THE UNIVERSITY OF QUEENSLAND
AUSTRALIA

Signals in Motion: Determining How Signal Transduction is Mechanically Coupled Through Type-I Cytokine Receptors

Michael S. P. Corbett
B.Sc., M.Sc.

A thesis submitted for the degree of Doctor of Philosophy at
The University of Queensland in 2016
School of Chemistry & Molecular Biosciences

Abstract

The ability of a cell to recognise and respond to external stimuli is essential for cell survival and growth. Type-I cytokine receptors regulate many inflammatory, homeostatic, and growth and development signalling pathways. For example, the erythropoietin receptor is the main driver of red blood cell production from stem cells. All type-I cytokine receptors consist of an extracellular ligand-binding domain, a single helical transmembrane domain and an intracellular domain that associates with kinase/transcription factor signalling pathway, for example the JAK2/STAT5 pathway. Despite being extensively studied and structures of the extracellular domains of many of these receptors being known, the precise mechanism by which these receptors couple the event of ligand-binding to intracellular activation is not known. Indeed, multiple mechanisms have been proposed from experimental and crystallographic data for different receptors. For example, the activation of the growth hormone receptor is proposed to involve a relative rotation of two receptor chains, while in the case of the erythropoietin receptor the chains were suggested to separate in a scissor-like mechanism. In part, this is due to the limits of resolution achievable by experimental approaches as well as a lack of appreciation for the dynamical properties of the receptor proteins and the membrane environment. This thesis focuses on the use of fully atomistic molecular dynamics simulations to investigate the mechanism of activation for the type-I cytokine receptors. In particular MD simulations are used to probe whether conformations of the receptors observed crystallographically are representative of physiological conformations. Further to this, the quality of the X-ray crystal structures is tested by recreation of the crystal lattice in MD simulations. The effect of membrane composition is investigated by comparing the behaviour of the transmembrane domain in bulk and cholesterol-enriched raft-like bilayers. Finally, receptor dimers are examined for active and inactive conformations in raft-like bilayers containing sphingomyelin. Force field parameters for the lipid sphingomyelin that were used to build raft-like bilayers were also generated and validated. The work draws into question several of the mechanistic models proposed for the type-I cytokine receptors and demonstrates the difficulty in proposing a detailed mechanism of action based on limited sets of structural data. In particular, they highlight the limitations of the use of structural data in proposing a model when only a part of the receptor is considered. Results from the simulations also suggest that the lipid composition can influence the structure and behaviour of the receptors. Taken together the work shows the importance of considering not only the ligand-binding domain of the receptor but also the restrictions imposed by the transmembrane domain and structural influences caused by the membrane when proposing a mechanism of activation for this important class of cell surface receptor.

Declaration by author

This thesis is composed of my original work, and contains no material previously published or written by another person except where due reference has been made in the text. I have clearly stated the contribution by others to jointly-authored works that I have included in my thesis.

I have clearly stated the contribution of others to my thesis as a whole, including statistical assistance, survey design, data analysis, significant technical procedures, professional editorial advice, and any other original research work used or reported in my thesis. The content of my thesis is the result of work I have carried out since the commencement of my research higher degree candidature and does not include a substantial part of work that has been submitted to qualify for the award of any other degree or diploma in any university or other tertiary institution. I have clearly stated which parts of my thesis, if any, have been submitted to qualify for another award.

I acknowledge that an electronic copy of my thesis must be lodged with the University Library and, subject to the policy and procedures of The University of Queensland, the thesis be made available for research and study in accordance with the Copyright Act 1968 unless a period of embargo has been approved by the Dean of the Graduate School.

I acknowledge that copyright of all material contained in my thesis resides with the copyright holder(s) of that material. Where appropriate I have obtained copyright permission from the copyright holder to reproduce material in this thesis.

Publications during candidature

1. M. S. P. Corbett, D. Poger, and A. E. Mark, ``Revisiting the scissor-like mechanism of activation for the erythropoietin receptor," *FEBS Letters*, vol. 590, no. 18, pp. 3083--3088, 2016 (Reference [1]).

Publications included in this thesis

1. M. S. P. Corbett, D. Poger, and A. E. Mark, ``Revisiting the scissor-like mechanism of activation for the erythropoietin receptor," *FEBS Letters*, vol. 590, no. 18, pp. 3083--3088, 2016, incorporated as Chapter 2.

Contributor	Statement of contribution
Author MSPC (candidate)	Wrote the paper (60%) Performed the analysis (60%)
Author DP	Wrote the paper (20%) Performed the analysis (20%)
Author AEM	Wrote the paper (20%) Performed the analysis (20%)

Contributions by others to the thesis

1. Chapter 3 - Conceptualization M.S.P.C (candidate) (30%), D.P. (30%) and A.E.M. (40%); Investigation M.S.P.C (100%); Formal analysis M.S.P.C (100%); Visualization M.S.P.C (100%); Writing M.S.P.C (40%), D.P. (40%) and A.E.M. (20%); Supervision D.P. (50%) and A.E.M. (50%).

2. Chapter 4 - Conceptualization M.S.P.C (candidate) (50%) and A.E.M. (50%); Investigation M.S.P.C (100%); Formal analysis M.S.P.C (100%); Visualization M.S.P.C (100%); Writing M.S.P.C (80%), D.P. (10%) and A.E.M. (10%); Supervision D.P. (50%) and A.E.M. (50%).

3. Chapter 5 - Conceptualization M.S.P.C (candidate) (80%), D.P. (10%) and A.E.M. (10%); Investigation M.S.P.C. (100%); Formal analysis M.S.P.C. (100%); Visualization M.S.P.C. (100%); Writing M.S.P.C.(60%), D.P. (20%) and A.E.M. (20%); Supervision D.P. (50%) and A.E.M. (50%).

4. Chapter 6 - Conceptualization M.S.P.C (candidate) (40%) and D.P. (60%); Investigation M.S.P.C. (100%); Formal analysis M.S.P.C. (80%) and D.P. (20%); Visualization M.S.P.C. (90%), M.S. (5%) and B.C (5%); Writing M.S.P.C.(60%), D.P. (20%) and A.E.M. (20%); Supervision D.P. (50%) and A.E.M. (50%).

5. Chapter 7 - Conceptualization M.S.P.C (candidate) (80%), D.P. (10%) and A.E.M. (10%); Investigation M.S.P.C. (100%); Formal analysis M.S.P.C. (100%); Visualization M.S.P.C. (100%);

Writing M.S.P.C. (60%), D.P. (20%) and A.E.M. (20%); Supervision D.P. (50%) and A.E.M. (50%).

Statement of parts of the thesis submitted to qualify for the award of another degree

None.

Acknowledgements

To Alan I would like to offer my deepest thanks for your supervision throughout this project. Your support and advice was always appreciated and invaluable.

David, you were a great supervisor and have become a good friend. Merci de m'avoir enseigné qu'il existe plus de 16 couleurs.

A huge thanks to the past and present members of the MD Group for your friendship and support. Thanks to all the people in the 3^{rd} Floor for making it an enjoyable workplace. I wish you all the best.

Finally, to Sara and Theo (and little bubble), thank you, thank you, thank you, with all my love!

Keywords

cytokine, erythropoietin, growth hormone, receptor, signalling, molecular dynamics, sphingomyelin, lipid rafts

Australian and New Zealand Standard Research Classifications (ANZSRC)

ANZSRC code: 030402, Biomolecular Modelling and Design, 40%
ANZSRC code: 030406, Proteins and Peptides 30%
ANZSRC code: 060112, Structural Biology (incl. Macromolecular Modelling), 30%

Fields of Research (FoR) Classification

FoR code: 0601, Biochemistry and Cell Biology, 60%
FoR code: 0202, Atomic, Molecular, Nuclear, Particle and Plasma Physics, 30%
FoR code: 0307, Theoretical and Computational Chemistry, 10%

Contents

List of Figures

List of Tables

Acronyms and Units

ATB Automated Topology Builder

BPTI bovine pancreatic trypsin inhibitor

°C degree Celsius

CHOL cholesterol

CSF2Rα granulocyte-macrophage colony-stimulating factor receptor subunit α

DOPC 1,2-dioleoyl-*sn*-glycero-3-phosphocholine

DHPC 1,2-diheptanoyl-*sn*-glycero-3-phosphocholine

DPC *n*-dodecylphosphocholine

DPPC 1,2-dipalmitoyl-*sn*-glycero-3-phosphocholine

ECD extracellular domain

EMP erythropoietin mimetic peptide

EPO erythropoietin

EPOR erythropoietin receptor

fs femtosecond

GH growth hormone

GHR growth hormone receptor

GROMACS GROningen MAchine for Chemical Simulation

GROMOS GROningen MOlecular Simulation

IL interleukin

IL-R interleukin-receptor

IL-4Rα interleukin-4 receptor subunit α

IL-5Rα interleukin-5 receptor subunit α

IL-7Rα interleukin-7 receptor subunit α

JAK Janus kinase

K Kelvin

nm nanometer

NMR nuclear magnetic resonance

MD molecular dynamics

mEpoR mouse erythropoietin receptor

μ**L** microlitre

PC phosphatidylcholine

PCs phosphatidylcholines

PDB Protein Data Bank

PL placental lactogen

ps picosecond

POPC 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphocholine

PRL prolactin

PRLR prolactin receptor

PSM *N*-palmitoyl-D-*erythro*-sphingomyelin

QM quantum mechanics

RMSD root mean square deviation

rPRLR rat prolactin receptor

sEPOR soluble erythropoietin receptor

SM sphingomyelin

SPC simple-point charge

SSM *N*-stearoyl-D-*erythro*-sphingomyelin

STAT signal transducer and activator of transcription

TICR type-I cytokine receptor

TMD transmembrane domain

TPO thrombopoietin

TPOR thrombopoietin receptor

Chapter 1

Introduction

1.1 Signalling Through the Cell Plasma Membrane

1.1.1 Type-I cytokine receptors

A cell requires the ability to sense and monitor its surroundings to respond to and survive environmental changes. In cell communities, for example microbial biofilms and higher order organisms, such as humans, cells are required to work in concert and it is necessary for an individual cell to recognise and respond to signals released into the environment from other cells. Cytokines are a large family of proteins that act as chemical messengers to coordinate the function of cells during inflammation, immune responses, foetal and neonatal growth and development, and to maintain homeostasis. Cells perceive cytokines through protein receptors that are embedded in the plasma membrane. These receptors are able to selectively recognise and bind specific cytokines and elicit a cellular response. The signal of the cytokine binding is transferred via the receptor through the plasma membrane into the cytoplasm. Intracellular secondary chemical messengers are then activated, for example protein kinases, and continue the signalling pathway that in many cases ultimately results in changes in gene expression. An example of this process is the binding of the cytokine erythropoietin (EPO) to the erythropoietin receptor (EPOR) on the surface of stem cells in the bone marrow to regulate the formation of erythroblasts during erythropoiesis [2, 3]. Understanding how cytokine receptors transfer the activation signal into the cell could help explain how these receptors become dysregulated in diseases, such as chronic inflammation and cancer, and aid in the development of drugs designed to activate or inactivate specific signalling pathways.

Type-I cytokine receptors are a large family of cytokine receptors that comprise interleukin receptors, growth factor receptors and haematopoietic cytokine receptors (Figure 1.1). A sub-family of the haematopoietic cytokine receptors, referred to as the growth hormone receptor (GHR) family, consists of the growth hormone (GH), prolactin (PRL), EPO and thrombopoietin (TPO) receptors. The proteins of the GHR family are, in general, the smallest of the type-I cytokine receptors, but still contain all the features of a typical type-I cytokine receptor. These are: (i) an extracellular ligand-binding domain, (ii) a single-pass transmembrane domain, and (iii) an intracellular domain associated with a Janus kinase (JAK) coupled to a signal transducer and activator of transcription (STAT) sig-

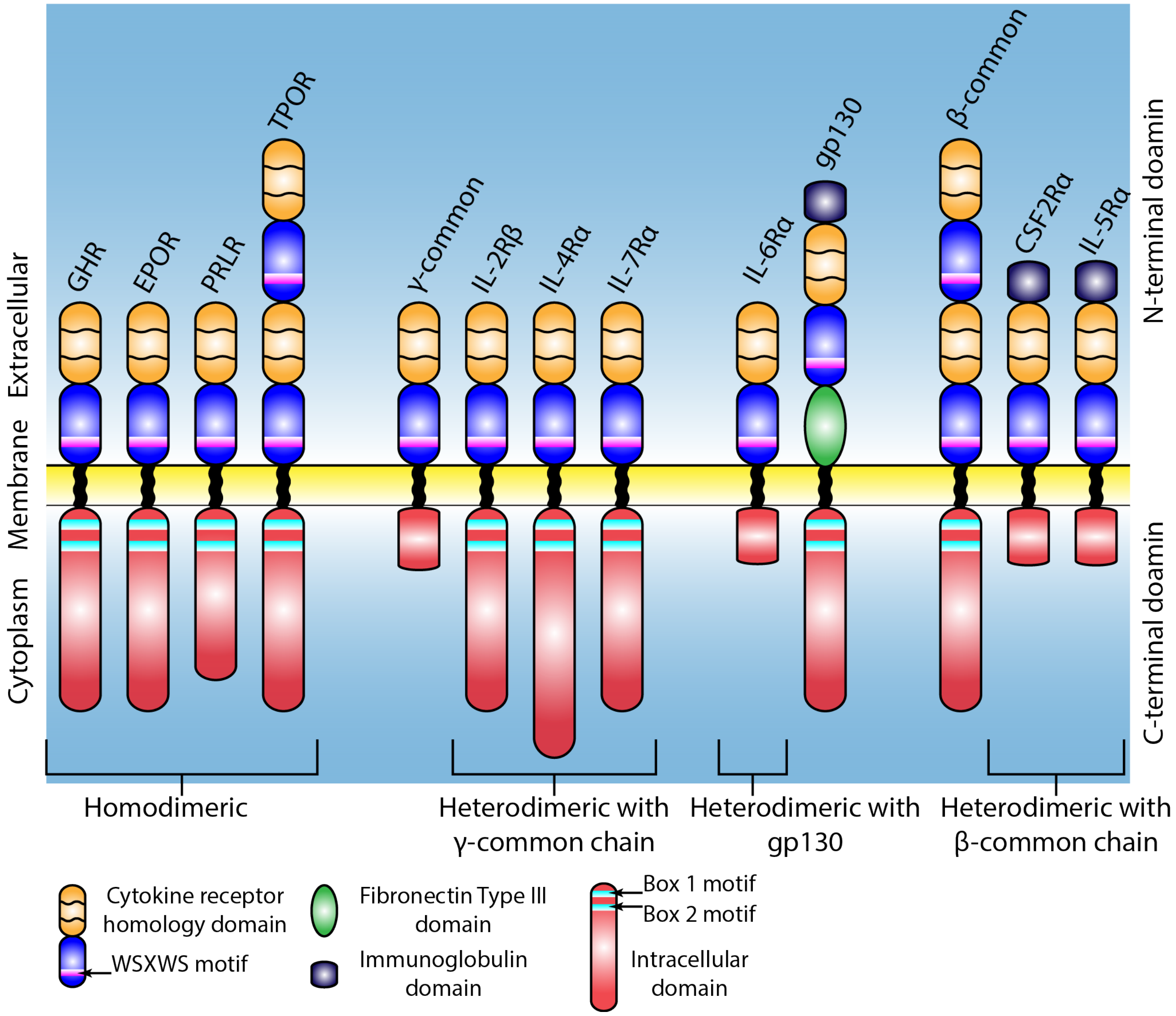


Figure 1.1: Representative members of the type-I cytokine receptor family. The ligand-binding extracellular domain of these receptors comprises a cytokine receptor homology domain, which itself is consists of two fibronectin type-III sub-domains with an immunoglobulin-like fold. The extracellular domain may also comprise additional fibotrnectin type-III or immunoglobulin domains. The receptors all have a single pass transmembrane domain that continues into the cytosolic domain. Conserved motifs amongst the receptors are the structurally important WSXWS motif in the extracellular domain and the Box 1 and 2 motifs in the cytosolic domain required for JAK binding [13].

nalling pathway. The GHR family also differ from other members of the type-I cytokine receptors, in that they are active as homodimers, whereas, for example, the interleukin-6 receptor is active as a heterodimer with gp-130 [4]. Currently, there is no atomic-resolution structure of a full-length type-I cytokine receptor. Instead, as in the case of the EPOR, only X-ray crystallographic structures of the soluble extracellular domains (referred to as soluble erythropoietin receptor (sEPOR), residues 1-226) [5, 6, 7, 8, 9] and an NMR structure of the transmembrane domain in a micelle [10, 11] have been reported. To date only the prolactin receptor (PRLR) has been been modelled as full-length receptor by piecing together X-ray crystal data for the extracellular domain and NMR data for the transmembrane and intracellular domains [12]. However, the use of this model either as a single chain or in a dimer complex to study the motions involved in the activation of the receptor has not been performed.

As first shown in 1991 in relation to the GHR, in this receptor family a single cytokine interacts

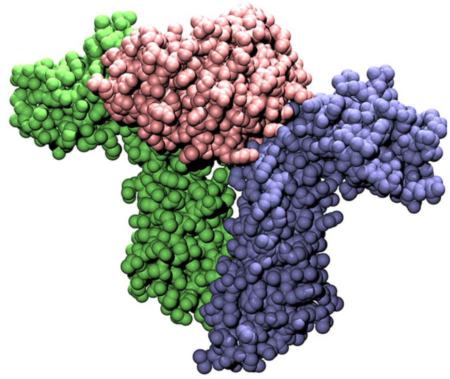


Figure 1.2: A Corey, Pauling and Koltun (CPK)-ball presentation of the X-ray crystal structure of the growth hormone cytokine (pink) bound in between two extracellular domains of the growth hormone receptor (green and blue). Protein Data Bank entry 3HHR.

with two independent receptor chains (Figure 1.2) to form an activated complex [14, 15]. The formation of such an activated complex results in the transphosphorylation of the receptor cytosolic domains at key tyrosine residues by the JAK and initiates one of the STAT pathways -- in the case of the GHR and EPOR, JAK2 and the STAT5 pathway. This transphosphorylation is in contrast to the members of the tyrosine kinase receptor family, e. g. the epidermal growth factor receptor, that auto-phosphorylate upon activation. Despite these similarities in the structure of the type-I cytokine receptors, the nature of ligand binding and the fact that all type-I cytokine receptors activate a JAK/STAT pathway, multiple mechanisms of activation have been proposed for the type-I cytokine receptors. These models include ligand-induced dimerisation [16], a relative rotation of the receptor chains [17, 18] and a cross-action scissor-like motion of the receptor chains [8] (Figure 1.3). It is not currently feasible to directly observe the movements of these receptors on the cell surface to elucidate the mechanism of action. In order to be able to observe the dynamical behaviour and properties of these receptors and to study the initial steps in receptor activation in atomic detail computational techniques, such as molecular dynamics (MD) simulations, will need to be employed.

As an example of a prototypical type-I cytokine receptor, Figure 1.4A shows the crystal structure of the complex between EPO (red) and two sEPOR protomers (green and blue). As can be seen, each soluble erythropoietin receptor (sEPOR) consists of two fibronectin type-III domains, labelled D1 (residues 1-117) and D2 (residues 122-226), represented in light and dark blue and green respectively. These are joined by a four-residue linker region depicted as yellow and magenta spheres (residues 118-121) [5, 6, 7, 8]. The cytokine EPO is a member of the interleukin-2 cytokine family with the typical coiled-coil four-helix bundle tertiary structure. To date, two alternative X-ray crystal structures of sEPOR bound to EPO (sEPOR$_2$-EPO) have been solved (Protein Data Bank (PDB) entries 1EER and 1CN4) [7]. In both cases the overall arrangement of the molecules and the binding interfaces between EPO and the two sEPOR molecules are similar [7]. As evident from Figure 1.4A, there are two EPO:EPOR interfaces, referred to as site 1 and site 2 (S1 and S2 respectively) [7]. For the sEPOR, the affinity at S1 is 1000 fold greater than at S2 (dissociation constants of 1 nM and 1 μM respectively) [19]. Nevertheless, occupation of both the S1 and S2 sites is required for receptor activation. For example, the mutation R103A in EPO abolishes binding at S2 resulting in an antagonistic ligand [16]. The mechanism by which the binding of EPO to the EPOR forms an activated complex has

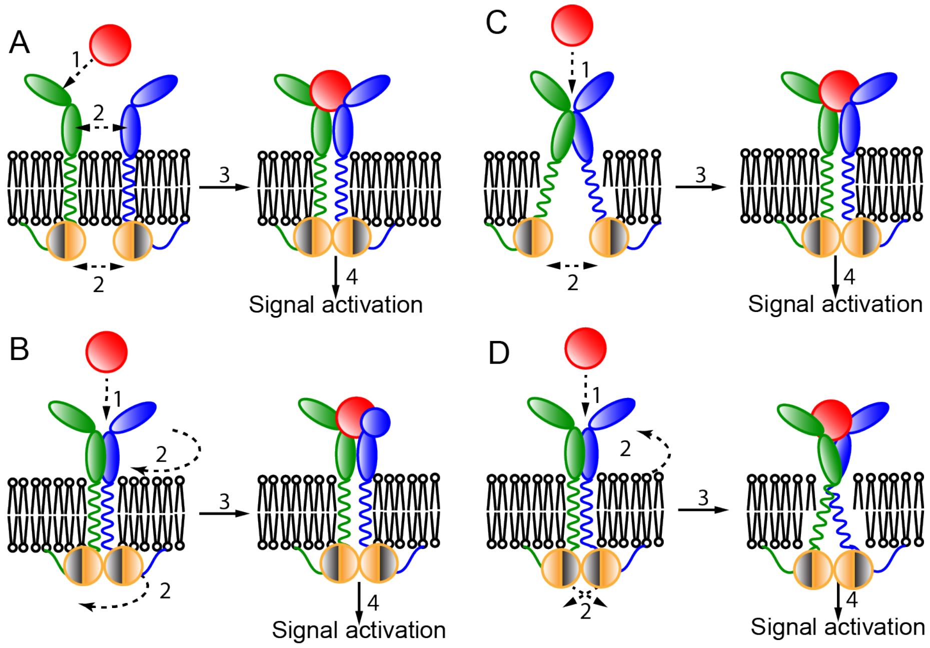


Figure 1.3: Proposed mechanisms of activation for the type-I cytokine receptors: (A) ligand-induced dimerisation, (B) relative rotation of the receptor chains and (C) cross-action scissor-like motion. (D) Proposed rotation-based mechanism of activation of the related growth hormone receptor [18]. The receptor chains A and B are coloured green and blue respectively, and the cytokine is coloured red. The orange spheres represent the protein kinase domain of the JAK protein associated with the cytosolic domains of the receptor that require the correct alignment of two domains in order to transphosphorylate.

been the subject of continuing debate with a number of alternative models having been proposed [16, 8, 20, 21, 22]. The main models of activation are ligand-induced dimerisation (Figure 1.3A) [16], a change in the relative orientation of the receptor chains (Figure 1.3B) [21], and cross-action scissor-like motion of the receptor chains (Figure 1.3C) [8, 20, 22].

