrophin Receptor Cleavage by α- and γ-Secretases Is Required for Neurotrophin-mediated Proliferation of Brain Tumor-initiating Cells. *J. Biol. Chem.* **289**, 8067–8085 (2014).

281. Zahavi, E. E. *et al.* The receptor tyrosine kinase TrkB signals without dimerization at the plasma membrane. *Sci. Signal.* **11**, (2018).

282. Marchetti, L. *et al.* Fast-diffusing p75 NTR monomers support apoptosis and growth cone collapse by neurotrophin ligands. *Proc. Natl. Acad. Sci.* **116**, 21563–21572 (2019).

283. Vilar, M. *et al.* Ligand-independent signaling by disulfide-crosslinked dimers of the p75 neurotrophin receptor. *J. Cell Sci.* **122**, 3351–3357 (2009).

284. Vilar, M. *et al.* Activation of the p75 Neurotrophin Receptor through Conformational Rearrangement of Disulphide-Linked Receptor Dimers. *Neuron* **62**, 72–83 (2009).

285. Bateman, A. *et al.* UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res.* **49**, D480–D489 (2021).

286. Kumar, M. *et al.* The Eukaryotic Linear Motif resource: 2022 release. *Nucleic Acids Res.* **50**, D497–D508 (2022).

287. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. E. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **10**, 845–858 (2015).

288. Nugent, T. & Jones, D. T. Transmembrane protein topology prediction using support vector machines. *BMC Bioinformatics* **10**, 159 (2009).

289. Hofmann K. and Stoffel, W. TMbase-a database of membrane spanning proteins segments. *Biol. Chem. Hoppe Seyler* **347**, 166 (1993).

290. Krogh, A., Larsson, B., von Heijne, G. & Sonnhammer, E. L. . Predicting transmembrane protein topology with a hidden markov model: application to complete genomes. *J. Mol. Biol.* **305**, 567–580 (2001).

291. Bernhofer, M. *et al.* PredictProtein - Predicting Protein Structure and Function for 29 Years. *Nucleic Acids Res.* **49**, W535–W540 (2021).

292. Drozdetskiy, A., Cole, C., Procter, J. & Barton, G. J. JPred4: a protein secondary structure prediction server. *Nucleic Acids Res.* **43**, W389–W394 (2015).

293. Buchan, D. W. A. & Jones, D. T. The PSIPRED Protein Analysis Workbench: 20 years on. *Nucleic Acids Res.* **47**, W402–W407 (2019).

294. Combet, C., Blanchet, C., Geourjon, C. & Deléage, G. NPS@: Network Protein Sequence Analysis. *Trends Biochem. Sci.* **25**, 147–150 (2000).

295. T., A. K. CFSSP: Chou and Fasman Secondary Structure Prediction server. *Zenodo* (2013). doi:10.5281/zenodo.50733

296. Bocharov, E. V. *et al.* Spatial Structure and pH-dependent Conformational Diversity of Dimeric Transmembrane Domain of the Receptor Tyrosine Kinase EphA1. *J. Biol. Chem.* **283**, 29385–29395 (2008).

297. MacKenzie, K. R., Prestegard, J. H. & Engelman, D. M. A Transmembrane Helix Dimer: Structure and Implications. *Science (80-.).* **276**, 131–133 (1997).

298. Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **126**, 014101 (2007).

299. Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A. & Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **81**, 3684–3690 (1984).

300. Wassenaar, T. A., Ingólfsson, H. I., Böckmann, R. A., Tieleman, D. P. & Marrink, S. J. Computational Lipidomics with insane : A Versatile Tool for Generating Custom Membranes for Molecular Simulations. *J. Chem. Theory Comput.* **11**, 2144–2155 (2015).

301. Wassenaar, T. A., Pluhackova, K., Böckmann, R. A., Marrink, S. J. & Tieleman, D. P. Going Backward: A Flexible Geometric Approach to Reverse Transformation from Coarse Grained to Atomistic Models. *J. Chem. Theory Comput.* **10**, 676–690 (2014).

302. Barducci, A., Bussi, G. & Parrinello, M. Well-tempered metadynamics: A smoothly converging and tunable free-energy method. *Phys. Rev. Lett.* **100**, 1–4 (2008).

303. Tribello, G. A., Bonomi, M., Branduardi, D., Camilloni, C. & Bussi, G. PLUMED 2: New feathers for an old bird. *Comput. Phys. Commun.* **185**, 604–613 (2014).

304. Psachoulia, E., Fowler, P. W., Bond, P. J. & Sansom, M. S. P. Helix−Helix Interactions in Membrane Proteins: Coarse-Grained Simulations of Glycophorin A Helix Dimerization †. *Biochemistry* **47**, 10503–10512 (2008).

305. Javanainen, M., Martinez-Seara, H. & Vattulainen, I. Excessive aggregation of membrane proteins in the Martini model. *PLoS One* **12**, e0187936 (2017).