The involvement of two receptor chains interacting with a single cytokine molecule in the activated type-I cytokine receptor complexes led initially to the widespread assumption that receptor activation was the result of ligand-induced dimerisation (Figure 1.3A). Indeed, EPO has been observed to induce the dimerisation of the sEPOR in solution [19]. The proposal that dimerisation was key to activation was supported by the fact that antibodies generated against the extracellular domains of the GHR and EPOR could activate the respective receptors [23, 24], and mutagenesis studies of the EPOR that showed certain cysteine mutants capable of forming intermolecular disulphide bridges were constitutively active [25]. However, other studies have shown that in the absence of cytokine, the receptors could be found in the form of inactive dimers and oligomers on the cell surface [26, 27, 17]. The extent to which the receptor exists as a preformed dimer continues to be debated. Recent single-molecule fluorescence measurements suggest the EPOR may be primarily monomeric on the cell surface in the absence of EPO [9]. That the receptors can exist as preformed dimers on the cell surface has led some to suggest that the activation of the receptor is triggered by difference in the structure of the dimer between the apo (unbound) and the holo dimer (cytokine-bound) rather than by the process of dimerisation itself [20, 27, 17, 28, 29].

Experimentally, the activation of the EPOR dimer after EPO binding would appear to involve a reorientation of the receptor molecules. Isothermal titration calorimetry and circular dichroism data

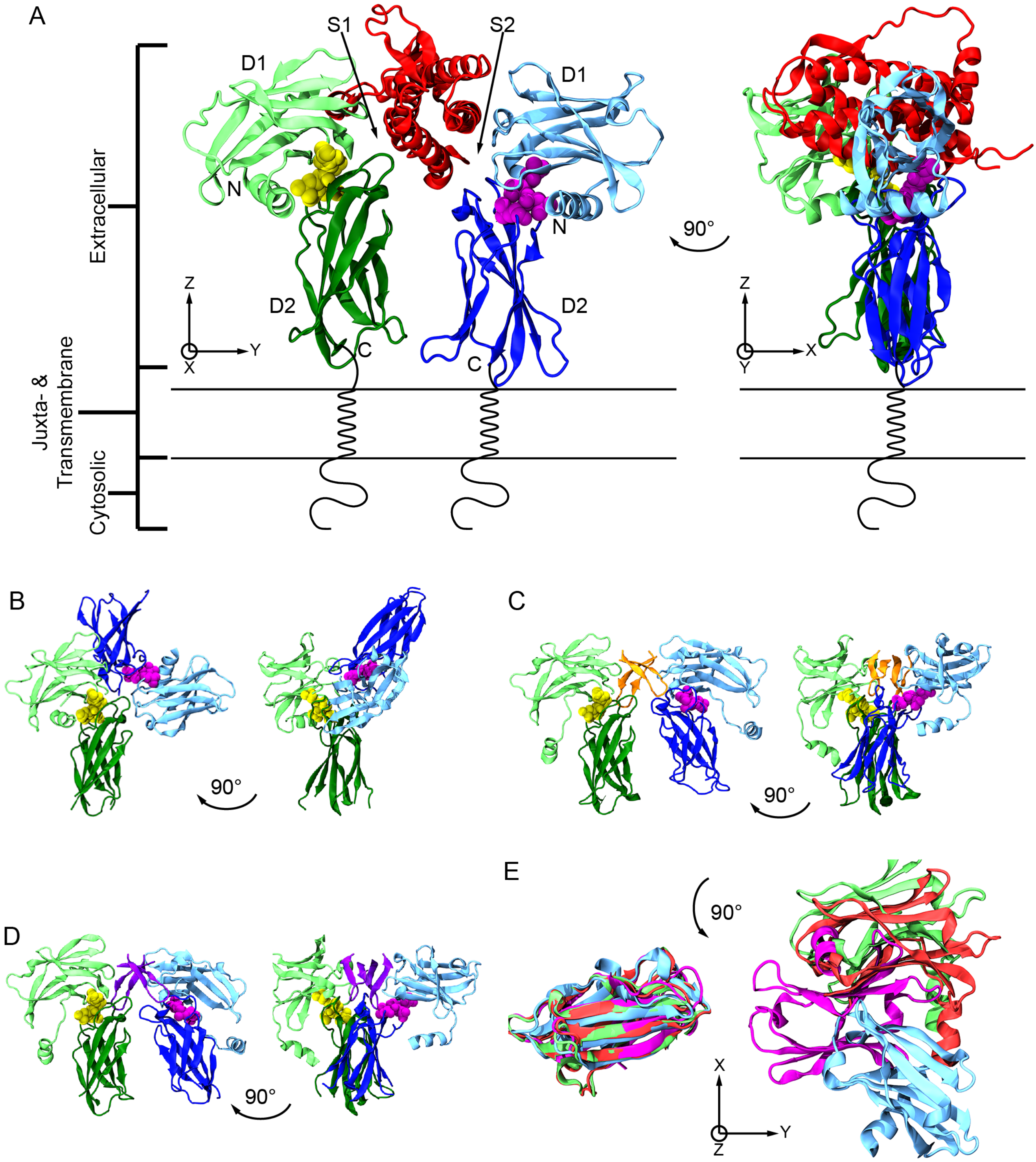


Figure 1.4: Crystal structures of the soluble EPOR extracellular domain dimers. The extracellular domain is shown bound to EPO (A), in the apo form (B), bound to an agonistic mimetic peptide EMP1 (C) and bound to an anagonist mimetic peptide EMP33 (D) (PDB entries 1EER, 1ERN, 1EBP and 1EBA, respectively)[5, 6, 7, 8]. (A-D) Chains A and B are shown in green and blue, respectively, and the ligand is shown in (A) red, (C) orange and (D) purple. The four residues in the hinge region connecting the N-terminal D1 domain and the C-terminal D2 domain within a chain are represented as yellow and magenta spheres in chain A and chain B respectively. Note, the extracellular juxtamembrane region and the transmembrane and cytosolic domains are depicted schematically in (A). (E) Top view (looking down onto the surface of the plasma membrane) of the two D1 domains in the EPOR dimer from the crystal structures of the EPOR bound to EPO (blue), EMP1 (red), EMP33 (green) and in the apo state (magenta). The structures in (B-E) have been aligned on the receptor chain A of (A) as a reference.

of the EPO binding to a chimeric construct consisting of an sEPOR$_2$ dimer fused to an antibody Fc domain suggest a conformational change within the dimer [30]. Furthermore, studies that were designed to induce a relative rotation between the two transmembrane helices within the EPOR dimer [21], and similar chimeric constructs with the GHR [18], could generate constitutively active or inactive dimers. This suggests regulation of the activity of the dimer occurs through particular orientations of the two receptor chains rather than merely bringing the two chains into close proximity. However, the precise nature of the structural transition that leads to receptor activation upon ligand binding in the case of the type-I cytokine receptors is not known. In the GHR, ligand binding is hypothesised to result in the rotation and reorientation of the two extracellular domains with respect to each other within the dimer, leading to the partial separation of the cytosolic domains that brings together the protein kinase domains of JAK2 [18] (Figure 1.3D). This model is based on mutagenesis and Förster resonance energy transfer (FRET) studies using fluorophore-labelled cytosolic domains. This model is also supported by simulation studies of the holo and apo forms of the GHR extracellular domains [31], which showed a spontaneous rotation of the extracellular domains on removal of the cytokine, in line with experiment. Similar results were also obtained in the case of the prolactin receptor (PRLR) [32].

The rotation between the GHR molecules was proposed to be mediated in part via contacts between the two D2 domains of the extracellular domains [17]. Whether a similar mechanism occurs in the case of the EPOR is unclear. In the EPO-bound crystal structures of the sEPOR dimer, the two D2 domains do not form extensive contacts (Figure 1.4A). In fact, contacts in this region only occur in one of the alternative crystal forms. For example, in the PDB entry 1EER the two D2 domains form hydrogen-bonding interactions between Ser135 and Glu134, while in PDB entry 1CN4 the two domains do not make contact [7]. This is in contrast to the extensive contacts observed in the crystal structures of the homodimeric GHR bound to the growth hormone [33] and the prolactin receptor bound to the placental lactogen or prolactin [34, 35]. It is all the more puzzling that the sEPOR chains are arranged in a similar manner to the receptor chains in other structures of cytokine-bound type-I cytokine receptors. Intriguingly, based on the crystal structure of the apo dimer of the sEPOR (Figure 1.4B), dimerisation was suggested to occur via the EPO-binding sites in the extracellular domains with the termini of the D2 domains remaining separated in the inactive dimer [8]. A similar arrangement of the receptor chains dimerising via the cytokine binding cite was observed for the apo dimer of the interleukin-7 receptor subunit α (IL-7Rα) [29]. However, if these apo dimers are truly representative of the inactive dimer on the cell surface it is unclear how the cytokine would initially bind and drive the dissociation of the receptor chains to allow binding to the cytokine binding sites to then drive a rearrangement between the two receptor chains.

Two further crystal structures of the sEPOR in complex with an erythropoietin mimetic peptide (EMP) have been solved (Figures 1.4C and D) [5, 6]. Specifically, the agonist EMP1 and antagonist EMP33. These EMPs are a part of a series of peptides that are 18-20 amino acids in length and are unrelated to EPO. In these EMP-bound complexes precise differences in the arrangement of the receptor chains were suggested to account for the activity of the ligands. To illustrate these different arrangements an overlay of the positions of the D1 domains of all the structures is shown in Figure

1.4E. The second sEPOR chain in the EMP33-bound dimer was calculated to be rotated 15° with respect to the second chain in the EMP1-bound dimer when first fitted on a single sEPOR chain [6]. In the EPO-bound structure, there is a greater asymmetry with the two chains at an angle of about 120° [7], and in the cross-like structure of the apo the D1 domains are almost parallel but inverted [8]. These differences in the arrangement of the receptor chains in these X-ray crystal structures have not only been suggested to account for the differences in the activity of the bound ligands but have also been interpreted as inactive and active conformations of the receptor dimer (in particular the apo and EMP1-bound EPOR dimers respectively) along a single trajectory involving a cross-action scissor-like motion (Figure 1.3C) [8]. The cross-action scissor-like motion would require the two D1 domains to move apart resulting in the two D2 domains (separated in the structure Figure 1.4B) being brought closer together (as seen in Figures 1.4A, C and D) and allowing the interaction of the two cytoplasmic domains [8, 20]. This would imply that EPOR operated via a fundamentally different mechanism to that proposed for the widely studied GHR [18]. Interestingly, it was recently inferred from the crystal structures of the sEPOR bound to three alternative diabodies that an activated complex of the sEPOR could be formed with the D2 C-termini separated by over 13 nm [9]. Such an arrangement would be in stark contrast to the active dimers proposed based on both the EPO and EMP1 complexes and challenge the proposal that the separation of the D2 domains in the apo dimer was significant in the regulation of the inactive dimer. The use of MD simulations to study these receptors can help resolve any doubt as to whether the differences in the structures observed crystallographically are truly significant differences that exist beyond the crystalline environment.

It is important to know whether these conformations of the receptors observed in the crystal structures are unique conformations in the presence and absence of agonistic and antagonistic ligands as they have been given physiological significance when proposing mechanisms of activation such as the cross-action scissor-like model for the EPOR. This becomes even more important when experimental data are interpreted in light of these models. For example, the cross-action scissor model was compatible with *in vivo* enzyme complementation experiments [20]. Specifically, chimeric constructs consisting of the EPOR extracellular and transmembrane domains linked to complementary fragments of dihydrofolate reductase (DHFR) by a flexible linker of 5, 10, or 30 amino acids were used to probe the activation of the receptor by fluorescence techniques. It was found that in the absence of EPO or EMP1, the DHFR fragments did not associate when linked to the EPOR by 5 amino acids whereas the longest 30-amino-acid linker was associated with functional DHFR. Addition of EPO or EMP1 was able to generate a functional DHFR with the 5-amino-acid linker. From this, it was claimed that only the cross-action scissor-like motion could account for the difference in activation of EPOR. Intriguingly, a similar protein fragmentation complementation assay, in which the full-length EPOR was fused to alternative β-galactosidase fragments, suggested little to no EPOR dimerisation in the absence of EPO or ligand [9]. It also suggested EMP33 could act as an agonist at sufficiently high concentrations. It is unclear however, if this later study captured the event of activation via ligand-induced dimerisation or the activation of a preformed dimer. In these assays detection of the threshold of β-galactosidase complementation is only detected at concentrations of EPO above 100 pM whereas phosphorylation of the EPOR could already be detected at concentrations of EPO as low as

10 pM (see fig. 1 of Ref. [9]). Together these data indicate that considering the mechanism of activation to be solely ligand-induced dimerisation is overly simplistic. The validity of the cross-action scissor-like mechanism has also been brought into question by immunofluorescence experiments using EPOR-PRLR heterodimers which suggest that interactions between the transmembrane domains play an important role in the dimerisation of the apo receptor [27], and studies of the self-association of the EPOR transmembrane domains in isolation [36, 37]. The diverse range of dimer configurations proposed to be active or inactive conformations and conflicting experimental data used to support these models raises an important question of whether it is valid to interpret these structures observed crystallographically as physiologically meaningful. Given the conflicting reports in the literature it is unclear which mechanism is likely in the case of the EPOR and whether the type-I cytokine receptors do indeed have a similar mechanism of activation across the family. In order to understand the dynamics involved in receptor activation MD will need to be employed. The EPOR provides a novel opportunity to study the effects of ligands with varying activities on receptor activation as the extracellular domain has been crystallised with several different agonistic and antagonistic ligands.

Further complexity in the activation of the type-I cytokine receptors arises when the effects of the membrane are considered. As noted above, although only one model of a full length receptor has been generated based on experimental data, attempts at linking the extracellular domain obtained from X-ray crystal data to a modelled transmembrane domain have been done for the GHR and PRLR [18, 32]. In this case, only the PRLR system has been studied by MD simulations [32]. In both instances however, the transmembrane domain was only considered in a pure phosphatidylcholine (PC) lipid bilayer. It has been observed for the EPOR that upon stimulation with EPO, the receptor colocalises with membrane rafts and that rafts are required for signalling [38]. It is therefore necessary to examine and compare the structure and behaviour of these receptor complexes in alternative membrane environments including a raft environment. MD simulations is a method of choice to yield exquisite details on the structure and dynamics of these receptors in different membrane environments. This however requires the development and validation of lipid force field parameters prior to investigating the properties of raft-like model membranes and protein behaviour in raft-like membranes.

1.1.2 Membrane rafts and sphingomyelins

The type-I cytokine receptors transfer the activation signal through the plasma membrane of the cell. The plasma membrane consists of a lipid matrix in which cell-surface proteins along with a wide variety of other proteins are embedded or associated. Within the membrane are patches of ordered lipids ranging between 20-200 nm in diameter and referred to as membrane rafts. Rafts organise and compartmentalise cellular processes and potentially regulate protein activity [39, 40]. These membrane rafts are generally richer in sphingolipids and cholesterol compared to the bulk membrane composed predominantly of PCs. The higher degree of order in these rafts is suggested to result from the intrinsic features of the sphingolipids compared to the PCs, as well as favourable interactions between sphingolipids and cholesterol. Initially recognised for having important roles in membrane trafficking [39], membrane rafts have been implicated as essential to the functioning of numerous receptor and transporter proteins, for example the T-cell antigen receptor and CD20 [41, 42]. Indeed, the EPOR, which co-localises with membrane rafts upon cytokine binding, loses the ability to signal when cholesterol is depleted from the cell [38]. This suggests involvement of the membrane environment in modulating type-I cytokine receptor structure and function during activation. While potential differences in the conformation of the receptor complex in different lipid environments may be observable via techniques such as cryo-electron microscopy, the use of MD provides a unique opportunity to study the dynamics and interactions of the receptor and lipid components in atomic detail. Of particular interest is how these receptors behave in the presence of an order-enhancing lipid such as the sphingolipid sphingomyelin.

A major component of membrane rafts is the sphingolipid family of lipids that is based on the molecule sphingosine. Sphingomyelins are the predominant phospholipid in this family and are generated by the addition of a phosphocholine group and an acyl chain to the sphingosine backbone (Figure 1.5). Within the sphingomyelins the length of the acyl chain varies with tissue type, for example the acyl chains palmitoyl and stearoyl are predominantly found in erythrocytes and nervous tissue respectively [43, 44, 45]. The major chemical differences between a typical sphingomyelin, such as *N*-palmitoyl-D-*erythro*-sphingomyelin (PSM) (Figure 1.5), and a glycerophospholipid, such as 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphocholine (POPC) (Figure 1.5B), are the degree of acyl chain saturation, configuration of the carbon-carbon double bond and the number of H-bond donor and acceptor groups (for completeness the structure of the sterol cholesterol is also shown in Figure 1.5C). Sphingomyelins tend to have a higher degree of saturation in the acyl chain, and the double bond in the sphingosine is in the *trans* configuration. In contrast POPC has an unsaturated carbon chain with the double bond in the *cis* configuration. Sphingomyelins also have both both H-bond donor (hydroxyl and amide hydrogens) and H-bond acceptor (amide carbonyl oxygen) group, while glycerophospholipids only have the H-bond acceptor (carbonyl oxygen) groups. These properties promote tighter packing of the lipids and result in a higher gel-to-liquid-crystalline melting temperature allowing the lipids to phase-separate at physiological temperatures with the bulk membrane and form the rafts. As well as having a higher degree of order than the bulk membrane, rafts tend also to be thicker. What remains unclear is whether membrane proteins, in particular the type-I cytokine receptors, require the presence of a particular lipid molecule or the general structural properties of membrane rafts (e. g. ordering and thickness) in order to function. Studying the type-I cytokine receptors by MD simulations in bulk and raft-like membrane is essential to elucidate the critical regulatory properties of the membrane environment on the receptors in atomic detail.

1.2 Molecular Modelling and Dynamics

1.2.1 Molecular dynamics

Despite extensive research into the activation of the type-I cytokine receptors the precise structural changes in the receptor complex after ligand binding that lead to activation have not been determined. This is due in part to the limits of resolution achievable by experimental approaches. To understand the activation of these receptors as a mechanical process requires the use of computational tools, in

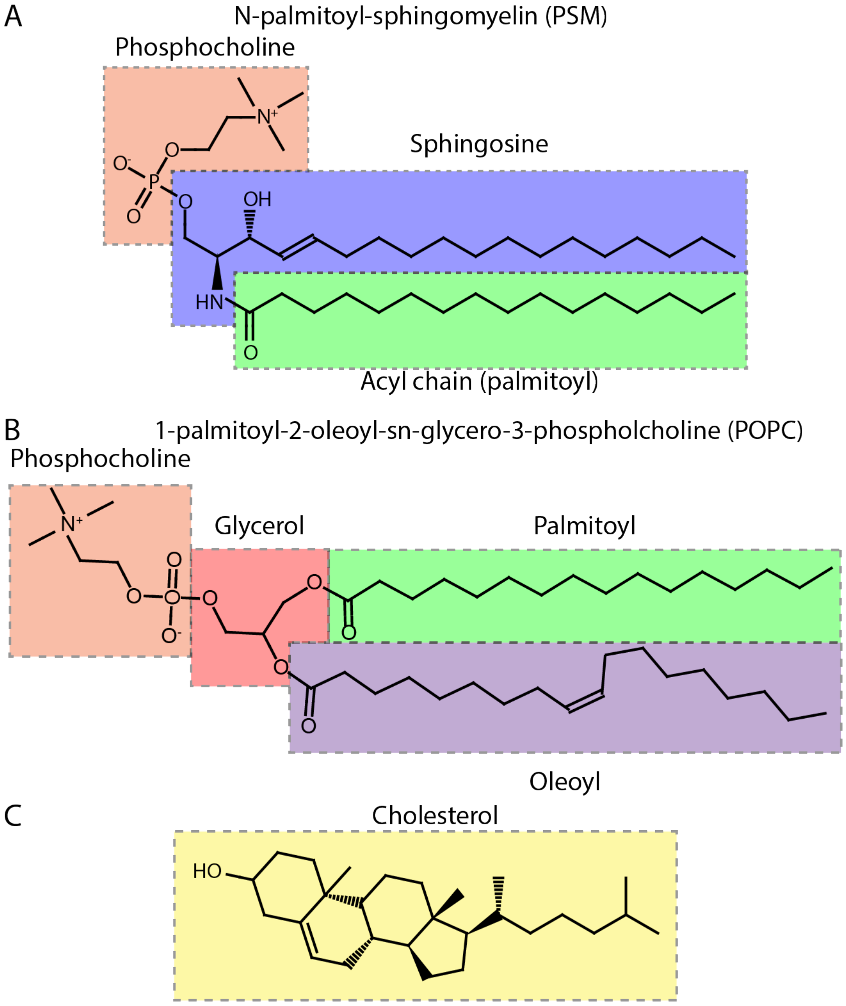


FIGURE 1.5: Molecular structure of a sphingomyelin (A), a phosphatidylcholine (B) and cholesterol (C). Highlighted are the phosphocholine headgroup (orange), sphingosine backbone (blue), glycerol backbone (red) and acyl chains (green and purple).

particular molecular dynamics (MD) simulations techniques. MD essentially treats a protein as a series of interconnected soft spheres with empirically derived and theoretical functions to describe the interactions between the particles. By assigning all the atoms in the system initial velocities, Newton's Equations of motion are solved by finite difference methods to propagate the system through time, which generates a trajectory that describes how the system changes with time. The receptors can therefore be modelled in a range of different environments, e. g. solution, crystalline or membrane-embedded, to examine their dynamical behaviour and how the environment influences the receptor complex. This section will provide an overview of MD theory [46, 47] beginning with Newton's Laws of Motion:

1. When viewed in an initial reference frame, an object either remains at rest or continues to move at a constant velocity, unless acted upon by an external force.

2. The vector sum of the external forces **F** on an object is equal to the mass m of that object multiplied by the acceleration vector **a** of the object: $\mathbf{F} = m\mathbf{a}$.

3. When a body exerts a force on a second body, the second body simultaneously exerts a force equal in magnitude and opposite in direction on the first body.

From the second law it can be shown the rate of change in the position of a particle with mass m_i along the coordinate x_i will have the directional force $\mathbf{F}_{x_i}$.

$$\frac{d^2x_i}{dt^2} = \frac{\mathbf{F}_{x_i}}{m_i} \tag{1.1}$$

The total force acting on the particle changes with its position and the positions of any interacting particles. The interactions between particles during MD are described by a force field. Classical force fields use empirically derived and theoretical functions to calculate the forces from bonded and non-bonded interactions between atoms in the electronic ground state. Using the GROMOS forcefield [48] as an example, the total potential energy (V_{total}) for a system is calculated as the sum of bonded and non-bonded terms, represented schematically in Figure 1.6. The bonded terms consist of covalent bonds (V_b), bond angles (V_a), dihedral angles (V_d) and improper dihedral angles (V_i). Non-bonded terms consist of van der Waals interactions, which are the short-range repulsion and long-range dispersion interactions approximated using the Lennard-Jones potential (V_{LJ}), and Coulomb interactions (V_C) between charges or partial charges on atoms.

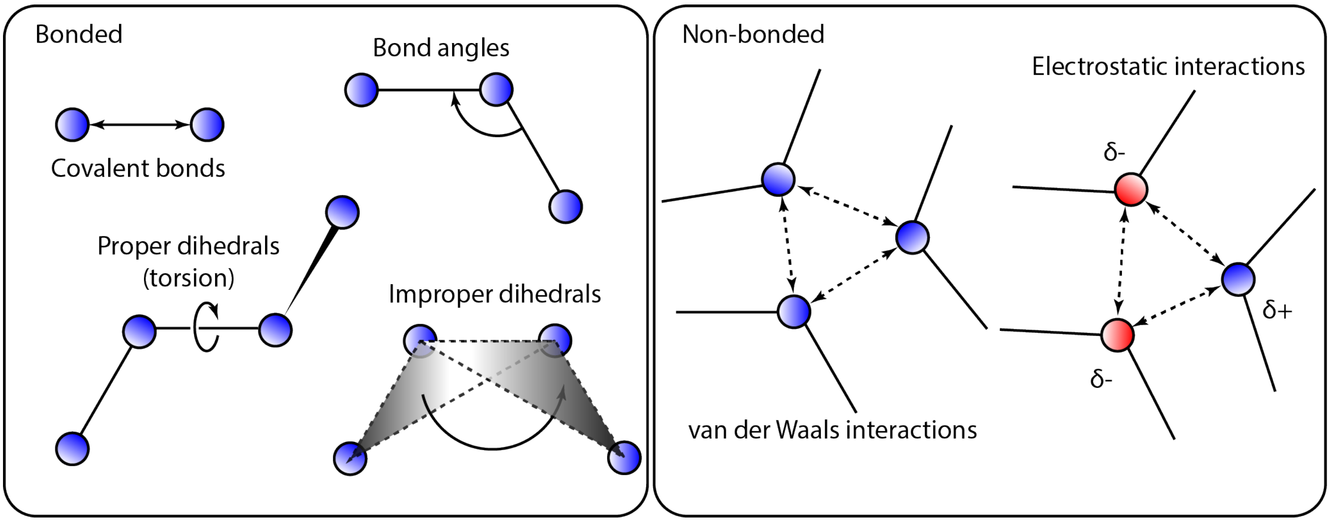


FIGURE 1.6: Illustration of the bonded and non-bonded terms used in the GROMOS forcefield.

$$V_{total} = V_b + V_a + V_d + V_i + V_{LJ} + V_C \tag{1.2}$$

V_b has a harmonic potential of the form:

$$V_b(\mathbf{r}_i, \mathbf{r}_j) = \frac{1}{2}k_b(r - b_0)^2 \tag{1.3}$$

where k_b is the bond stretching force constant, b_0 is the bond length at equilibrium and r is the distance between the two atoms (i, j) with the coordinates $\mathbf{r}_i$ and $\mathbf{r}_j$:

$$r = |\mathbf{r}_i - \mathbf{r}_j| \tag{1.4}$$

V_a has a harmonic potential of the form:

$$V_a(\mathbf{r}_i, \mathbf{r}_j, \mathbf{r}_k) = \frac{1}{2}k_\theta(\theta - \theta_0)^2 \tag{1.5}$$

where k_θ is the bond angle force constant, θ_0 is the bond angle at equillibrium and θ is the angle between atoms (i, j, k) with the coordinates $\mathbf{r}_i$, $\mathbf{r}_j$ and $\mathbf{r}_k$ calculated from the dot product:

$$\theta = \arccos\frac{\mathbf{r}_{ij} \cdot \mathbf{r}_{jk}}{\|\mathbf{r}_{ij}\|\|\mathbf{r}_{jk}\|} \tag{1.6}$$

V_d maintains the torsional angle between the atoms (i, j, k, l) and is given by the periodic function:

$$V_d(\phi) = k_\phi(1 + \cos(n\phi - \phi_0)) \tag{1.7}$$

where k_ϕ is the dihedral force constant, ϕ_0 is the dihedral angle at equilibrium, n is the periodicity (number of maxima and minima) of the potential between 0 and 2π, and ϕ is the angle between the two planes (i, j, k) and (j, k, l) is calculated from the dot product:

$$\phi = \arccos\frac{\mathbf{o}_{ijk} \cdot \mathbf{p}_{jkl}}{\|\mathbf{o}_{ijk}\|\|\mathbf{p}_{jkl}\|} \tag{1.8}$$

where $\mathbf{o}_{ijk}$ and $\mathbf{p}_{jkl}$ are the normals to the planes.

V_i maintains chirality and planarity of groups of four atoms i, j, k, l and has the harmonic form:

$$V_i = \frac{1}{2}k_\xi(\xi - \xi_0)^2 \tag{1.9}$$

where ξ is the angle between the planes (i, j, k) and (j, k, l), ξ_0 is the improper dihedral force constant and ξ_0 is the dihedral angle at equilibrium. The angle ξ is calculated using the dot product in Equation (1.8)

The van de Waals interactions between two atoms i, j are approximated using a Lennard-Jones potential of the form:

$$V_{LJ}(r) = 4\epsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right] \tag{1.10}$$

where ϵ is the measure of attraction between the two atoms i and j, σ is the distance at which the intermolecular potential between atoms i and j is zero, and r is the distance between the atoms (Equation (1.4)).

The Coulomb potential between two atoms i, j has the form:

$$V_C(r) = f_{el}\frac{q_i q_j}{\varepsilon_r r} \tag{1.11}$$

where q_i and q_j are the charges on the atoms, $f_{el} = (4\pi\varepsilon_0)^{-1}$, ε_0 is the physical constant of the dielectric permittivity in vacuum, ε_r is the relative dielectric permittivity and r is the distance between the atoms (Equation (1.4)).

To begin the simulation, the atoms in the system are assigned initial velocities using a Maxwell-Boltzmann distribution centered around the temperature (T) of reference:

$$p(v_{ix}) = \Big(\frac{m_i}{2\pi k_B T}\Big)^{\frac{1}{2}} \exp\Bigg[-\frac{1}{2}\frac{m_i v_{ix}^2}{k_B T}\Bigg] \tag{1.12}$$

where $p(v_{ix})$ is the probability that an atom i with mass m_i has a velocity v_{ix} in the x direction. k_B is the Boltzmann constant.

The system is then propagated through time by using a discrete time step (δt). The total force on each particle in the system at time t is calculated as the vector sum of its interactions with all other particles. The calculated forces are used to determine the acceleration of the particles. Combining the new accelerations with the known positions and velocities at time t the new positions and velocities at $t + \delta t$ are calculated. The forces at time $t + \delta t$ are then used to calculate the new positions and velocities at time $t + 2\delta t$ *ad infinitum* [46].

The MD simulation software package GROMACS [49, 50] implements the *leap-frog* algorithm [51], a variant of the *Verlet algorithm* [52] for integrating the equations of motion. The underlying assumption of these algorithms is the positions, velocities and accelerations can be approximated by Taylor series expansions:

$$\mathbf{r}(t+\delta t) = \mathbf{r}(t) + \delta t\mathbf{v}(t) + \frac{1}{2}\delta t^2\mathbf{a}(t) + \frac{1}{6}\delta t^3\mathbf{b}(t)\ldots \tag{1.13}$$

$$\mathbf{v}(t+\delta t) = \mathbf{v}(t) + \delta t\mathbf{a}(t) + \frac{1}{2}\delta t^2\mathbf{b}(t)\ldots \tag{1.14}$$

$$\mathbf{a}(t+\delta t) = \mathbf{a}(t) + \delta t\mathbf{b}(t)\ldots \tag{1.15}$$

v is the velocity (first derivative of the positions with respect to time), **a** is the acceleration (second derivative) and **b** the third derivative. The Verlet algorithm obtains the new positions $\mathbf{r}(t + \delta t)$ by adding the positions and accelerations at time t and the previous positions and accelerations at time $t - \delta t$:

$$\mathbf{r}(t+\delta t) = \mathbf{r}(t) + \delta t\mathbf{v}(t) + \frac{1}{2}\delta t^2\mathbf{a}(t) + \frac{1}{6}\delta t^3\mathbf{b}(t) + O(\delta t^4) \tag{1.16}$$

$$\mathbf{r}(t-\delta t) = \mathbf{r}(t) - \delta t\mathbf{v}(t) + \frac{1}{2}\delta t^2\mathbf{a}(t) - \frac{1}{6}\delta t^3\mathbf{b}(t) + O(\delta t^4) \tag{1.17}$$

$$\mathbf{r}(t+\delta t) = 2\mathbf{r}(t) - \mathbf{r}(t-\delta t) + \delta t^2\mathbf{a}(t) + O(\delta t^4) \tag{1.18}$$

The velocities are obtained from the Verlet algorithm by dividing the difference in positions at $t + \delta t$ and $t - \delta t$ by δt (Equation (1.19)), or estimated during the half-step $t + \frac{1}{2}\delta t$ (Equation (1.20)):

$$\mathbf{v}(t) = \frac{[\mathbf{r}(t+\delta t) - \mathbf{r}(t-\delta t)]}{2\delta t} \tag{1.19}$$

$$\mathbf{v}(t+\frac{1}{2}\delta t) = \frac{[\mathbf{r}(t+\delta t) - \mathbf{r}(t)]}{\delta t} \tag{1.20}$$

A major limitation of the Verlet algorithm is that the velocities are not used in the propagation of the positions and as a consequence it is not possible to regulate the temperature of the system during a simulation. To address this issue a number of other algorithms have been proposed. For example, the leap-frog algorithm uses the velocities at time $t - \frac{1}{2}t$ and accelerations at time t to calculate the velocities at time $t + \frac{1}{2}t$ which are used to update the positions $\mathbf{r}(t + \delta t)$:

$$\mathbf{v}(t + \frac{1}{2}\delta t) = \mathbf{v}(t - \frac{1}{2}\delta t) + \delta t\mathbf{a}(t) \tag{1.21}$$

$$\mathbf{r}(t + \delta t) = \mathbf{r}(t) + \delta t\mathbf{v}(t + \frac{1}{2}\delta t) \tag{1.22}$$

Many biophysical properties are dependent on the temperature and pressure, for example phase transitions in lipid bilayers. Several algorithms can be implemented to maintain the temperature and pressure in molecular dynamics simulation. The Berendsen thermostat uses the concept of an external heat bath that can provide or remove heat as required [53]. The velocities are scaled by a factor λ such that the rate of change is proportional to the difference between temperature of the bath (T_{bath}) and the temperature of the system (T).

$$\lambda^2 = 1 + \frac{\delta t}{\tau_T}\Big(\frac{T_{bath}}{T(t)} - 1\Big) \tag{1.23}$$

Similarly, the Berendsen barostat couples the pressure to a pressure bath by scaling the atomic coordinates which is effectively scaling the volume of the system [53]. The pressure scaling factor is:

$$\lambda = 1 - \kappa\frac{\delta t}{\tau_P}(P - P_{bath}) \tag{1.24}$$

where κ is the isothermal compressibility, usually of water ($\approx 4.5 \times 10^{-5}\text{atm}^{-1}$). τ_T and τ_P are coupling constants that determine how strong the coupling is between the scaling factor and the time step.

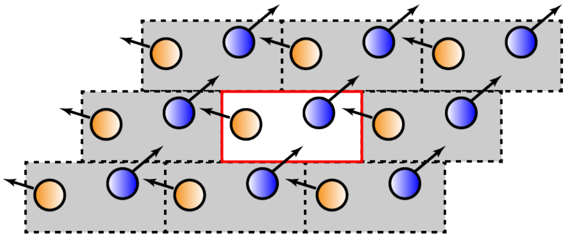


FIGURE 1.7: Periodic boundary conditions are created by surrounding a system (red box) with translated copies (dashed boxes) of itself in all directions. Any movement of a molecule outside the box is translated to the other side of the box as indicated by the arrows.

To eliminate edge of a box artefacts in molecular dynamics simulations the molecules are subjected to *periodic boundary conditions*. The system is placed within a space-filling unit cell or box and then surrounded by translated copies of itself, which in effect mimics an infinitely large system (Figure 1.7). This removes the effect of molecules hitting artificial boundaries or walls in the system and is not only beneficial but essential in simulations of environments such as lipid bilayers and protein crystals. A cost of using periodic boundary conditions is the system must be sufficiently large enough to prevent a molecule from seeing or interacting with its periodic image.

By careful and thoughtful construction of a molecular system, MD simulations provide a tool to study biological process in atomic detail. For example, by exploiting the effects of the periodic boundary conditions in two dimensions one can transform a small patch of a lipid bilayer into what can essentially be considered a membrane. Furthermore, a crystal lattice can be created by simply simulating a single unit cell under periodic conditions in three dimensions. MD is therefore an appropriate and powerful tool to use to probe some of the questions that arise from the literature on the type-I cytokine receptors: (i) are crystal structures of the EPOR appropriate models for active and inactive conformations of the receptor? (ii) does the membrane environment influence the receptor behaviour? (iii) how do the receptors mechanically send a signal from the extracellular domain through a membrane via the transmembrane region?

1.2.2 X-ray crystallography

Molecular modelling aims to provide a (relatively) simple atomic description of a molecular system that can be used for behaviour and structure prediction. While MD simulations can investigate the dynamics of a protein, it is heavily reliant on the initial coordinates of the models for the molecules used to build the system. By far the most common technique for generating a model of a protein structure is X-ray crystallography, which has been used to generate over 90% of the more than 100,000 structures in the Protein Data Bank. To obtain a model of a protein by X-ray crystallography, the protein must first be purified and concentrated into a supersaturated solution that can be forced to phase separate into a protein-rich phase (the protein crystal) and a saturated protein solution. This separation can be caused, for example, by gradually evaporating the solvent. The protein crystal is comprised of the protein molecule periodically arranged in space with, on average, 50% solvent in large channels between the stacked protein molecules. Structural information about the atom positions within the crystal is generated from exposure of the crystal to intense X-ray radiation from multiple angles with the scattered reflections detected (Figure 1.8) [54]. The X-ray diffraction pattern can be described using Bragg's law (Equation (1.25)):

$$n\lambda = 2d\sin\theta \tag{1.25}$$

At a particular incidence angle (θ) of a radiation beam with a wavelength λ, two reflecting planes within the crystal will produce constructive interference when separated by a certain distance (d) resulting in intense peaks of reflected radiation. In Equation (1.25) n is an integer.

Each reflection is indexed according to the number of times a set of reflecting planes intersects the respective $\boldsymbol{a}$, $\boldsymbol{b}$ or $\boldsymbol{c}$ axis of the unit cell ((hkl) where h, k and l are all integers). The unit cell being the smallest number of asymmetric units that can reproduce the crystal lattice by only using translational operations (Figure 1.9). Also measured is the intensity of the reflection, which is proportional to the square of the structure factor for the reflection. The structure factor (F_{hkl}) for a given reflection is a summation of the scattering factor (f_j) over all atoms j with their x, y, z fractional coordinates (Equa-

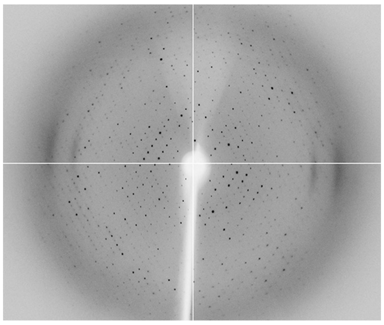


FIGURE 1.8: Bragg diffraction giving rise to an X-ray diffraction pattern. Image of X-ray diffraction from RPP1 NdA TIR domain, courtesy of Adam Bentham.

tion (1.26)). f_j is the scattering factor for atom j and is dependent on the atom type and diffraction angle of the reflection hkl.

$$F_{hkl} = \sum_{j=1}^{atoms} f_j e^{2\pi \mathrm{i}(hx_j + ky_j + lz_j)} \tag{1.26}$$

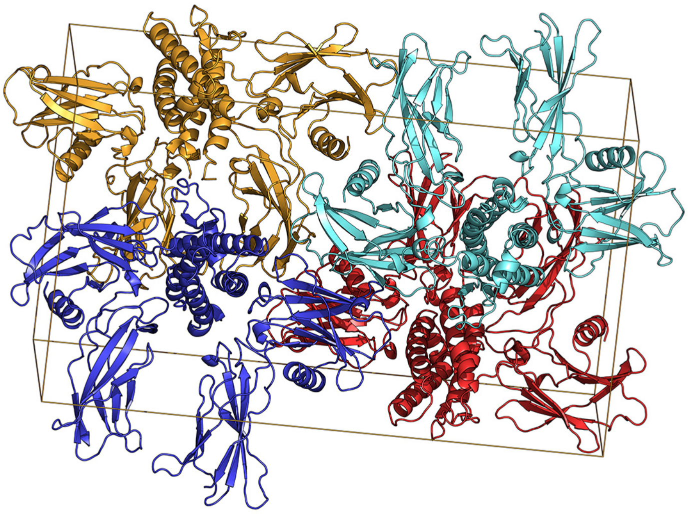


FIGURE 1.9: A unit cell of a crystal of the erythropoietin receptor in complex with erythropoietin (PDB entry 1EER) [7]. The crystal lattice is generated by translational operations of the unit cell. The unit cell itself is generated by translational and rotational operations (symmetry operations) performed on the asymmetric unit. The four asymmetric units coloured red, blue, green and yellow each contain two receptor chains and one erythropoietin molecule.

The structure factors are thus dependent on the local electron density at a given point in space. It becomes evident from Equation (1.26) that if the structure is known the structure factors can be calculated. X-ray crystallography deals with the inverse problem of determining the structure (or electron density (ρ)) from the structure factors. By applying a Fourier transformation to Equation (1.26) the following formula is obtained:

$$\rho(x, y, z) = \frac{1}{V} \sum_h \sum_k \sum_l F_{hkl} e^{-2\pi i(hx+ky+lz)} \tag{1.27}$$

This provides the electron density (ρ) as the sum of the structure factor (F_{hkl}) in electrons divided by the volume (V) of the unit cell. However, Equation (1.27) is expressed in terms of complex structure factors and the experimental data only provides the magnitude of the structure factors $|F_{hkl}|$.

Alternatively, this problem can be expressed by noting the absolute value of the complex vector F_{hkl} can be determined experimentally but not its phase α_{hkl}. The phase is determined by the distance between the crest of the emitted wave and a point of reference common to all the emitted X-rays. If the phases can be estimated, for example by proposing a given structure of the protein, they can be combined with the magnitudes to generate the electron density (Equation (1.27)):

$$\rho(x, y, z) = \frac{1}{V} \sum_h \sum_k \sum_l |F_{hkl}| e^{-2\pi i(hx+ky+lz-\alpha_{hkl})} \tag{1.28}$$

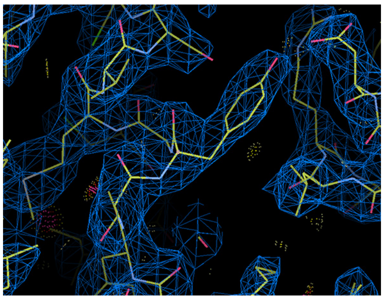


Figure 1.10: Electron density map generated during refinement. Image courtesy of Shane Horsefield.

One method for estimating the phases is multiple isomorphous replacement where there are two crystals of the same protein, except in one a specific atom within the protein has been replaced by a heavier atom, for example mercury. This method exploits the fact that the heavy metal dominates the observed diffraction pattern and thus the location of the heavy atoms can be obtained directly. These can then be used to solve the rest of the structure using Equation (1.29).

$$F_{hkl}^{H+P} = F_{hkl}^{P} + F_{hkl}^{H} \tag{1.29}$$

Where F_{hkl}^{H+P} is the reflection from a crystal containing a heavy metal, F_{hkl}^{P} is the same reflection from a crystal without the heavy metal and F_{hkl}^{H} is a theoretical reflection of a crystal containing only the heavy metal atoms and at the same locations as in the $H + P$ crystal. By determining the location of the heavy metal atoms in the unit cell from multiple isomorphous crystals, the structure factor amplitude and phases can be calculated for F_{hkl}^{H}. These in turn can be used to determine the phases for the F_{hkl}^{H+P} and F_{hkl}^{P} using vector addition.

Once the electron density map has been generated a crystallographic model is made by building the known protein sequence into the electron density (Figure 1.10). Subsequent rounds of refinement can be carried out by systematically adjusting the atomic positions and the amplitudes of the reflections by comparing the theoretical reflections calculated from the model to the experimental data. X-ray crystal structure of a protein can then be presented as stylised traces of the backbone atoms to illustrate protein folding or as a space filling model such as the Corey, Pauling and Koltun (CPK) model that reveals information about the surface of the protein (Figure 1.11).

Unlike crystals of, for example, sodium chloride (NaCl) where each sodium and chloride ion in the unit cells are essentially identical, each protein molecule in the unit cell varies slightly. This variation is a result of static disorder, in which the protein molecules have a different conformation or arrangement in each unit cell, and dynamic disorder as a result of thermal motion. These slightly

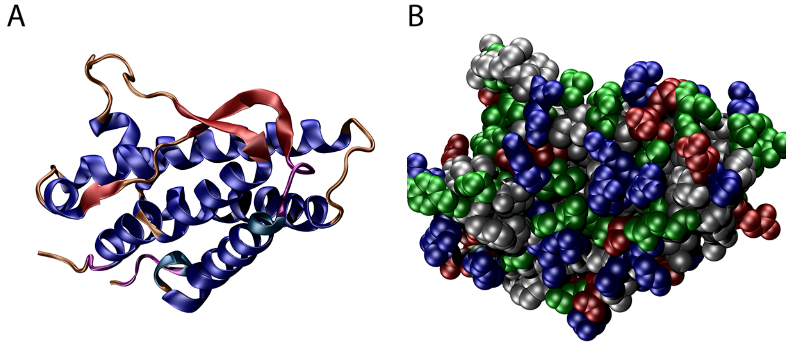


Figure 1.11: Representation of the crystal structure of EPO as (A) a cartoon highlighting secondary structure elements, or (B) a CPK model revealing a surface of the protein. Both (A) and (B) are taken from the same view of EPO from the Protein Data Bank entry 1EER [7].

alternative arrangement of atoms affect the scattering of the waves emanating from the crystal. To limit the effect of thermal motion the protein crystal is usually kept under liquid nitrogen during data collection, however flexible regions of the protein are frozen in multiple configurations and have a higher degree of uncertainty compared to more rigid regions. This means that the model structure will not only be an average structure of the unit cell over time during X-ray diffraction but also spatially within the crystal. To give a relative measurement of the atomic displacement of each heavy atom in the protein crystal, a B-factor is calculated for each atom. The B-factor is often interpreted as indicating increased areas of flexibility or dynamics. However, whilst X-ray crystallography is a powerful tool for determining protein structure, it only provides very limited information about protein dynamics. In fact analysis of multiple crystal structures indicated that atoms involved in crystal contacts had lower B-factors than those without contacts indicating that the assignment of B-factors is heavily dependent on the packing within the crystal [55]. This makes it difficult to discern in some cases if a region of a protein is more flexible than another by considering the X-ray structure alone.