306. Stark, A. C., Andrews, C. T. & Elcock, A. H. Toward Optimized Potential Functions for Protein–Protein Interactions in Aqueous Solutions: Osmotic Second Virial Coefficient Calculations Using the MARTINI Coarse-Grained Force Field. *J. Chem. Theory Comput.* **9**, 4176–4185 (2013).

307. Ingólfsson, H. I. *et al.* Computational Lipidomics of the Neuronal Plasma Membrane. *Biophys. J.* **113**, 2271–2280 (2017).

308. Niemelä, P. S., Ollila, S., Hyvönen, M. T., Karttunen, M. & Vattulainen, I. Assessing the nature of lipid raft membranes. *PLoS Comput. Biol.* **3**, e34 (2007).

309. Chan, R. B. *et al.* Comparative Lipidomic Analysis of Mouse and Human Brain with Alzheimer Disease. *J. Biol. Chem.* **287**, 2678–2688 (2012).

310. Lee, J. *et al.* CHARMM-GUI Membrane Builder for Complex Biological Membrane Simulations with Glycolipids and Lipoglycans. *J. Chem. Theory Comput.* **15**, 775–786 (2019).

311. Jurrus, E. *et al.* Improvements to the APBS biomolecular solvation software suite. *Protein Sci.* **27**, 112–128 (2018).

312. Artim, S. C., Kiyatkin, A. & Lemmon, M. A. Comparison of tyrosine kinase domain properties for the neurotrophin receptors TrkA and TrkB. *Biochem. J.* **477**, 4053–4070 (2020).

313. Lucarelli, E., Kaplan, D. & Thiele, C. . Activation of trk-A but not trk-B signal transduction pathway inhibits growth of Neuroblastoma cells. *Eur. J. Cancer* **33**, 2068–2070 (1997).

314. Madhavi Sastry, G., Adzhigirey, M., Day, T., Annabhimoju, R. & Sherman, W. Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. *J. Comput. Aided. Mol. Des.* **27**, 221–234 (2013).

315. Schrödinger Release 2020-2: Protein Preparation Wizard; Epik, Schrödinger, LLC, New York, NY, 2020; Impact, Schrödinger, LLC, New York, NY; Prime, Schrödinger, LLC, New York, NY, 2020.

316. Olsson, M. H. M., Søndergaard, C. R., Rostkowski, M. & Jensen, J. H. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions. *J. Chem. Theory Comput.* **7**, 525–37 (2011).

317. Harder, E. *et al.* OPLS3: A Force Field Providing Broad Coverage of Drug-like Small Molecules and Proteins. *J. Chem. Theory Comput.* **12**, 281–296 (2016).

318. Schrödinger Release 2020-2: SiteMap, Schrödinger, LLC, New York, NY, 2020.

319. Halgren, T. A. Identifying and Characterizing Binding Sites and Assessing Druggability. *J. Chem. Inf. Model.* **49**, 377–389 (2009).

320. Halgren, T. New Method for Fast and Accurate Binding-site Identification and Analysis. *Chem. Biol. Drug Des.* **69**, 146–148 (2007).

321. Schrödinger Release 2020-2: Maestro, Schrödinger, LLC, New York, NY, 2020.

322. Schrödinger Release 2020-2: LigPrep, Schrödinger, LLC, New York, NY, 2020.

323. Farid, R., Day, T., Friesner, R. A. & Pearlstein, R. A. New insights about HERG blockade obtained from protein modeling, potential energy mapping, and docking studies. *Bioorg. Med. Chem.* **14**, 3160–73 (2006).

324. Sherman, W., Day, T., Jacobson, M. P., Friesner, R. A. & Farid, R. Novel Procedure for Modeling Ligand/Receptor Induced Fit Effects. *J. Med. Chem.* **49**, 534–553 (2006).

325. Sherman, W., Beard, H. S. & Farid, R. Use of an Induced Fit Receptor Structure in Virtual Screening. *Chem. Biol. Drug Des.* **67**, 83–84 (2006).

326. Schrödinger Release 2020-2: Induced Fit Docking protocol; Glide, Schrödinger, LLC, New York, NY, 2020; Prime, Schrödinger, LLC, New York, NY, 2020.

327. Schrödinger Release 2020-2: Prime, Schrödinger, LLC, New York, NY, 2020.

328. Jacobson, M. P., Friesner, R. A., Xiang, Z. & Honig, B. On the Role of the Crystal Environment in Determining Protein Side-chain Conformations. *J. Mol. Biol.* **320**, 597–608 (2002).

329. Jacobson, M. P. *et al.* A hierarchical approach to all-atom protein loop prediction. *Proteins* **55**, 351–67

(2004).

330. Charou, D. *et al.* Comprehensive characterization of the neurogenic and neuroprotective action of a novel TrkB agonist using mouse and human stem cell models of Alzheimer's Disease. *bioRxiv* 2023.05.08.539797 (2023). doi:10.1101/2023.05.08.539797