A common method that has been believed to yield details of the dynamics within proteins is to solve multiple crystal structures of the same protein in different conditions. Indeed the structure of hen egg white lysozyme has been know for over 50 years, yet more recent structures produced under high-pressure have revealed novel conformations of the catalytic residues [56]. In the case of the EPOR alternative conformations of the receptor dimer solved under different conditions have been considered by some to be snapshots of the receptor complex along a single action pathway and used to propose a mechanism of action for the EPOR [8]. This assumes all the conformations of the protein within the crystal are not only of physiological significance but that straight (non-physical) linear interpolations between two crystal structures can be taken as dynamical processes. Specifically, the sEPOR-EMP1 complex involves two large receptor molecules and two small ligand molecules arranged within a crystalline environment. The individual contributions of the ligand molecules and the crystal packing forces to the overall arrangement of the complex are not known and whether the complex has a similar arrangement outside of a crystal has not yet been determined. While there are many metrics that can be used during the generation of a model structure from X-ray data to increase the reliability of the structure, such as Ramachandran plots, clash scores and R-values, these do not provide information on the biological significance of an X-ray crystal structure. To study the dynamics of proteins in atomic detail and determine physiologically significant conformations the use of computational techniques such as MD simulations are required.

1.3 Aims and Objectives of this Thesis

The overarching aim of this thesis is to better understand how the type-I cytokine receptors mechanically couple the event of a ligand binding to the extracellular domain to activation of the intracellular domains through a cell membrane.

Specific aims and objectives:

1. The first aim of the thesis is to determine whether the X-ray crystal structures of only the extracellular domain from the EPOR are appropriate models of the active and inactive conformations of the receptor complex. This will be achieved through a re-examination of the crystal structures and MD simulations of the receptor complex in solution bound to a series of agonistic and antagonistic ligands.

2. The second aim of the thesis to to ensure the models of the X-ray crystal structures of the EPOR have been modelled appropriately and that the force field is suitable for simulating these proteins. This will be achieved by recreating the crystal lattice of the models containing the EPOR as well as a high resolution crystal structure of bovine pancreatic trypsin inhibitor (BPTI).

3. The third aim of the thesis is to understand the influences of the membrane and lipid composition has on the behaviour of the type-I cytokine receptors. This will be achieved by simulation of the transmembrane domain in a variety of lipid bilayers. This will also require the generation and validation of parameters for sphingomyelin to be used in the raft-like bilayers.

4. The final aim of the thesis is to determine how the transmembrane domains are orientated in a dimer complex and how they are coupled to the extracellular domain. This will be examined by modelling the transmembrane domains in chimeric constructs such as the Jun and Put3 dimers. The linkage between the two domains will be studied by modelling the linker as a loop or continuation of the transmembrane domain helix.

1.4 Thesis Outline

Chapter 1: *Introduction.*

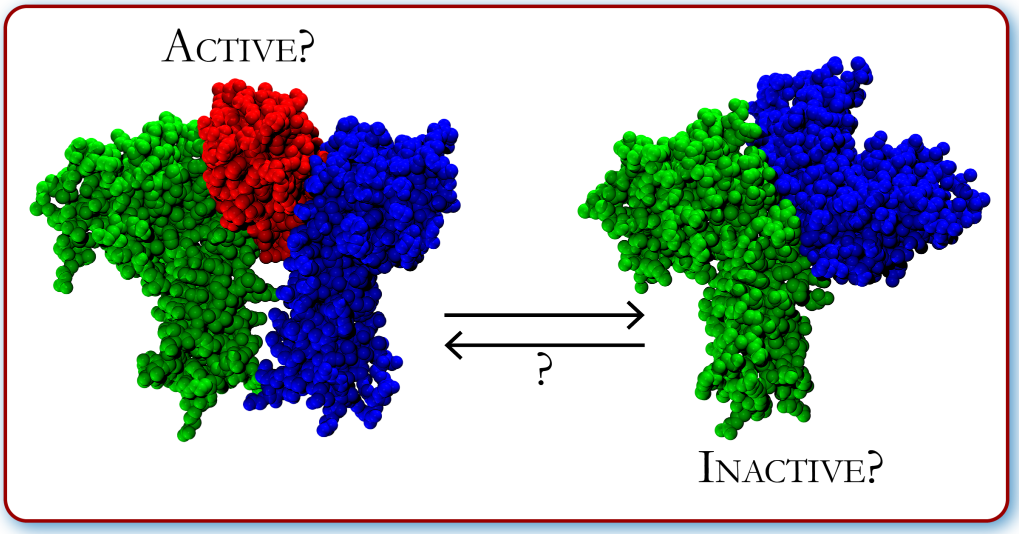
Chapter 2: *Revisiting the Scissor-Like Mechanism of Activation for the Erythropoietin Receptor.* This chapter reviews the support for the scissor-like mechanism and re-examines the crystal structures upon which the model was originally based.

Chapter 3: *Do X-ray Crystal Structures of Protein Complexes with Agonistic and Antagonistic Ligands Truly Represent Active and Inactive Conformations? The Erythropoietin Receptor a Case Study.* In this chapter the X-ray crystal structures of the EPOR are examined free in solution.

Chapter 4: *Can the Models of the X-ray Crystal Structures for the Erythropoietin Receptor Reproduce the Crystal Lattice?* In this chapter the X-ray crystal structure containing the EPOR are used to recreate the crystal lattice the stability of which is tested using MD simulations.
Chapter 5: *Does the Membrane Environment Affect the Structural Properties of the Type-I Cytokine Receptor Transmembrane Domains?* In this chapter the influence of cholesterol in the membrane on the transmembrane domain from several type-I cytokine receptors is examined.
Chapter 6: *Generation and Validation of Palmitoyl Sphingomyelin Parameters and Lipid Raft-Like Bilayers.* This chapter is dedicated to generating and validating parameters for sphingomyelin that are to be used in raft-like bilayer for MD simulations.
Chapter 7: *What are Active Conformations of a Type-I Cytokine Receptor Transmembrane Domain Dimer?* This chapter examines transmembrane domain dimers of the IL-7Rα, GHR and EPOR.
Chapter 8: *Conclusions and perspectives*

Chapter 2

Revisiting the Scissor-Like Mechanism of Activation for the Erythropoietin Receptor



2.1 Does the Erythropoietin Receptor Activate via a Scissor-like Mechanism?

2.1.1 History of the scissor-like model

In 1999, an article published in *Science* by Livnah *et al.* [8] proposed a potential mechanism for the activation of the erythropoietin receptor (EPOR) in which it was suggested that the binding of the cytokine induced a scissor-like motion of the two receptor domains within the receptor dimer. This model was based on an interpretation of differences between crystal structures of the EPOR dimer with and without the agonist erythropoietin mimetic peptide 1 (EMP1) (Figure 2.1). The article was also published back-to-back with an experimental study by Remy *et al.* [20] who used a protein fragment complementation assay to suggest the unliganded receptor could exist as a preformed dimer on the cell surface in a conformation in which the associated Janus kinase (JAK)2 molecules did not become transphosphorylated (activated). It was one of the first models to be proposed that could account for how the binding of a cytokine to a preformed receptor dimer might lead to activation. At the time, it was novel and attracted considerable attention and, according to the Web of Science, the work of Livnah *et al.* [8] has been cited over 430 times. This scissor-like model continues to be propagated. Not only is it still being reproduced in literature reviews [57] and cited to describe the intracellular domains of the receptor dimer moving into proximity of each other upon erythropoietin binding [58], but has in recent years also been used to interpret data on bacterial two-component systems [59] and even on T cell receptors [60]. For example, McElroy *et al.* [29] used the work of Livnah *et al.* [8] on EPOR to propose a scissor-like model to interpret crystal structures for the interleukin-7 receptor subunit α (IL-7Rα) (a related type-I cytokine receptor) with direct parallels being drawn between the two systems. This is concerning as the scissor-like model appears to be being invoked without proper consideration as to whether the proposed model is physiologically reasonable, or whether the model as originally proposed by Livnah *et al.* [8] is even compatible with what is currently known regarding the EPOR and other type-I cytokine receptors.

The scissor-like model of activation was proposed following a suggestion that preformed receptor dimers of the surface of the cell would be required to explain why erythropoietin binding to both the high- and low-affinity sites was required for activation [19]. At the time it was widely believed that signalling was initiated by the bringing together of the intracellular JAK2 molecules associated with the receptor cytosolic domains, and that these would need to be separated in the apo (unliganded) state. The structural basis on which the scissor-like model was proposed is illustrated in Figure 2.1, which is a recreation of fig. 1 of Livnah *et al.* [8]. On the left of Figure 2.1A is a backbone trace of the X-ray crystal structure of the apo dimer showing that the extracellular domains of two receptor molecules (chains A and B) adopt a cross-shaped arrangement in which the membrane-proximal D2 subdomains ($D2_A$ and $D2_B$) are widely splayed. In this arrangement the two equivalent C-terminal residues (Thr220) from each chain, to which the transmembrane helices are connected, are separated by 7.3 nm. In the left panel of Figure 2.1B, is the backbone trace of the X-ray crystal structure of the EPOR dimer bound to the agonist EMP1. In this structure the $D2_A$ and $D2_B$ subdomains lie close

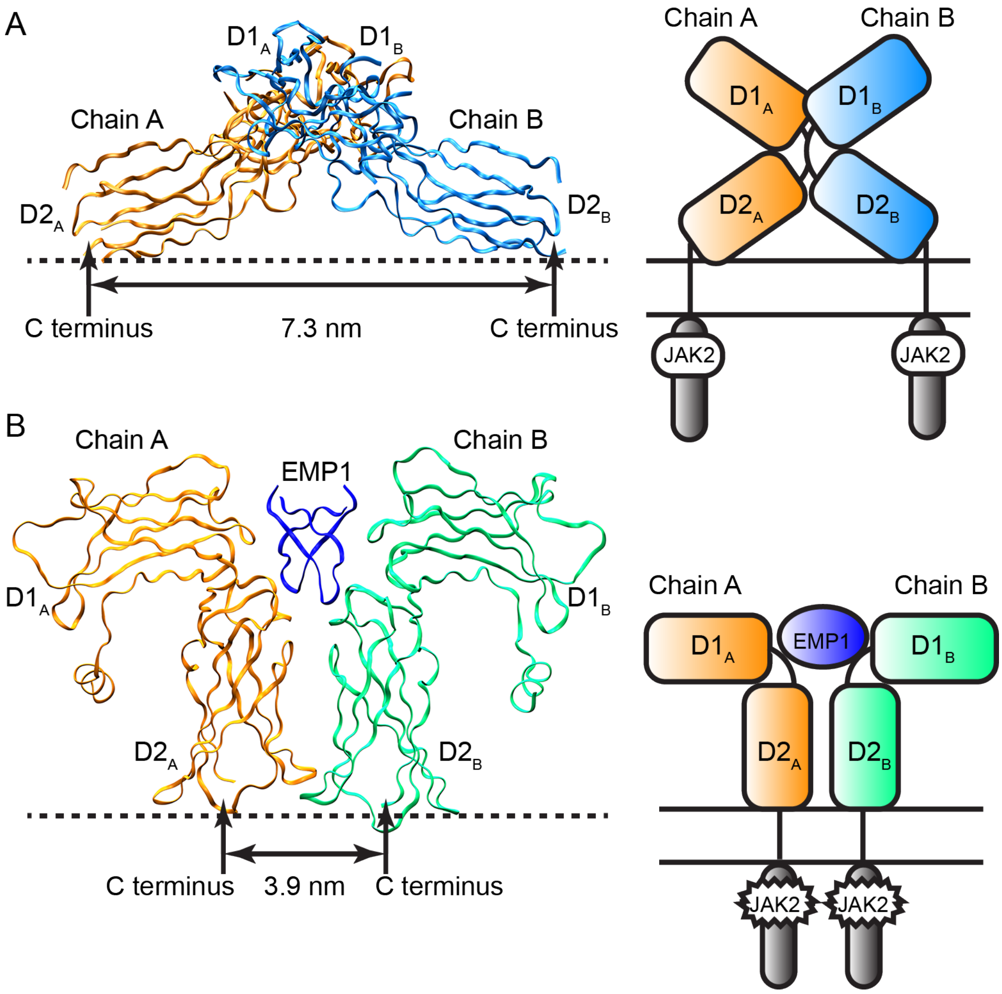


Figure 2.1: The projection of the crystal structure as presented in fig.1 1 from Livnah *et al.* [8] of (A) the apo EPOR dimer and (B) the EMP1-bound EPOR dimer illustrating the proposed scissor-like motion. Chain A is coloured orange in both (A) and (B) while chain B is coloured blue in (A) and green in (B). The black arrows indicate the membrane-proximal C-terminal ends of the D2 subdomains in each EPOR molecule. The distances between these two residues were calculated as 7.3 and 3.9 nm for the apo and EMP1-bound dimer respectively [8].

together with the C-terminal residues (Thr220) separated by just 3.9 nm [8]. On the right of Figure 2.1A and B is a recreation of the schematic from fig. 1 of Livnah *et al.* [8] of the receptor complex in relation to the membrane surface. As presented, it is easy to imagine the intracellular region and the associated JAK molecules being held apart by the extracellular domain of the receptor, and that the binding of ligand would lead to a reorganisation of the receptor molecules bringing the $D2_A$ and $D2_B$ subdomains closer together via a scissor-type motion. This could allow the transphosphorylation of the JAK2 molecules and subsequent signalling via the signal transducer and activator of transcription (STAT)5 pathway [8].

2.1.2 Experimental data conflicting with the scissor-like model

However, there are several difficulties with the scissor-like model for the EPOR. The model is based primarily on the differences between the structure of the apo dimer and the structure of the agonist EMP1. When the model was proposed, the structure of the EPOR bound to erythropoietin and also the structure of the EPOR bound to the antagonist EMP33 were also known. In the EMP33-bound structure, the EPOR chains are aligned in a manner more similar to the EMP1-bound structure than to the apo dimer. Despite the EMP33 complex being inactive, the D2 subdomains do not have the wide separation observed in the apo form. In fact, the distance between the C-terminal residues (Thr220) in the EMP33-bound structure is also 3.9 nm. The explanation offered by Livnah *et al.* [8] to account for the fact the EMP33-bound complex was inactive was that this complex is not as symmetric as that of EMP1. This explanation is difficult to reconcile with the fact that at the time it was known that in the EPO-bound EPOR structure the two receptor chains are asymmetrically arranged in the complex [7]. The authors also mentioned that the receptor-receptor contacts in the apo dimer were very similar to those used to bind EMP1, EMP33 and erythropoietin [8]. Atanasova and Whitty [61] noted this receptor-receptor contact surface occludes the binding site making it difficult to understand how erythropoietin or the EMPs bind to the preformed dimer in such a state.

Further difficulties with the scissor-like model for the EPOR have emerged over the last decade. First, the model assumes that it is the bringing together of the JAKs which leads to activation, while extensive studies on the closely related growth hormone receptor (GHR) have demonstrated that transphosphorylation is associated with the partial separation of the JAK2s which form a tight complex in the inactive state of the receptor [18]. Given the structural similarity between the extracellular domains of these two receptors and that signalling occurs through JAK2 in both cases, it is unlikely that inhibition and activation of the JAK2 is occurring through fundamentally different mechanisms. Indeed, Waters *et al.* [62] have created chimeras where the GHR extracellular domain has been replaced with the extracellular domain of the EPOR. These chimeras could activate GHR signaling in response to the agonist EMP1 but not the antagonist EMP33. This would suggest the transmembrane and intracellular regions of the type-I cytokine receptors are being activated by a conserved mechanism.

Second, in the structure of the apo dimer the D2 subdomains are widely splayed. This implies that the extracellular domains of the EPOR must self-associate strongly to maintain both this conformation and drive the formation of the inactive dimer on the surface of the cell. Experimentally it has been shown that the extracellular domains interact weakly and that the formation of the preformed dimers appears to be driven primarily by the interactions between the transmembrane domains [27]. In addition, mutagenesis studies in which cysteine residues were introduced into the transmembrane domains in order to crosslink the receptors by inducing disulfide bridges, have demonstrated that simply bringing the D2 subdomains together is insufficient for receptor activation [63].

2.1.3 Re-examination of the erythropoietin receptor crystal structures

A final difficulty with the scissor-like model for EPOR also becomes apparent if one re-examines the original structures on which the model was based. The projections of the two crystal structures shown in Figure 2.1, and in fig. 1 from Livnah *et al.* [8] give the impression that the transition between the two alternative arrangements of the receptor extracellular domains is associated with a motion of the transmembrane domains of approximately 3 to 4 nm. This was reinforced by the simple 2D schematic of the mechanism provided by Livnah *et al.* [8]. What is not captured in either of the projections or the original schematic is that the C-terminus of chain B lies on alternative sides of the chain A in the two structures. The differences between the arrangements of the receptors in the two structures can be readily seen in the alternative projections shown in Figure 2.2. From the side projections in Figure 2.2A it can be seen that in the proposed inactive apo dimer, the $D2_B$ subdomain projects behind and away from the $D2_A$ subdomain. In contrast, the same projection of the proposed active EMP1-bound dimer, shown in Figure 2.2B, has the $D2_B$ subdomain projecting to the front of the $D2_A$ subdomain. Furthermore, the $D2_B$ subdomain is not in direct contact with the $D2_A$ subdomain as suggested in the projection shown in Figure 2.1B. The significance of $D2_B$ subdomains lying on opposite sides of chain A is that these two alternative arrangements of the EPOR molecules cannot be generated by simply changing the distance between the D2 subdomains by just 3 to 4 nm through a scissor-like motion. This can be clearly seen in Figure 2.2C, which shows an overlay of the two structures fitted on chain A. As shown in Figure 2C the linear distance between the C-terminal residues (Thr220) of the $D2_B$ subdomains from the apo and EMP1-bound dimers is 8.3 nm. Assuming that the receptor complex could undergo a scissor-like motion, following the associated radial path required on the surface of the membrane would mean the transmembrane domains attached to the C-termini of the D2 subdomains would have to move a combined distance of over 13 nm through a membrane environment (Figure 2.3). Considering the relative size of the EMP1 to the EPOR extracellular domain and the attached cytosolic region and JAK2 molecules this scissor-like motion seems impossibly large.

2.2 Conclusions

When the scissor-like model for the activation of the EPOR was initially proposed over 15 years ago, it was novel and could explain much of the experimental data available at the time. Its rapid acceptance reflected the fact that it could reconcile the widely held assumption that transphosphorylation required the JAKs to be brought together and the suggestion that the receptors could reside on the surface of cells as inactive, preformed dimers. The crystallographic studies on which the model was based represented

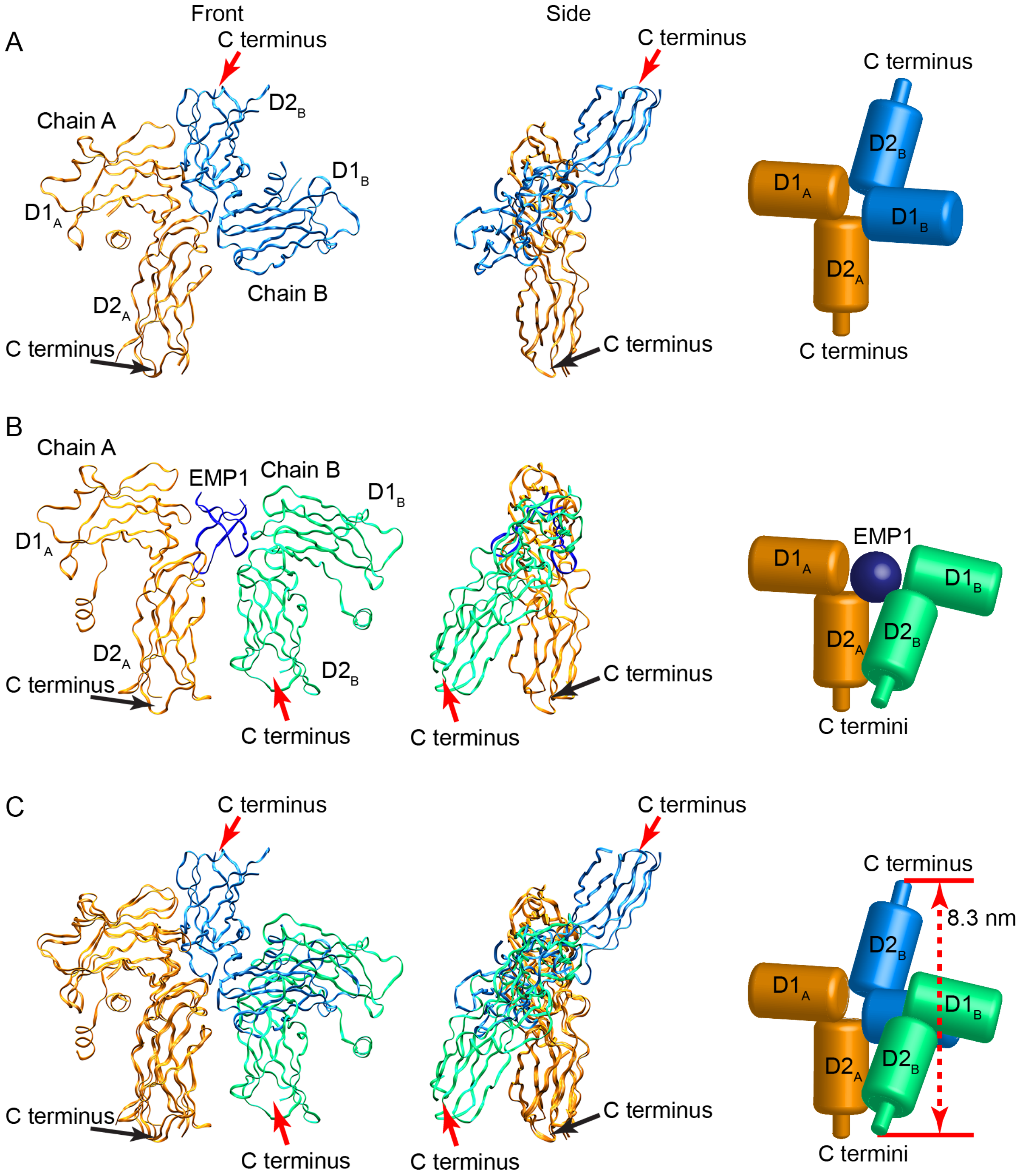


FIGURE 2.2: Front and side projections of (A) the apo EPOR dimer and (B) the EMP1-bound EPOR dimer after superimposition of the two structures onto chain A. The third panels of (A) and (B) are schematics of the rear and forwards projection of the D2$_B$ subdomain relative to chain A in the apo and EMP1-bound crystal structures respectively. (C) The overlay of the two structures shown in (A) and (B). The calculated linear distance between the α-carbon of Thr220 in the apo (blue) and EMP1-bound EPOR (green) chain Bs is 8.3 nm. The arrows indicate the membrane-proximal C-terminal ends of the D2 subdomains in each EPOR molecule. In (B) and (C) the EMP1 molecules have been omitted for clarity.

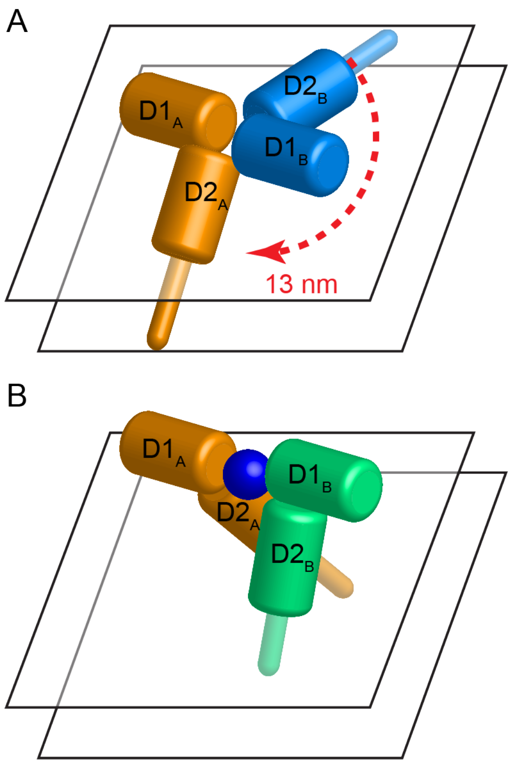
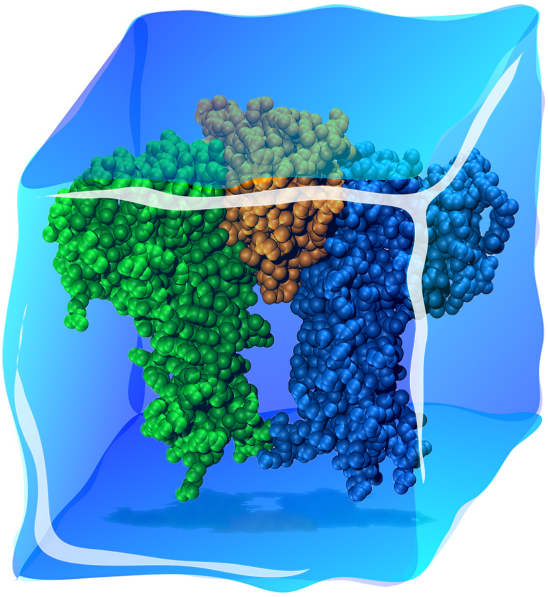


FIGURE 2.3: A 3D schematic of (A) the apo dimer transitioning via a scissor-like motion to (B) the EMP1-bound dimer in a membrane. The combined radial distance between the C-termini of the two structures is over 13 nm. The two planes indicate the surface of the membrane.

a major scientific advance and a unique source of high-resolution structural data. However, at the same time the model was highly idealised. As shown in Figure 2.2, the motions proposed do not directly account for the differences between the two crystal structures on which the model was based. In addition, one might question whether it should even be expected that the structures of the extracellular domains with and without cytokine truly represent active and inactive conformations of the receptor. Such an assumption not only implies that the transmembrane domains and the membrane itself play only a minor role in the organisation of the extracellular domains but also neglect the possible effects of crystal packing. The problem illustrated here is not in the proposal of the original model or the presentation of overly simplistic schematics. The problem is that despite an accumulation of evidence against the scissor-like model for the activation of the EPOR over the last decade [18, 27, 63], it has not been critically evaluated and revised in light of new findings and observations and continues to be propagated and used as a basis to interpret other systems [29]. We hope that by presenting the evidence above it can be seen that it is highly unlikely that the activation of EPOR by binding of cytokine occurs via a scissor-like model, and that this interpretation of the EPOR structures should not be used to justify proposing this model in respect to other receptor systems. More generally, our analysis of the EPOR system illustrates the potential pitfalls when proposing detailed mechanistic models based on the selective interpretation of crystal structures, especially in cases of protein complexes such as the EPOR, where only part of the protein is considered and the systems are not in a representative (membrane) environment.

Chapter 3

Do X-ray Crystal Structures of Protein Complexes with Agonistic and Antagonistic Ligands Truly Represent Active and Inactive Conformations? The Erythropoietin Receptor a Case Study.



3.1 Introduction

The erythropoietin receptor (EPOR) is a cell-surface protein receptor and a member of the type-I cytokine receptor family. This family also includes the growth hormone, prolactin, thrombopoietin and some interleukin receptors. As first shown in 1991 in relation to the growth hormone receptor (GHR), the activated complex for a type-I cytokine receptor is formed from the interaction of two independent receptor chains with a single cytokine (Figure 1.2) [14, 15, 33]. The formation of such an activated complex results in the phosphorylation of the receptor cytosolic domains at key tyrosine residues by a Janus kinase (JAK) and initiates one of the Signal Transducer and Activator of Transcription (STAT) pathways -- in the case of the EPOR, JAK2 and the STAT5 pathway. The full-length EPOR molecule comprises: (i) a ligand-binding extracellular domain (residues 1-226), (ii) a single-pass transmembrane domain (residues 227-249), and (iii) a cytosolic domain (residues 250-484). Currently, there are no atomic-resolution structures of the full-length EPOR. Instead, only X-ray crystallographic structures of the soluble EPOR extracellular domain (referred to as sEPOR, residues 1-226) in unbound [8] and ligand-bound states [5, 6, 7, 9], and NMR structures of the transmembrane domain in a micelle [10, 11] have been reported. Figure 1.4 shows the crystal structure of the EPO-bound (Figure 1.4A) and apo sEPOR dimers (Figure 1.4B). As can been seen, each sEPOR consists of two fibronectin type-III domains, labelled D1 (residues 1-117) and D2 (residues 122-226), that are joined by a four-residue linker region (residues 118-121) [5, 6, 7, 8].

Owing to the medical importance of erythropoiesis, the EPOR has been the target of considerable pharmaceutical development. A wide range of mimetic peptides with agonistic and antagonistic activity have been identified, some of which are in widespread use in the clinic. In contrast to EPO, which has a coiled-coil four-helix bundle structure, these EPO mimetic peptides (EMPs) are only 18-20 amino acids in length and adopt a hairpin structure consisting of two short antiparallel β-strands connected by a disulfide bridge close to the termini [5, 6]. The sEPOR dimer has been crystallised in the presence of two EMPs: the agonist EMP1 [5] and the antagonist EMP33 [6]. In both cases the crystal complex contains two EMP molecules that form a dimer, which in turn binds two sEPOR chains arranged as a quasisymmetric dimer (Figure 1.4C and D). Like EPO, the EMPs bind in between the two sEPOR protomers. Overall, the structures of the two EMP-bound complexes are similar. However, if the structures are superimposed using a single receptor D2 domain as a reference, there is a rotation of 15° of the second sEPOR protomer in EMP33-bound sEPOR complex relative to the EMP1-bound sEPOR [6]. An overlay of D1 domains from the EPO, EMP1, EMP33-bound and apo crystal structures after superimposition onto a single D1 domain is shown in Figure 1.4E. This variation in the orientation between the two EMP-bound sEPOR complexes is important to note as it was suggested that the difference in dimer configuration accounted for the difference in proliferative properties [6].

Based on the crystal structures of the apo dimer of sEPOR, dimerisation of the EPOR chains was suggested to occur via the EPO-binding sites in the extracellular domains with the termini of the D2 domains remaining separated (Figure 1.4B), thereby keeping the transmembrane and cytosolic domains apart [8]. Furthermore, based on differences in the relative arrangement between the two protomers and the distance between the two D2 domains between the apo and EMP1-bound sEPOR dimers, it was proposed that the activation of EPOR involved a cross-action scissor-like motion (Figure 1.3C) [8]. In this case, the two D1 domains move apart resulting in the two D2 domains being brought closer together (as seen in Figures 1.4A, C and D), allowing the interaction between cytoplasmic domains [8, 20]. However, our reanalysis in Chapter 2 of the scissor-like mechanism in the light of the differences between the apo and EMP1-bound sEPOR dimers has questioned whether this model is physiologically possible on the cell surface [1]. Specifically, the motions proposed do not directly account for the differences between the two crystal structures and lead to an unrealistic rearrangement within the EPOR dimer that neglects the role of the transmembrane domains and the membrane itself.

This leads to the more general question of whether the relative arrangement of the extracellular domains within the sEPOR dimer as observed crystallographically are truly representative of the complexes with agonists and antagonists free in solution, and indeed whether they are representative of active and inactive complexes more generally. To address this, atomistic molecular dynamics simulations have been used to examine the structure of the sEPOR dimer in the apo state, bound to EPO and bound to a range of agonistic and antagonistic EMPs in the absence of crystal packing effects.

3.2 Materials and Methods

3.2.1 Systems simulated

The sEPOR dimer (sEPOR$_2$) was simulated: (i) in presence of the wild-type erythropoietin (EPO) cytokine, (ii) in the absence of EPO, (iii) in a complex with a series of six EPO mimetic peptides (EMPs), and (iv) in the presence of the EPO mutant R103A. The coordinates of sEPOR$_2$ in the presence of EPO were taken from the crystal structure of the EPO-bound human sEPOR dimer (Protein Data Bank (PDB) entry 1EER) [7]. The structure of the apo form of the sEPOR dimer was derived in two ways. First, directly from the crystal structure of the apo sEPOR$_2$ (PDB entry 1ERN) and second from the sEPOR$_2$-EPO complex by removing EPO either from the initial crystal structure (PDB entry 1EER) or after 20 ns of simulation. These apo dimers are referred to as apo-sEPOR$_2$, apo-sEPOR$_2$-0 and apo-sEPOR$_2$-20, respectively. Coordinates of the sEPOR$_2$ with the mutant EPO R103A were also based on the PDB entry 1EER. The EPO R103A mutant was obtained by deleting the supernumerary side-chain atoms of residue 103 in the EPO molecule. The structures of the sEPOR$_2$ bound to two copies of the EMPs 1, 6, 7, 8 and 16 were derived from the crystal structure of sEPOR$_2$-EMP1$_2$ (PDB entry 1EBP) [5]. The coordinates of the structure of sEPOR$_2$ complexed with two EMP33 ligands were taken from the crystal structure of sEPOR$_2$-EMP33$_2$ (PDB entry 1EBA) [6]. Residues within the EMP ligands that were not observed crystallographically (residues G1, G2, G19 and G20 in EMP1, and G1, G2, G19 and G20 in EMP33) were modelled using PyMOL [64]. Residue 4 in EMP6, EMP7 and EMP8 (A4, T4 and F4, respectively) was substituted for Y4 in EMP1 using PyMOL. The sequences of the EMPs used in this study are provided in Table 3.1. The sEPOR chains in all systems simulated included all residues from D8 to P226 with the N- and C-termini acetylated or amidated respectively. Side chains or residues not observed crystallographically were modelled using PyMOL. Force field parameters for the 3,5-dibromotyrosine in EMP33 were obtained using the ATB [65]. Side chains and

Table 3.1: Amino acid sequence of the EMPs used in this study.

EMP	Sequence	Activity
1	`GGTYSCHFGPLTWVCKPQGG`	Agonist
6	`GGTASCHFGPLTWVCKPQGG`	Antagonist
7	`GGTTSCHFGPLTWVCKPQGG`	Agonist
8	`GGTFSCHFGPLTWVCKPQGG`	Agonist
16	`GGTYSCHFGPLTWVCKPQ`	Agonist
33	`GGTXSCHFGPLTWVCKPQGG`	Antagonist

X is 3,5-dibromotyrosine.
Sequence and activity data obtained from Livnah *et al.* (1998) [6].

Table 3.2: Overview of the systems simulated.

System	Receptor:ligand ratio	Initial PDB structure	Repeats	Simulation time (ns)
Crystal structures				
sEPOR$_2$-EPO	2:1	1EER	5	80
apo-sEPOR$_2$	2:0	1ERN	3	50
sEPOR$_2$-EMP1	2:2	1EBP	3	50
sEPOR$_2$-EMP33	2:2	1EBA	3	50
Derived structures				
apo-sEPOR$_2$-0	2:0	1EER	3	50
apo-sEPOR$_2$-20	2:0	[a]	3	50
sEPOR$_2$-EPO R103A	2:1	1EER	3	80
sEPOR$_2$-EMP6	2:2	1EBP	3	50
sEPOR$_2$-EMP7	2:2	1EBP	3	50
sEPOR$_2$-EMP8	2:2	1EBP	3	50
sEPOR$_2$-EMP16	2:2	1EBP	3	50

[a] The initial coordinates for the sEPOR$_2$-20 system were derived from the configuration of the sEPOR$_2$-EPO simulation 2 at 20 ns.

the N- and C-termini of the ligands (EPO and EMPs) were protonated as appropriate for pH 7. Each system was then placed in a box of water and sufficient Na$^+$ and Cl$^-$ ions were added to neutralise the system and achieve a physiological salt concentration (150 mM). The ions were added by selecting water molecules at random and replacing them by either Na$^+$ or Cl$^-$ ions. The systems examined in this work are summarised in Table 3.2.

3.2.2 Simulation parameters

All simulations of the receptor complex in solution were performed using the GROMACS simulation package 3.3.3 [49] in conjunction with the GROMOS 54A7 united-atom force field [48]. Each system was simulated under periodic conditions in a truncated octahedral box. The pressure was coupled to an isotropic pressure bath and maintained at 1 bar by weakly coupling the system to a pressure bath [53] using an isothermal compressibility of 4.6×10^{-5} bar^{-1} and a coupling constant $\tau_P = 1$ ps. The temperature of the system was maintained by independently coupling the protein and water together with ions to an external temperature bath at 298 K with a coupling constant $\tau_T = 0.1$ ps using a Berendsen thermostat [53]. All bond lengths within the protein were constrained using the LINCS algorithm [66]. Water was included explicitly in the simulations using the Simple-Point Charge (SPC) model [67] and constrained using the SETTLE algorithm [68]. Nonbonded interactions were evaluated using a twin-range cutoff scheme: interactions falling within the 0.8-nm short range cutoff were calculated every step whereas interactions within the 1.4-nm long cutoff were updated every three steps, together with the pair list. A reaction-field correction was applied to the electrostatic interactions beyond the long-range cutoff [69], using a relative dielectric permittivity constant of $\epsilon_{\mathrm{RF}} = 62$ as appropriate for SPC water [70]. Polar hydrogen atoms in the protein were replaced by virtual interaction sites, the positions of which were constructed at each step from the coordinates of the heavy atoms to which they are attached as described by Feenstra *et al.* [71]. This allowed a 4-fs time step to be used without affecting the thermodynamic properties of the system significantly. All systems were energy-minimised. The temperature of each system was then gradually increased from 50 K to 298 K in 50-K steps over 120 ps to further relax the system and obtain the starting configurations used in the simulations. All systems were simulated three times starting from different initial velocities, except sEPOR$_2$-EPO which was simulated 5 times.

3.2.3 Analysis

3.2.3.1 Root mean square deviation

The root mean square deviation (RMSD) of the coordinates of the backbone with respect to a reference structure was calculated after performing a least-squares fit of the backbone atoms to the reference structure. Due to the extended nature of the N-terminal region of the sEPOR protomers in the initial EMP-bound structures, only residues 27-220 were used for the RMSD calculation for these structures.

3.2.3.2 Conformations of the apo and ligand-bound sEPOR dimers

The overall relative arrangement of the sEPOR protomers within the sEPOR$_2$ was characterised by the angle θ between the sEPOR protomers and the distance d between the D2 domains. θ was calculated as the angle between the normals of the planes representing each of the sEPOR protomers. The planes were defined by the centres of mass of the backbone atoms in D1 and D2 domains and the four-amino acid-hinge region (amino acids Val118-Leu121) at each time point. d was calculated as the distance between the centres of mass of the backbone atoms in D2 domains at each time point.

3.3 Results

3.3.1 Structure of EPOR in the apo and bound dimers

Changes within the structure of the individual receptor chains in the sEPOR$_2$ dimers bound to EPO, EMP1, EMP33 and in its apo form were examined by calculating the root mean square deviation (RMSD) of the atomic positions of the backbone atoms with respect to their corresponding initial crystal structure. An overlay of the backbone trace of the chain A and chain B structures from the simulations for each of these systems superimposed onto the four-residue linker is shown in Figure

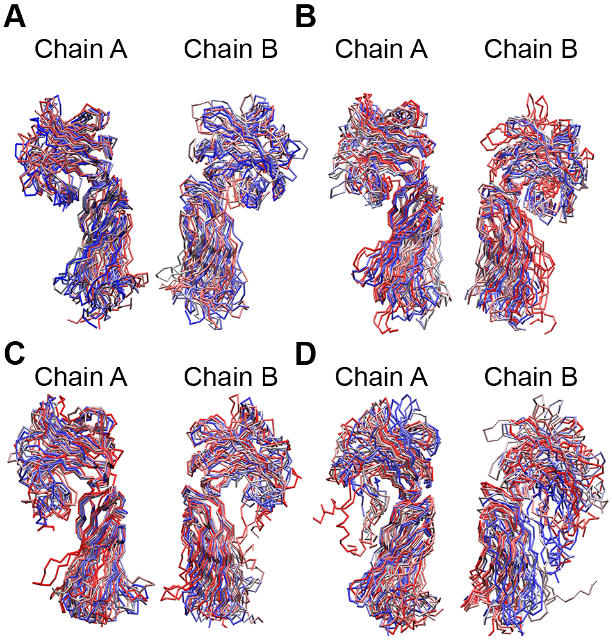


FIGURE 3.1: Overlays of the individual sEPOR chain conformations obtained from all simulations of $sEPOR_2$-EPO (A), apo-$sEPOR_2$ (B), $sEPOR_2$-EMP1 (C) and $sEPOR_2$-EMP33 (D). In each case the chain has been superimposed onto the backbone atoms of the hinge residues Val118-Leu121.

3.1. Overall, the structure of the individual sEPOR chains underwent little change during any of the simulations indicating that the structure of the D1 and D2 domains and their relative arrangement were stable and relatively rigid. The average RMSD values for chains A and B for each simulation are listed in Table 3.3. Regardless of the presence or absence of a ligand, or the nature of the ligand, the RMSD of the backbone typically ranged between 0.26 and 0.41 nm. Such variation is not unexpected considering the individual chains are elongated and consist of two sub-domains, and that the exposed termini are flexible. Note, in one replicate of the EMP33-bound dimer chain B displayed greater fluctuations within the chain in part because of the N-terminal region not folding under the D1 domain as in all other simulations, leading to a poorer superimposition of the chain conformations (Figure 3.1D) and a higher average RMSD of 0.61 nm.

Changes in the structure of the $sEPOR_2$ dimers were investigated by calculating the RMSD over the entire receptor dimer backbone after first superimposing onto chain A in the respective initial crystal structure as a reference. The time evolution of the overall RMSD for the dimer in each simulation is shown in Figure 3.2. The RMSD varied between 1.0 and to 3.0 nm. These larger values arose from changes in the relative orientations of the receptor chains. There were also marked differences between the four systems. Whereas the apo dimer (Figure 3.2B) remained close to the starting crystal structure, the ligand-bound $sEPOR_2$ dimers varied substantially between the different runs, especially in the case of the EPO-bound dimer (Figure 3.2A). The sharp increase in the RMSD of up to 3.0 nm at around 30-40 ns in a simulation of the EMP1-bound system (Figure 3.2C) was caused by a relative rotation of the chain B around the chain A that eventually rotated back leading to a relative orientation of the two chains that produced RMSD values similar to those observed in the two other simulations. As noted earlier for the individual chains, in one of the replicates involving the EMP33-bound system (Figure 3.2D), fluctuations of the N-terminal region caused fluctuations within the chain which in turn

Table 3.3: Atomic position root mean squared deviation (RMSD) values of the individual chains in the sEPOR$_2$.

		RMSD (nm)	
		Chain A	Chain B
sEPOR$_2$-EPO	1	0.41	0.37
	2	0.41	0.32
	3	0.41	0.35
	4	0.41	0.34
	5	0.40	0.30
	average	*0.41±0.06*	*0.34±0.05*
apo-sEPOR$_2$	1	0.28	0.28
	2	0.29	0.31
	3	0.26	0.32
	average	*0.28±0.04*	*0.30±0.05*
sEPOR$_2$-EMP1	1	0.28	0.31
	2	0.33	0.29
	3	0.30	0.28
	average	*0.30±0.06*	*0.29±0.04*
sEPOR$_2$-EMP33	1	0.27	0.32
	2	0.30	0.41
	3	0.30	0.61
	average	*0.29±0.05*	*0.44±0.16*

Values are calculated over all the backbone atoms for the apo and EPO-bound protomers, or from residue L27 for the EMP-bound protomers, after a least squares fit to the backbone atoms in the respective chain of the crystal structure.

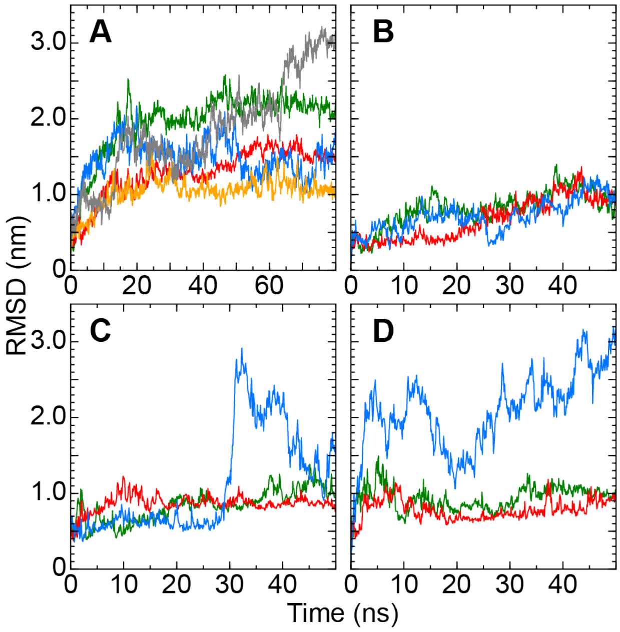


FIGURE 3.2: RMSD values for the sEPOR$_2$ dimer calculated over the backbone atoms after first performing a least squares fit onto the backbone atoms of the chain A. sEPOR$_2$-EPO (A), apo-sEPOR$_2$ (B), sEPOR$_2$-EMP1 (C) and sEPOR$_2$-EMP33 (D). Independent simulations 1-5 are colored green, red, blue, orange and grey.

led to a large overall RMSD.

3.3.2 Changes in the crystal structures of the apo, EPO-, EMP1- and EMP33-bound sEPOR dimers in solution

The conformations adopted by the sEPOR$_2$ dimers in solution were examined in terms of the relative arrangement of the two protomers within each dimer by monitoring an angle θ and a distance d between the two chains (Figure 3.3A). Each sEPOR chain was represented as a plane defined by the centres of mass of the D1 and D2 domains and the centre of mass the four linking hinge residues (Val118-Leu121). The angle between these planes (θ) and the distance (d) between the centres of mass for the D2 domains were calculated for each trajectory and compared.

In four out of the five simulations of the EPO-bound dimer, θ sampled angles between 70-180° and d varied between 2.0-3.6 nm (Figure 3.3B). Overall there was an increase in θ and decrease in d from the initial values of 105° and 3.0 nm, respectively (indicated by a black dot in the panel). In the fifth simulation (grey dots in Figure 3.3B), d increased to over 5.6 nm and θ fell to between 60-90°, resulting in a separation of the D2 domains while the D1 domains remained associated. In the apo dimer (Figure 3.3C), d stayed close to its initial value of 4.5 nm (3.9-4.9 nm) whereas θ ranged from 10 to 100°. In the case of the agonistic EMP1- and antagonistic EMP33-bound sEPOR$_2$ dimers (Figure 3.3D and E, respectively), the initial values of d in the crystal structures were similar (2.9 nm) but θ differed slightly (140° in sEPOR$_2$-EMP1 and 120° in sEPOR$_2$-EMP33). In the simulations, d sampled a range of values (1.9-4.1 nm in sEPOR$_2$-EMP1 and 2.1-3.3 nm in sEPOR$_2$-EMP33) and θ varied between 70-180° in both systems.

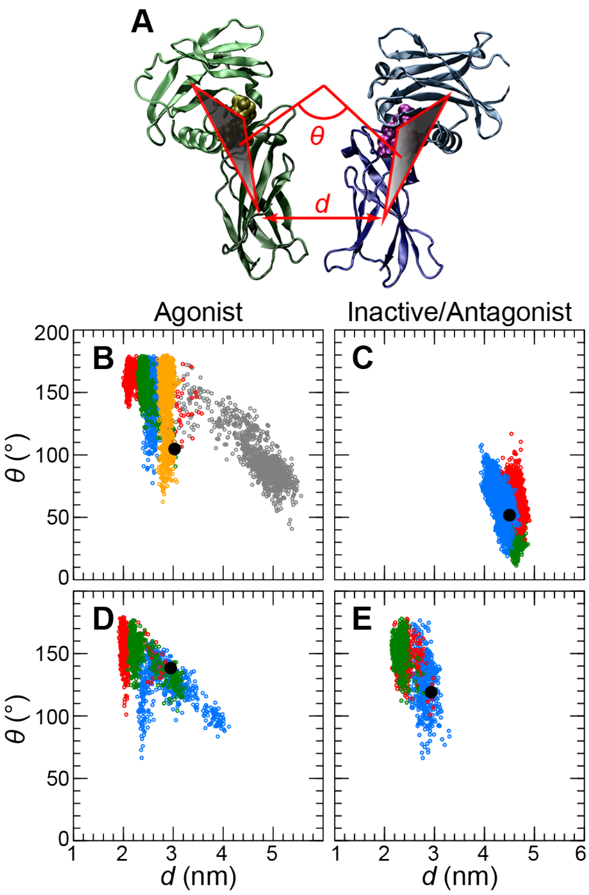


Figure 3.3: (A) The overall conformations of the sEPOR chains within the dimer were determined in terms of an angle between the chains (θ) and the distance between the centre of mass of the D2 domains (d). (B)-(E) scatter plots of θ against d for simulations of the sEPOR$_2$-EPO (B), sEPOR$_2$ (C), sEPOR$_2$-EMP1 (D) and sEPOR$_2$-EMP33 (E) starting from the crystal structures . Simulations 1-5 are coloured green, red, blue, orange, and grey respectively. The black dots represent the values of the initial conformations.

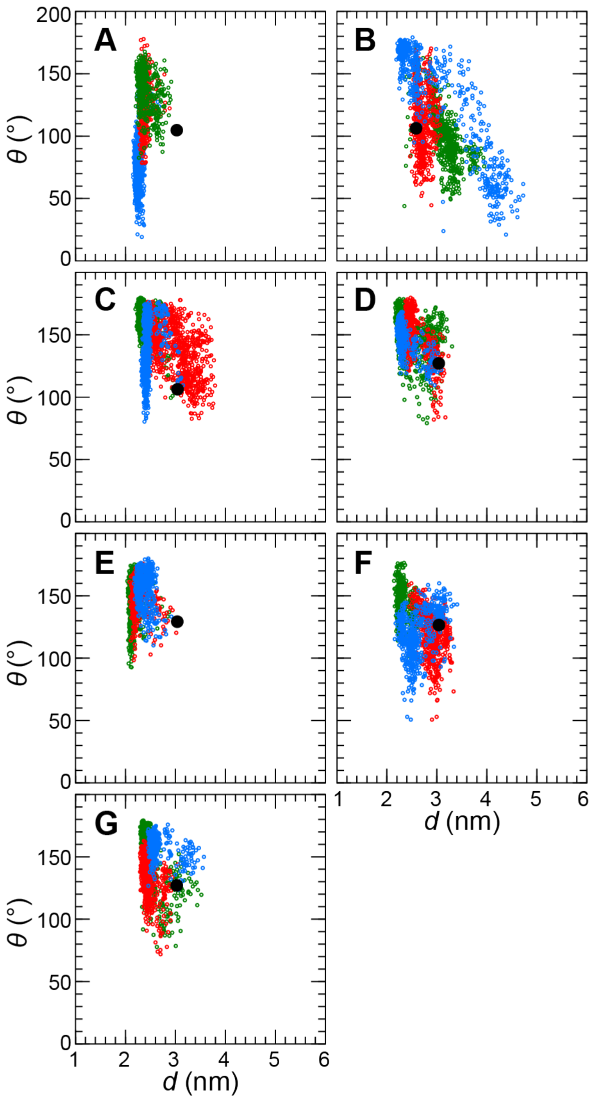


Figure 3.4: The overall conformation of the sEPOR chains in terms of θ and d (see Figure 3.3A) for the systems sEPOR$_2$-0 (A), sEPOR$_2$-20 (B), sEPOR$_2$-EPO R103A (C), sEPOR$_2$-EMP6 (D), sEPOR$_2$-EMP7 (E), sEPOR$_2$-EMP8 (F) and sEPOR$_2$-EMP16 (G). Simulations 1-3 are coloured green, red and blue respectively. The black dots represent the values of the initial conformations.

3.3.3 Changes in the structures of the sEPOR dimers in solution derived from the crystal structures

The variations in d and θ were also examined in a series of simulations involving alternative active and inactive complexes derived from either crystal structures or previous simulations. Specifically, two further apo forms of the sEPOR$_2$ were generated by removing the EPO from the EPO-bound sEPOR$_2$ in either the initial crystal structure (Figure 3.4A) or after 20 ns of simulation (Figure 3.4B). In addition, sEPOR$_2$ bound to a range of other agonistic and antagonistic ligands was also investigated. These included the antagonistic R130A mutant of EPO (Figure 3.4C), the antagonistic mimetic EMP6 (Figure 3.4D), and the agonistic EMP7 (Figure 3.4E), EMP8 (Figure 3.4F) and EMP16 (Figure 3.4G).

In the simulations of the apo complexes generated from the holo (EPO-bound) crystal structure (apo-sEPOR$_2$-0 and apo-sEPOR$_2$-20), θ ranged between 20-180° in both cases. In contrast the value of d varied significantly depending on the origin of the starting structure. In particular, in the apo form derived from the EPO-containing complex after the system had been relaxed for 20 ns (apo-sEPOR$_2$-20), d markedly increased reaching about 4.7 nm after 50 ns. This effectively meant the two sEPOR chains had disassociated (Figure 3.4B). In the case of the other bound forms, θ sampled a similar range of values in all simulations (60-180°) (Figure 3.4C-G). Likewise, d varied between 2.2 and 3.4 nm in these simulations except in a simulation of the EPO R130A mutant-bound dimer where d increased up to 3.8 nm, which again involved the disassociation of chain B from chain A (Figure 3.4C).

3.4 Discussion

The erythropoietin receptor (EPOR) is a primary regulator of erythropoiesis but has also been recently identified in neurons where it helps protect against brain injury [72]. Its mechanism of activation, in particular how the binding of a ligand on the extracellular domains is transmitted through the plasma membrane, has therefore been of prime interest for pharmaceutical development. A critical question in the activation of EPOR is whether differences in the relative arrangement of the soluble extracellular domains within the sEPOR dimers observed in the available crystal structures are truly indicative of the active or inactive states as has been widely assumed [5, 6, 7, 8, 22, 9].

When the crystal structures of the apo and EMP1-bound forms of the soluble extracellular domain of EPOR (sEPOR) were solved initially, it was proposed that distances between the termini observed crystallographically were distances that occurred on the cell surface. In this model, ligand binding would separate the D1 domains and bring the D2 domains close together (Figure 1.3C). This would then allow interactions between the cytosolic domains of the EPOR molecules and the JAK kinases [8]. However, this model was proposed without any discussion of how the transmembrane domains might restrain the extracellular domains when embedded in a membrane. Not only has it been claimed that the precise distance between the terminal regions is crucial to the activity state of the sEPOR dimers, but also the angle formed between the receptor chains [5, 6, 7, 8]. An overlay of the sEPOR dimers showing the relative arrangement of the D1 domains is displayed in Figure 1.4E. When the dimers are fitted onto the first same chain, the second chain in the EMP33-bound dimer is rotated by 15° with respect to the second chain in the EMP1-bound dimer [6]. The difference in this angle was claimed to account for the different activity of the ligand despite the distance of the C-termini between the sEPOR chains being practically identical. This would suggest that not only are the individual domains rigid but that the binding of either EPOR, EMP1, EMP33 or indeed any of the known agonists or antagonists would need to hold the two receptor chains in a very precise relative orientation and to generate sufficient mechanical force to separate (or bring together) the transmembrane helices and the two associated JAK2 domains that consist of over 1000 amino acids.

In the simulations presented here, the overall structure of the individual sEPOR molecules were maintained with average backbone RMSD values between 0.29-0.44 nm, in both the EPO-bound and unbound (apo) sEPOR dimer (Table 3.3). This is in line with the range of RMSD values (average RMSD of 0.25 nm$\pm$0.11 nm) between the different structures of the sEPOR obtained crystallographically, including the apo dimer [8], the dimer bound to EPO [7], EMP1 [5], EMP33 [6], a single sEPOR molecule bound to an antibody [73] or a diabody [9]. This indicates that the range of motion seen within the individual receptor molecules in the simulations is in line with the variations observed crystallographically. The range of motion observed in the simulations is also consistent with experimental studies that showed that the secondary structure of the receptor molecules in the EPOR dimer was not changed upon EPO binding using circular dichroism [30].

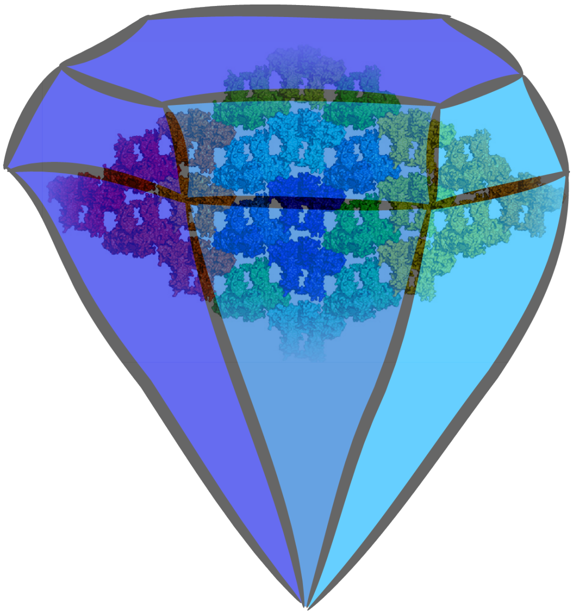
The simulations were also analysed in terms of the angle θ between the chains and the distance d between the two D2 domains. The aim was to determine whether the differences considered characteristic of activity based on the crystal structures were maintained in solution. It was found that the range of values of θ sampled by the dimer was similar across all simulations regardless of the nature (agonist or antagonist) of the ligand bound to the sEPOR dimer. In general, the D2 domains moved close together in the simulations, except in three cases (one agonist bound, one apo and one antagonist bound) where these domains separated (Figures 3.3B, 3.4B and C). As the EPO-binding site is occluded in the apo sEPOR dimer crystal structure, two further apo systems derived from EPO-bound conformations were investigated to determine if the structure of the apo dimer as observed in the crystal structure could spontaneously form in solution. In neither apo systems (apo-sEPOR$_2$-0 or apo-sEPOR$_2$-20) did the two sEPOR protomers relax to a relative arrangement close to that seen in the crystal structure. In fact, the states sampled in the simulations were not obviously dependent on the nature or the presence of the ligand. They were however, heavily influenced by the initial coordinates. This suggests that for these complexes the energy surface is relatively flat. Certainly it was not possible to determine any structural differences that would support any of the potential models of receptor activation. Indeed, in one simulation of the sEPOR$_2$-EPO system, the D2 domains separated, but the complex remained in contact via the D1 domains. This suggests the membrane, transmembrane domains and intracellular regions all potentially play essential roles.

3.5 Conclusion

X-ray crystallography remains the primary technique for determining protein structures with almost 90% of the structures deposited with the PDB being solved using crystallographic techniques. The aim of this study is not to cast aspersion on any specific structure but simply to highlight the difficulty and pitfalls when attempting to infer how a receptor or other protein may operate when only considering the structure of a protein within a crystal lattice. This is of particular importance in the case of membrane-associated proteins such as the EPOR and other cytokine receptors, where only the soluble extracellular domain is considered. In this case differences observed in the crystal were not maintained in solution. All of the systems examined readily sampled a broad range of structures. In short the differences observed in the crystals are unlikely to be physiologically significant. Clearly in the case of the EPOR the influence of the transmembrane domains and the membrane environment must be considered when attempting to propose a mechanism of activation. Recently, a series of crystal structures were published which suggest alternative arrangements of the sEPOR dimer bound to diabodies with a range of proliferative activity [9]. Despite not being attached to the transmembrane domain, much significance was placed on the distances in the domain D2 C-termini of these structure that were separated over 13 nm. Such large distances are in stark contrast to the values seen with the native ligand EPO and the small EMPs and would imply a mechanism of action very different from that proposed for closely related systems such as the growth hormone receptor. In light of the simulations presented here the physiological significance of the arrangement observed in these complexes should be treated with due caution.

Chapter 4

Can the Models of the X-ray Crystal Structures for the Erythropoietin Receptor Reproduce the Crystal Lattice?



4.1 Introduction

Protein X-ray crystallography is an invaluable technique in structural biology with over 90% of the structures deposited with the Protein Data Bank having been solved using this technique. The method requires high-quality protein crystals that are able to produce sharp, discrete spots in an X-ray diffraction pattern from an X-ray beam. From these diffraction patterns structure factors are generated and used to estimate the electron density [74]. If the resolution of the data is sufficiently high the placement of atoms into the electron density is unambiguous. However, using lower resolution data, which may be due to intrinsic or static disorder within the crystal lattice, the placement of atoms becomes less certain. In areas where the electron density cannot be determined atoms are frequently omitted from the final structure. The quality and reliability of a model produced from X-ray crystallography is directly related to the confidence with which atoms can be placed within the electron density. In order to validate the model quality, both global and local structure indicators are required. The global quality can be inferred from values such as the highest resolution shell, which is the smallest lattice spacing sampled and therefore the smallest distance or detail that can be resolved in the electron density, and the R-*free* value which is a statistical method to avoid over-fitting of the model. Local structure quality indicators check predominantly for geometric outliers from historic knowledge, e. g. plausibility of backbone torsion angles by checking for Ramachandran outliers as well as bad contacts (clash scores) [74].

In the case of the crystal structures that contain the soluble erythropoietin receptor (sEPOR) the models deposited to the Protein Data Bank are of reasonable resolution (1.90-2.80 Å) for protein complexes containing 3-4 separate protein/peptide chains and totalling between 500-600 amino acids in the asymmetric unit. However, the validation reports for these structures produced by the Protein Data Bank have scores in the lower percentiles in regard to the clashscore, Ramachandran outliers and side chain outliers (Figure 4.1). These reports would suggest that the local geometry of the proteins is not ideal and could be improved. That is, there may be alternative ways to place the atoms that could provide a better agreement with the structure factors. Visualisation of why these structures containing sEPOR have such poor validation reports can be seen when the X-ray crystal structures are coloured according to the recorded *B*-factor in the respective PDB entries (Figure 4.2). It becomes clear in the case of the sEPOR$_2$-EPO complex that a number of surface exposed residues, particularly in loop regions, have a high atomic displacement indicating these regions are less well defined (Figure 4.2A). This is less obvious in the sEPOR$_2$-EMP1 complex where the *B*-factors have been scaled and are dominated by the modeled loop regions for which no density was observed (Figure 4.2B) [5].

In both the crystal structures of the sEPOR$_2$-EMP1 and sEPOR$_2$-EMP33 complexes (PDB entries 1EBP and 1EBA respectively) the residues of the sEPOR that connect the N-terminal α-helix to the D1 domain (residues 21-23) have been modelled using low or no density [5, 6]. However, the authors noted that density was observed for these residues in an unpublished structure of the sEPOR with another EMP [75]. The trace obtained for the unpublished structure was used to model the chain for these residues in the sEPOR$_2$-EMP1 and sEPOR$_2$-EMP33 complexes. This led in both cases to the N-terminal α-helix being modelled as a domain swapped dimer with a neighbouring asymmetric unit (Figure 4.3) [75]. In the sEPOR$_2$-EPO complex crystal structure this α-helix is folded under the

D1 domain (see the sEPOR chains in Figure 1.4A). This immediately raises the question on whether these residues have been appropriately modelled in these complexes. It is essential to know when using these models for molecular dynamics (MD) that the motions observed in the simulations may be physiologically relevant and not just a result of spurious forces due to the inappropriate placement of atoms in the X-ray model.

A method to investigate the quality of these X-ray crystal structures containing the sEPOR, that simultaneously validates the use of MD simulations to study these proteins, is to recreate the crystal lattice of these crystal structures. The lattice effect can be achieved by simply creating the unit cell and combining it with periodic boundary conditions (Figure 1.7). By using this approach the protein conformations are effectively kept restrained in the crystalline environment without artificial restraints. This would allow identification, if any, of regions in the protein that are under strain owing to large forces; these large forces would be an indication of atoms being inappropriately modelled, for example from electrostatic interactions that are not considered during structure refinement.

In turn, if the crystal lattice remains stable, it would indicate that the simulation conditions, including the force field, are not the predominant driver of the motions observed in MD simulations. However, in order to validate the approach and rule out any bias or artefacts due to the simulation conditions, the force field and system of interest and their combinations, an additional, independent system needs to be tested. As a test case, the protein bovine pancreatic trypsin inhibitor (BPTI) will be used for this purpose as its available crystal structure (PDB entry 4PTI) has a high resolution (1.5 Å) and the PDB validation report has favourable scores (Figure 4.1E) [76]. Here we aim to determine the reliability of the models that contain the sEPOR derived from X-ray crystallography data. In parallel we will test the ability of the GROMOS54A7 force field to reproduce a protein crystal lattice.

4.2 Methodology

4.2.1 Systems simulated

The X-ray crystal structures of the BPTI (PDB entry 4PTI) and the sEPOR in complex with erythropoietin (EPO) (PDB entry 1EER) or EMP1 (PDB entry 1EBP) were simulated within a crystal lattice. Any missing atoms were modelled using PyMOL [64] to make the protein complete as recorded under the respective SEQRES entries in the respective PDB files. The sEPOR$_2$-EMP1 was also simulated with the N-terminal α-helix modelled folded under the D1 domain of the chain to which it is connected, similar to how it is folded in the sEPOR$_2$-EPO structure. This was achieved by taking the coordinates of the N-terminal α-helix (residues 1-20) from both the neighbouring asymmetric units and modelling the residues in the loop (residues 21-23) connecting the helix to the D1 domain. The coordinates of the complete asymmetric unit in conjunction with the recorded crystallographic symmetry operations were used to generate the unit cell in PyMOL. In all cases the unit cell was duplicated and translated in one dimension to prevent an asymmetric unit being in contact with its periodic image. For this set of proteins this resulted in eight copies of the asymmetric unit within a super-unit cell, see for example Figure 4.4. Side chains and the N- and C-termini were protonated as appropriate for pH 7.

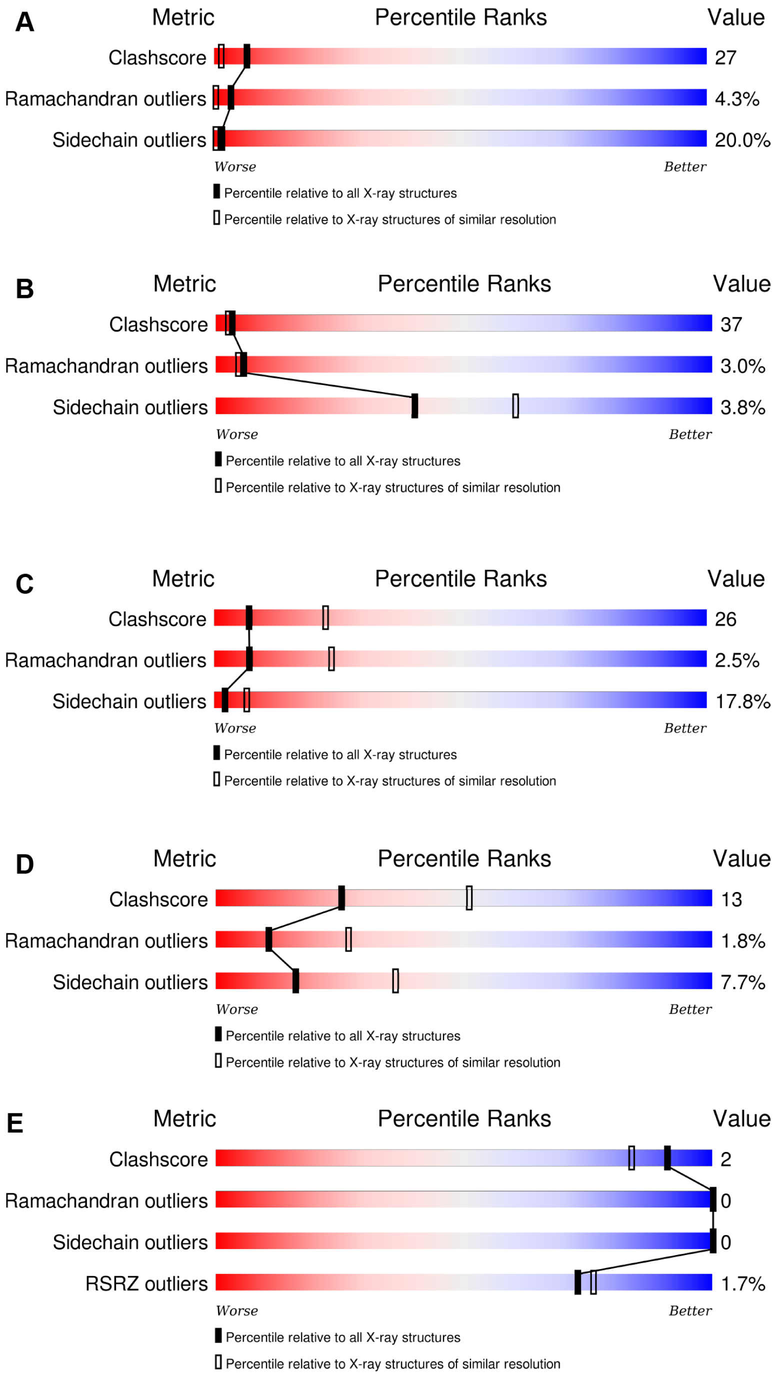


Figure 4.1: Validation reports from the Protein Data Bank for entries 1EER (sEPOR$_2$-EPO) (A), 1ERN (apo-sEPOR$_2$) (B), 1EBP (sEPOR$_2$-EMP1) (C), 1EBA (sEPOR$_2$-EMP33) (D) and 4PTI (BPTI) (E).

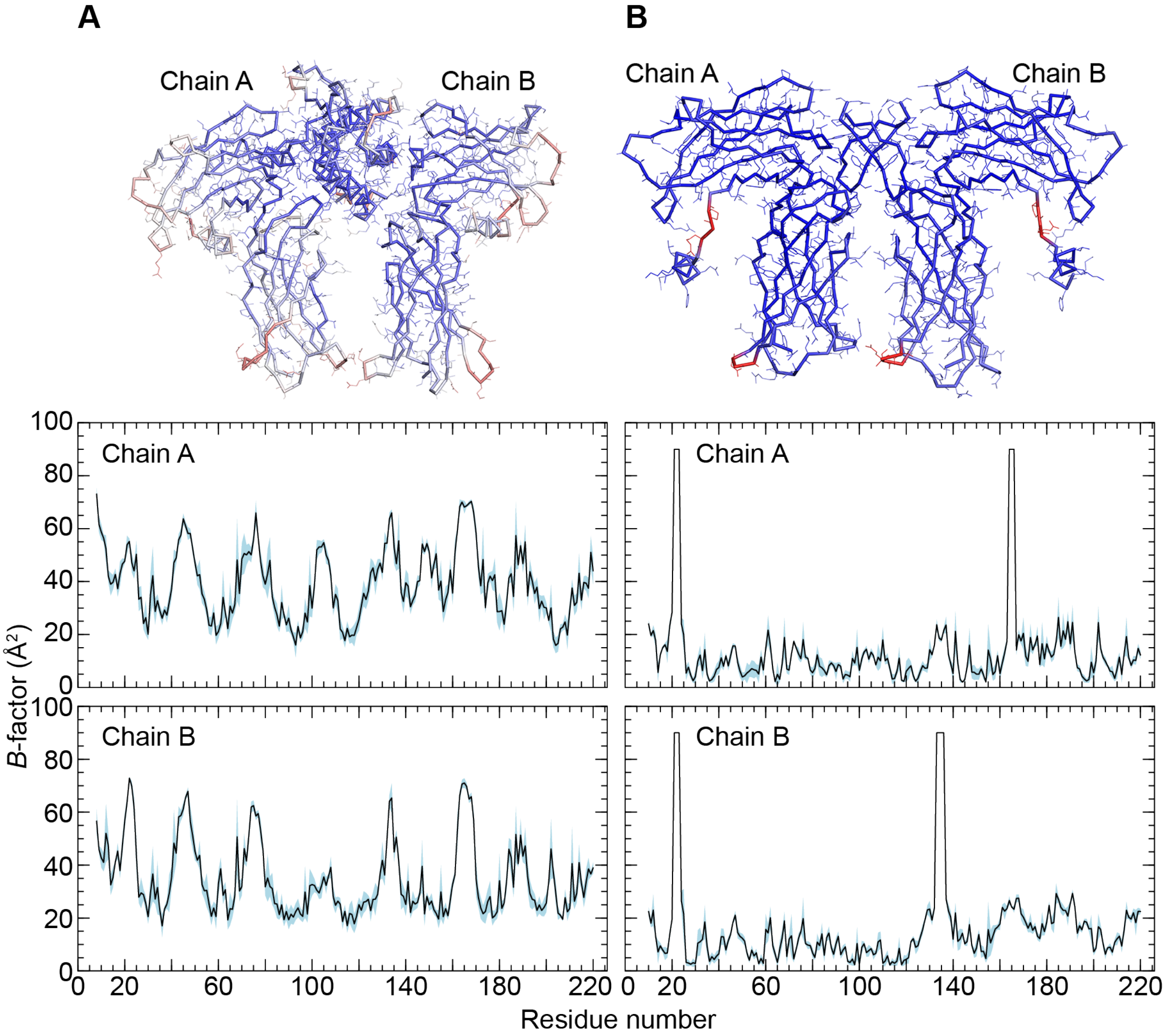


FIGURE 4.2: The X-ray crystal structures and average B-factor per residue for the sEPOR chains in the dimers bound to EPO (A) and EMP1 (B) (PDB entries 1EER and 1EBP respectively). Structures have been coloured according to B-factors reported in the respective PDB entries. Blue, 0 Å^2; red, 100 Å^2.

FIGURE 4.3: The N-terminal α-helix of the sEPOR is modelled as a domain swapped dimer in the sEPOR$_2$-EMP1 complex crystal. The loop consisting of residues 21-23 (yellow) did not have observable electron density [5].

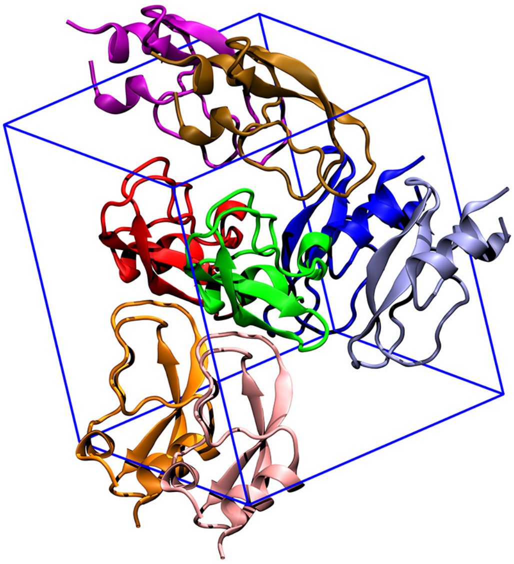


Figure 4.4: The super-unit cell of the BPTI consisting of eight replicates of the BPTI protein.

Each system was then placed in a box of water and sufficient Na^+ and Cl^- ions were added to obtain 100 mM and to neutralise the system . The ions were added by selecting water molecules at random and replacing them by either Na^+ or Cl^- ions.

4.2.2 Simulation parameters

All simulations were performed using the GROMACS simulation package 4.6 [50] in conjunction with the GROMOS 54A7 united-atom force field [48]. Each system was simulated under periodic boundary conditions in a rectangular box, the initial dimensions of which were obtained from the unit cell data in the respective PDB file. The temperature of the system was maintained by independently coupling the protein and, water and ions to an external temperature bath at 298 K with a coupling constant $\tau_T = 0.1$ ps using a Berendsen thermostat [53]. Pressure was maintained at 1 bar by weakly coupling the system to an anisotropic pressure bath [53] for using an isothermal compressibility of 4.6×10^{-5} bar^{-1} and a coupling constant $\tau_P = 1$ ps. All bond lengths within the protein was constrained using the LINCS algorithm [66]. Water was included explicitly in the simulations using the Simple-Point Charge model (SPC) [67] and constrained using the SETTLE algorithm [68]. Hydrogen atoms in the protein were replaced by virtual interaction sites, the positions of which were constructed at each step from the coordinates of the heavy atoms to which they are attached as described by Feenstra *et al.* [71]. This allowed a 4-fs time step to be used without affecting the thermodynamic properties of the system significantly. The Verlet cutoff scheme [52] was used with a minimum cutoff of 1.4 nm for short-range Lennard-Jones interactions and the real-space contribution to the smooth Particle Mesh Ewald algorithm using a fast Fourier transform grid spacing of 0.16 and an interpolation order of 6 [77, 78], which was used to compute long-range electrostatic interactions. A lattice sum method was used to better represent the crystalline environment.

4.2.3 System equilibration

Systems were energy-minimised and equilibrated under constant pressure with position restraints on Cα atoms that were observed in the electron density. Water molecules were added or removed during equilibration to obtain between 1-100 bar of pressure along the X, Y and Z-axes. Position restraints were initially imposed with an energy of 50 kJ $\cdot$ mol^{-1} $\cdot$ nm^{-2} and gradually removed over a period of 14 ns for BPTI; 6-8 ns for sEPOR$_2$-EMP1-crystal system or 18 ns for the sEPOR$_2$-EPO-crystal. In the case of the sEPOR$_2$-EMP1 system with the remodelled N-terminus, after an initial relaxation with position restraints the modelled atoms in residues 21-23 were gradually grown from soft-core potentials over a period of 0.2 ns with position restraints of 500 kJ$\cdot$mol^{-1} $\cdot$ nm^{-2} on the non-modelled atoms. After all position restraints were removed, the system was coupled to an anisotropic pressure bath. The final 2.5 ns of the unrestrained simulations was used in the analysis.

4.2.4 Analysis

4.2.4.1 Pressure

The pressure tensor P is calculated from the difference between the kinetic energy (E_{kin}) and the virial (Ξ):

$$P = \frac{2}{V}(E_{kin} - \Xi) \tag{4.1}$$

where V is the volume of the box. The virial (Ξ) tensor is defined as:

$$\Xi = -\frac{1}{2}\sum_{i<j} \mathbf{r}_{ij} \otimes \mathbf{F}_{ij} \tag{4.2}$$

4.2.4.2 Root mean square deviation

The root mean square deviation (RMSD) of the coordinates of the backbone atoms with respect to a reference structure was calculated after performing a least-squares fit of the backbone atoms to the reference structure. Only residues that were modelled in the respective PDB entries and had observed electron density were considered for the RMSD calculation. To calculate the RMSD per residue the last 2.5 ns of simulation for each asymmetric unit was first extracted and concatenated into a single trajectory. Each frame was then aligned by a least-squares fit of the backbone atoms to the reference structure and the RMSD per residue was calculated and averaged over all frames.

4.2.4.3 *B*-factors

The B-factors were calculated using Equation 4.3:

$$B = \frac{8\pi^2}{3}\langle u^2 \rangle \tag{4.3}$$

where u is the mean squared displacement of the atom from the reference structure. In all cases the displacement was calculated from the position of the atom in the X-ray crystal structure.

4.2.4.4 Conformations of the sEPOR dimers

The overall relative conformations adopted by the $sEPOR_2$ dimers in the crystals were characterised by the angle θ between the sEPOR protomers and the distance d, between the D2 domains (see Figure 3.3A). The angle θ was calculated as the angle between the normals of the planes representing each of the sEPOR protomers. The planes were defined by the centres of mass of the D1 and D2 domains and the four-amino acid-hinge region (amino acids Val118-Leu121) at each time point. The distance d was calculated as the distance between the centres of mass of the D2 domains at each time point. Analysis was performed over the last 2.5 ns for each asymmetric unit.

4.3 Results

4.3.1 BPTI crystal

Equilibration of the BPTI crystal was monitored via the pressure (Figure 4.5A). Between 0-5 ns the volume of the system was kept constant and water molecules removed until the average pressure in each direction was fluctuating between 0-100 bar. The system was allowed to further relax under isotropic pressure coupling with position restraints before going to fully anisotropic pressure coupling with no restraints. The dimensions of the unit cell remained constant under anisotropic conditions (Figure 4.5B), deviating less than ±4% of the box dimensions (Figure 4.5C). The average RMSD for the entire super-unit cell over the backbone atoms for the last 2.5 ns with respect to the initial X-ray crystal structure was 0.13 nm (Figure 4.5D). This low backbone RMSD is reflected in the RMSD per residue where only 7 of the 53 residues in BPTI had an average RMSD greater than 0.2 nm from the initial structure (Figure 4.5E). This can also be seen in average asymmetric unit structure coloured according to the B-factors, which shows the largest deviations occurred in the side chains of a few surface-exposed residues and at the termini (Figure 4.5F).

4.3.2 Crystals containing the sEPOR

Similarly to the BPTI crystal simulation, equilibration of the sEPOR containing crystal was monitored via the pressure (Figures 4.6A, 4.7A and 4.8A). Once the systems had reached between 1-100 bar and switched from constant volume to constant pressure the box dimensions deviated less than ±2% of the unit cell dimensions (panels B and C in Figures 4.6, 4.7 and 4.8). The average RMSD for the entire super-unit cell over the (non-modelled) backbone atoms for the last 2.5 ns for these crystals ranged between 0.24-0.32 nm (panels D in Figures 4.6, 4.7 and 4.8). The RMSD averaged for each residue for both sEPOR chains is shown in Panel F and G respectively in Figures 4.6, 4.7 and 4.8). Overall these sEPOR chains have relatively large RMSD values in particular at the termini and in the surface exposed loops. These regions of high deviation are reflected in the average asymmetric unit coloured according to the B-factors (panel H in Figures 4.6, 4.7 and 4.8).

The conformation of the sEPOR chains within each dimer was also monitored by the distance between the D2 domains (d) and angle between the chains (θ), as was done previously in Figure 3.3.

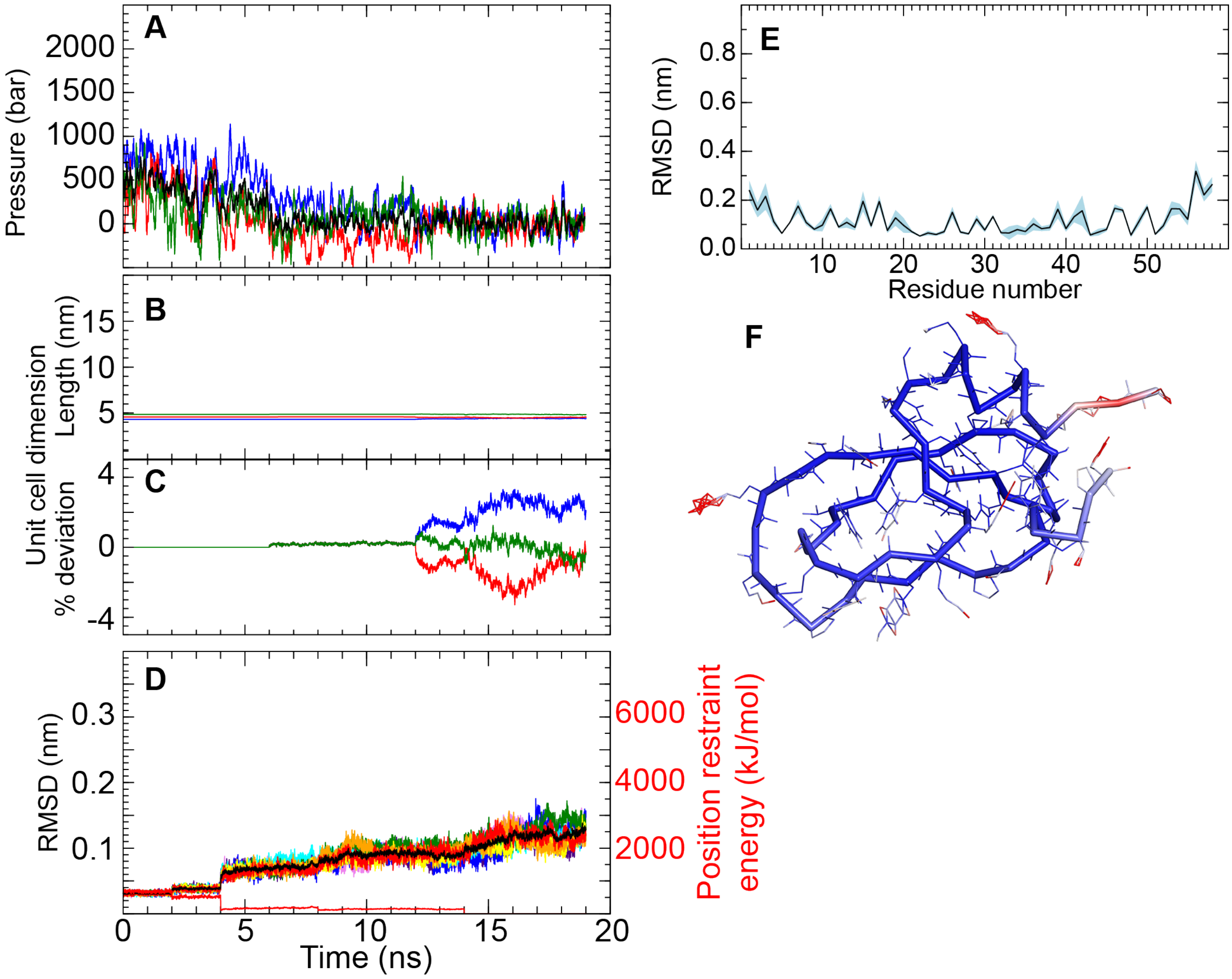


Figure 4.5: Results for the BPTI crystal lattice. (A) the total pressure of the unit cell (black) and along the X (blue), Y(red) and Z(green) axes. (B) fluctuations in the dimensions of the unit cell and deviation from the crystal values and (C) as a percentage along the X (blue), Y (red) and Z (green) axis. (D) The backbone RMSD (black) over the entire unit cell and the corresponding position restraint energy (red). (E) The average RMSD per residue from the initial crystal structure averaged over all asymmetric units within the crystal lattice. (F) The average asymmetric unit structure coloured according to the B-factor calculated from the simulation; blue, 0 Å^2; red, 100 Å^2.

In all cases d varied only over a small range of up to 0.6 nm across all dimers. The angle θ had a larger range of values between 80-150° in the EPO-bound dimers than the EMP1-bound dimers, which only ranged between 130-160°. Closer inspection of the EPO-bound dimer complexes revealed that in half of the dimer complexes θ increased while it decreased in the remaining half. This lead to average values for d and θ in the simulation of the EPO-bound dimer crystal (3.1 nm and 103.7°) that were near identical to the initial values (3.0 nm and 103.4°). In EMP1-bound dimer complexes the average for θ increased from the initial value of 130.1° to 139.6° and 141.2° in the originally submitted crystal structure and the structure with the swapped N-terminal respectively.

4.4 Discussion

X-ray crystallography not only provides beautiful images of proteins but also important insights into protein structure. Knowledge of the 3-dimensional structure for a protein is essential to understand how a protein functions and to explain how it can interact with other proteins or co-factors. Owing to the very nature of X-ray crystallography in which a crystal is formed from near-identical repeating units, the conformational space available to the protein is restricted and therefore limits information on the dynamical behaviour of the protein. Despite this limitation, detailed mechanisms of how proteins work have often been proposed based on a limited set of X-ray crystal structures. For example this has been done in the case of the erythropoietin receptor (EPOR) [8, 9]. In light of the work presented in Chapters 2 and 3, this approach to interpreting protein dynamics is questionable. To overcome the limits of protein X-ray crystallography MD simulations can be used to provide dynamical information about the protein. However, the accuracy and relevance of the models obtained in simulation is intrinsically limited by the quality of the initial structure employed and the reliability of the molecular force field. This study aimed to investigate and determine the quality of the protein structures containing the sEPOR. As a comparison we also examined the ability of the GROMOS54A7 force field to reproduce a well defined crystal lattice.

The crystal structure of BPTI (PDB entry 4PTI [76]) used in this study had the highest resolution at 1.5 Å of all the structures considered. It has a relatively small size (53 amino acids, compared to 225 for the sEPOR), and can be considered a high quality X-ray crystal structure given the low clash score and few Ramachandran and side chain outliers in the PDB validation report (Figure 4.1E). The quality of this structure is reflected in the ease of simulating the protein in a recreation of the crystal lattice as demonstrated by the relatively low position-restraint energies required to maintain the crystal structure (4.5D). The RMSD of 0.13 nm for the entire super-unit cell for BPTI was the lowest out of all the simulations. This RMSD value is comparable to, if not slightly lower than, the RMSD values produced from simulation of the protein ubiquitin in a crystal lattice, reported previously using an earlier version of the GROMOS force field [79, 80]. To give context to this value the distance a carbon-carbon bond is 0.154 nm. The RMSD observed is would almost entirely be due to thermal motion given that the simulations were performed at 293 K while the crystal structure was solved at 100 K. As evident from the B-factors the majority of the motion occurred in the exposed side chains and terminal regions (Figure 4.5F). Together these results reiterate the PDB validation report that

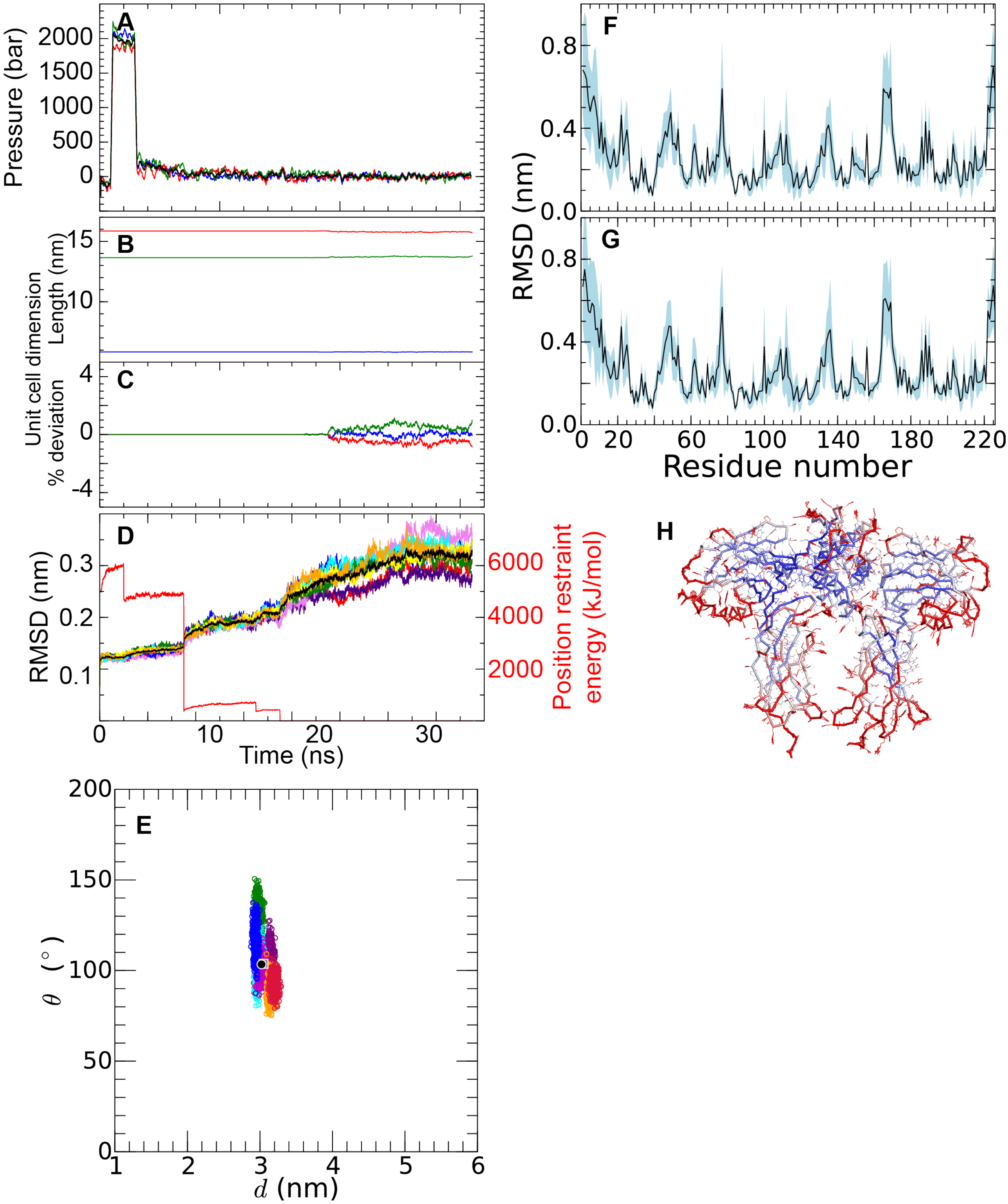


FIGURE 4.6: Results for the sEPOR$_2$-EPO crystal lattice. (A) the total pressure of the unit cell (black) and along the X (blue), Y(red) and Z(green) axes. (B) fluctuations in the dimensions of the unit cell and deviation from the crystal values and (C) as a percentage along the X (blue), Y (red) and Z (green) axis. (D) the backbone RMSD (black) over the entire unit cell and the corresponding position restraint energy (red). (E) Scatter plot of θ against d for the individual asymmetric units within the super-unit cell. Asymmetric units 1-8 coloured crimson, blue, orange, magenta, aqua, red, green and purple respectively. The initial crystal structure and average asymmetric unit values are shown as a black circle and grey square respectively. (F) and (G) the average RMSD per residue from the initial crystal structure averaged over all asymmetric units within the crystal lattice for the sEPOR chains A and B respectively. (H) the average asymmetric unit structure coloured according to the B-factor calculated from the simulation; blue, 0 Å^2; red, 100 Å^2.

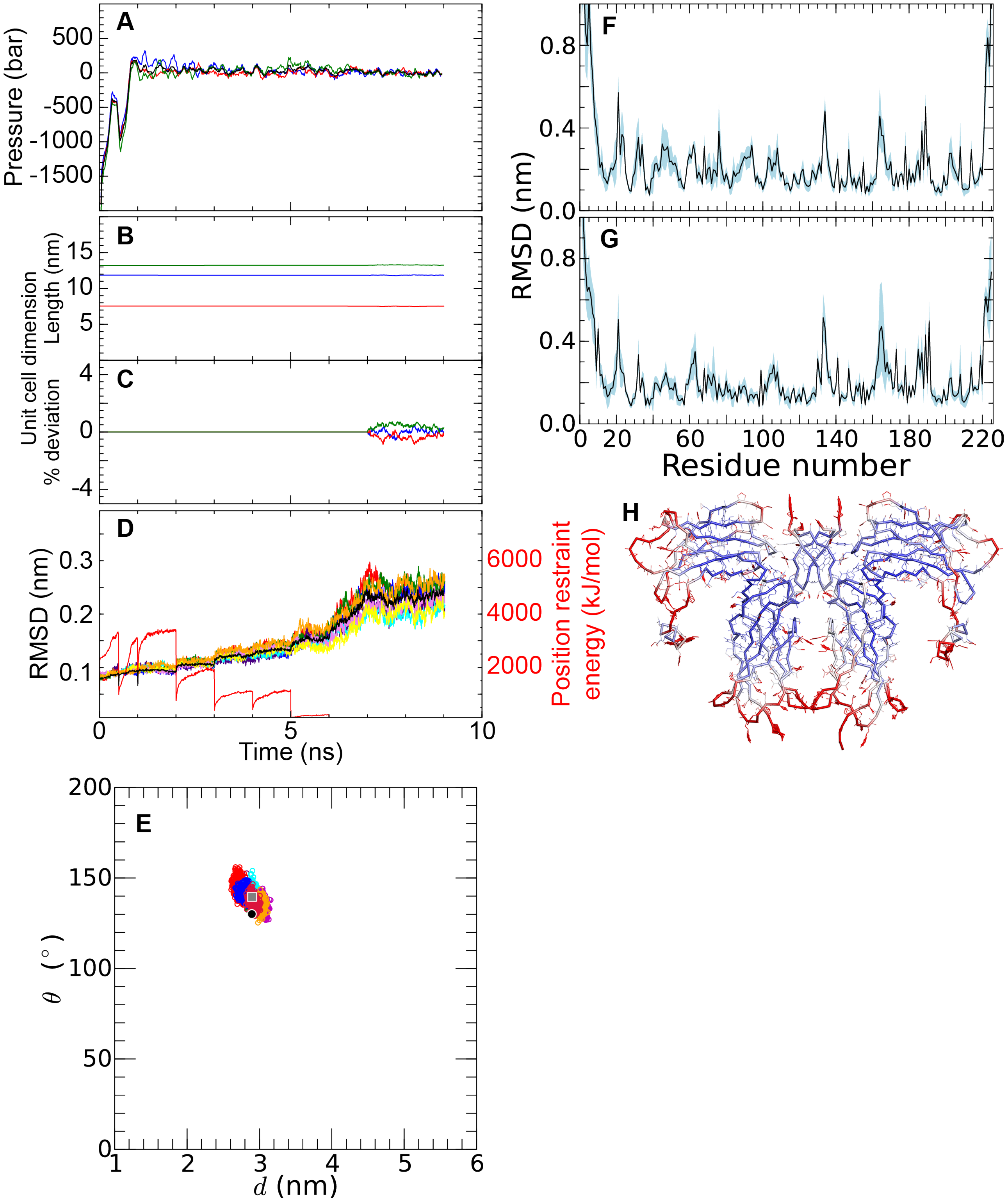


Figure 4.7: Results for the sEPOR$_2$-EMP1 crystal lattice. (A) the total pressure of the unit cell (black) and along the X (blue), Y(red) and Z(green) axes. (B) fluctuations in the dimensions of the unit cell and deviation from the crystal values and (C) as a percentage along the X (blue), Y (red) and Z (green) axis. (D) the backbone RMSD (black) over the entire unit cell and the corresponding position restraint energy (red). (E) Scatter plot of θ against d for the individual asymmetric units within the super-unit cell. Asymmetric units 1-8 coloured crimson, blue, orange, magenta, aqua, red, green and purple respectively. The initial crystal structure and average asymmetric unit values are shown as a black circle and grey square respectively. (F) and (G) the average RMSD per residue from the initial crystal structure averaged over all asymmetric units within the crystal lattice for the sEPOR chains A and B respectively. (H) the average asymmetric unit structure coloured according to the B-factor calculated from the simulation; blue, 0 Å^2; red, 100 Å^2.

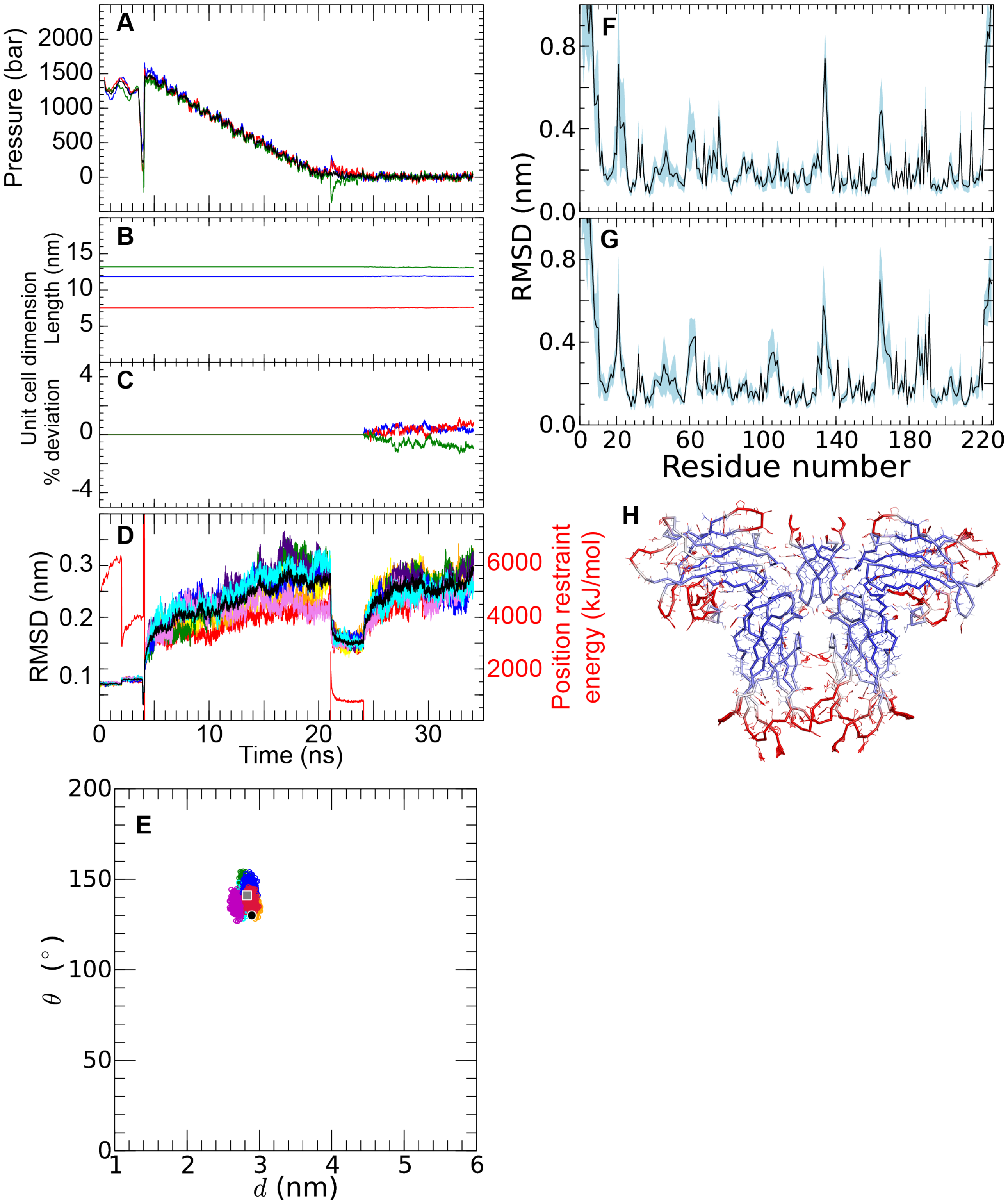


Figure 4.8: Results for the sEPOR$_2$-EMP1 crystal lattice with the swapped N-terminal region in sEPOR. (A) the total pressure of the unit cell (black) and along the X (blue), Y (red) and Z (green) axes. (B) fluctuations in the dimensions of the unit cell and deviation from the crystal values and (C) as a percentage along the X (blue), Y (red) and Z (green) axis. (D) the backbone RMSD (black) over the entire unit cell and the corresponding position restraint energy (red). (E) Scatter plot of θ against d for the individual asymmetric units within the super-unit cell. Asymmetric units 1-8 coloured crimson, blue, orange, magenta, aqua, red, green and purple respectively. The initial crystal structure and average asymmetric unit values are shown as a black circle and grey square respectively. (F) and (G) the average RMSD per residue from the initial crystal structure averaged over all asymmetric units within the crystal lattice for the sEPOR chains A and B respectively. (H) the average asymmetric unit structure coloured according to the B-factor calculated from the simulation; blue, 0 Å^2; red, 100 Å^2.

this X-ray crystal structure is of a high quality. Furthermore, this simulation proves the ability of the GROMOS54A7 force field to reproduce a protein crystal lattice. While somewhat crude in the fact that some of the solutes present experimentally in the crystalline medium were omitted in the simulations, e. g. missing polyethylene glycol or substituting Ca^{2+} for Na^{+}, that fact that no detrimental fluctuations in any of the proteins occurred indicates the methodology is reliable, but opportunities for refinement may exist.

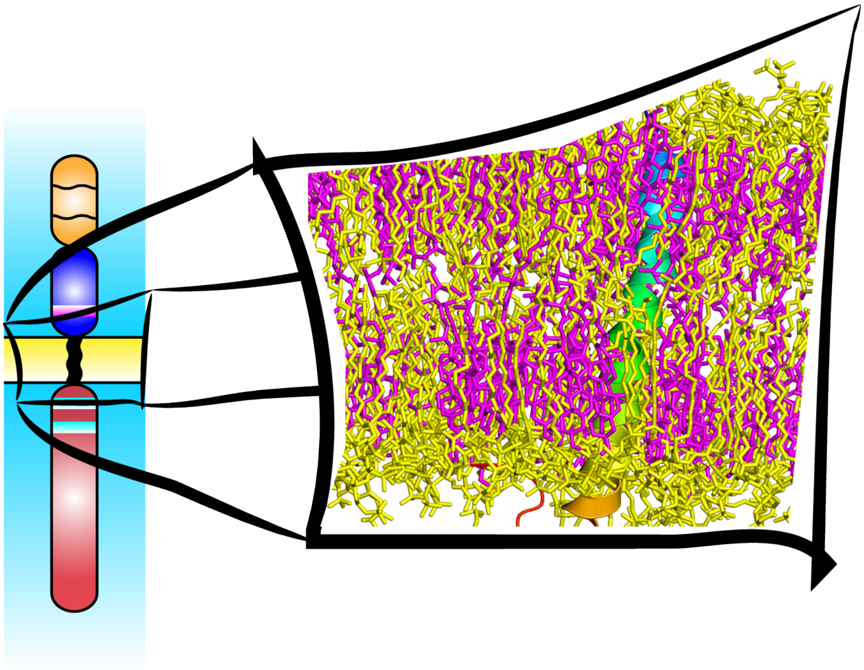
In the simulations with the sEPOR larger deviations in the RMSD values (0.24-0.32 nm) were observed over the entire super-unit cell than was seen with the BPTI crystal. Somewhat surprisingly the largest overall RMSD was calculated for the sEPOR_2-EPO crystal, which had the highest reported resolution of 1.9Å out of the three X-ray crystal structures containing the sEPOR. Intriguingly, this complex also deviated the greatest in regards to θ, however on average gave almost the exact value calculated for the initial structure. Furthermore, these changes in θ were not associated with significant deviations in the box dimensions (Figure 4.6C). A possible explanation for this result, is that the original authors enforced the symmetry of the average structure of a single sEPOR_2-EPO complex as the asymmetric unit, whereas the asymmetric unit may actually be composed of two complexes (or even more) with slightly different chain arrangements. Additionally, greater movement in this structure may be allowed due to a higher solvent content compated to the crystal containing the sEPOR-EMP1 dimers, approximately $1.2\times$ as many water molecules. In regard to the sEPOR-EMP1 crystal complexes, the results do not definitively prove the domain-swapped dimer of the N-terminal α-helix as originally published, or folded under the respective D1 domain, is a more stable form of the model than the other. The RMSDs from the original domain-swapped dimer gave the lower values, but the average results for θ and d where near identical in both systems, both deviate slightly from the initial crystal structure values. Only rerefinement of these models against the structure factors would be able to distinguish whether one model is a better fit than the other. Alternatively, given these structures were published over 20 years ago, resolving crystal structures of these complexes with modern techniques and equipment may generate models of higher quality without these structural uncertainties. One unexpected result that emerged from the comparison of the B-factors is the consistently high fluctuations observed in the lower half of the D2 domain. These loops and residues are proximal to the C-terminal that would be connected to the transmembrane domain inserted in a membrane. It is tempting to think that these residues are fluctuating because of missing interactions from either the transmembrane domain or lipid components of the membrane that would stabilise their positions.

4.5 Conclusion

This chapter aimed to validate a method used to reproduce protein crystals and thereby the use of the GROMOS54A7 force field in the simulation of systems containing the sEPOR. We believe the method was affirmed by the successful simulation of a stable BPTI crystal lattice. The same methodology was used to determine the quality of the X-ray crystal structures that contain the sEPOR molecule. While these simulations had fluctuations within the individual chains and overall arrangement of the complexes, on average they maintained the configuration of the asymmetric unit. These simulations support the PDB validation reports produced for these structures and that significant improvements could be made to the X-ray crystal structure. Unfortunately, further comparison of these crystal simulations to the experimental data would require the structure factor amplitudes, which are currently not publicly available. The authors of the X-ray crystal structures containing the sEPOR were contacted to request the structure factors, however this data was unable to be retrieved. This prevents the use of these simulations to further refine the models that could perhaps answer some the questions posed by these simulations such as, whether the sEPOR_2-EPO complex has a larger asymmetric unit, and whether the domain-swapped dimer occurs in all the EMP-bound sEPOR complexes.

Chapter 5

Does the Membrane Environment Affect the Structural Properties of the Type-I Cytokine Receptor Transmembrane Domains?



5.1 Introduction

In the previous Chapters 2-4 only the extracellular domain of the erythropoietin receptor (EPOR) has been considered in detail. However, as illustrated in Figure 1.1, the full-length EPOR, and all other type-I cytokine receptors, have the extracellular domain linked to an intracellular region via a single helical transmembrane domain that spans the plasma membrane. As suggested by the simulations presented in Chapter 3, to understand ligand-induced activation of these receptors, the restraints imposed by a transmembrane domain embedded in a membrane must also be considered. Further to this, it would also appear that the composition of the membrane, in particular the presence of cholesterol and membrane rafts, is critical to the activation of at least the growth hormone receptor (GHR), interleukin-7 receptor subunit α (IL-7Rα) and EPOR [81, 28, 38]. What remains to be determined is whether the receptors are merely co-localising with the cholesterol-rich rafts because that is where target kinases, for example Lyn kinase, are located [38], or whether the types of lipids within the rafts have a regulatory effect on the structure of the receptor transmembrane domains.

The transmembrane domain of the type-I cytokine receptors is predicted to be helical. This has been confirmed experimentally for the human and mouse EPOR [10, 11] and human prolactin receptor (PRLR) [12] in micelles by NMR spectroscopy. The helical nature of the transmembrane domain has been exploited in chimeric constructs in which the mouse erythropoietin receptor (mEpoR) and human GHR transmembrane domains are attached to leucine zippers from transcription factors [21, 17, 18]. These constructs suggested a critical role of the transmembrane domain in controlling the activation state of the receptor by orientating the intracellular domain. Despite knowing the secondary structure of the transmembrane domains, relatively little atomic resolution structural information is actually available on how these domains interact as a dimer, let alone with the lipids in the membrane. This has led some to employ computational techniques to find likely dimer conformations for the mEpoR and GHR [21, 82, 18]. Active and inactive conformations of the hGHR transmembrane dimer were proposed using a coarse-grained model of the transmembrane domain in a pure 1,2-dipalmitoyl-*sn*-glycero-3-phosphocholine (DPPC). Both studies that employed computational methods to determine low-energy dimer conformations of the mEpoR transmembrane domain were performed *in vacuo* [21, 82], that is there was no explicit treatment of the lipid molecules and water. This raises the question of whether these structures are physiologically meaningful conformations as different conformations may be obtained with lipid molecules interacting with the protein side chains. Indeed, it has been reported that depletion of cholesterol from a membrane can inhibit activation of the hGHR [81]. Whether this corresponds with a change in structure of the transmembrane domain in the different membrane environments has not been determined.

In order to better understand the behaviour and structure of the type-I cytokine receptor transmembrane domains we have investigated the transmembrane domains from several receptors as a monomer in a pure 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphocholine (POPC) bilayer and a mixed bilayer of POPC and cholesterol. We explored several structural features of the transmembrane domains and asked whether it is appropriate to ignore the effects of the membrane. Furthermore, we examined a range of receptors in order to verify whether the type-I cytokine receptors behave consistently across the family so that a single receptor could be used as a representative model for the whole family.

5.2 Methodology

5.2.1 Systems simulated

The transmembrane domains from the type-I cytokine receptors were investigated embedded in two types of model membrane: (i) a pure POPC bilayer and (ii) a POPC:Cholesterol bilayer. The type-I cytokine receptor sequences chosen for study were based on the availability of a corresponding X-ray crystal structure for the extracellular domain. The receptors examined and the corresponding UniProt sequence used were the human erythropoietin receptor (EPOR) P19235; the human growth hormone receptor (GHR) P10912; the rat prolactin receptor (rPRLR) P05710; the human granulocyte-macrophage colony-stimulating factor receptor subunit α (CSF2Rα) P15509; the human thrombopoietin receptor (TPOR) P40238; the human interleukin-4 receptor subunit α (IL-4Rα) P24394; the human interleukin-5 receptor subunit α (IL-5Rα) Q01344 and the human IL-7Rα P16871. The sequences were taken from the terminal residue observed in the corresponding X-ray crystal structure and included the Box-1 motif on the intracellular side of the transmembrane domain. Specifically the sequences and PDB entries were; GHR: P234-K287 (3HHR); IL-4Rα: H196-A244 (5E4E); IL-7Rα: N212-K260 (3UP1); rPRLR: N204-K251 (3NPZ); CSF2Rα: F296-K344(4RS1); IL-5Rα: H316-K359 (3VA2); EPOR: T220-E266 (1EER); and TPOR: T460-H511. An alignment of the sequences used that also includes the WSXWS motif is shown in Figure 5.1. The transmembrane domains were built with the N-terminal sequence up until the beginning of the transmembrane domain (as defined in the corresponding UniProt entry) as an unstructured loop, with the transmembrane domain and the remaining C-terminal sequence modelled as an ideal α-helix using PyMOL [64]. In one set of simulations of the EPOR the available nuclear magnetic resonance (NMR) spectroscopy structure was used for the initial structure [10]. The peptides were embedded into a membrane using the GROMACS program g_embed [83]. The peptides were first aligned with the α-helix of the transmembrane domain perpendicular to the plane of the membrane with the first residue after the end of the predicted transmembrane domain aligned to the average Z-coordinate of the phosphorus atom from POPC in the lower leaflet. The proteins were then resized in the XY directions (corresponding to the plane of the membrane) to 10% of their original size and to insert the α-helix any overlapping lipid and water molecules were deleted. The peptides were then grown stepwise to the original size [83]. An overview of the systems simulated is presented in Table 5.1

```
GHR      YGEFSEVLYVTLPQMSQFTCEEDFYFPWLLIIIFGIFGLTVM---LFVFLFSKQQRIKMLILPPVPVPK
IL-4RA   WSEWSPSTKWHNSYREPFE---QHLLLGVSVS--CIVILAVC---L--LCYVSITKIKKEWWDQIPNPA
IL-7RA   WSEWSPSYYFRTPEINNSSGEMDPILLTISILSFFSVALLVI---L--ACVLWKKRIKPIVWPSLPDHK
rPRLR    WSRWSQESSVEMPNDFTLKDTTVWIIV-------AILSAVICL-IMVWAVALKGYSMMTCIFPPVPGPK
PRLR     WSAWSPATFIQIPSDFTMNDTTVWISV-------AVLSAVICL-IIVWAVALKGYSMVTCIFPPVPGPK
CSF2RA   WSSWSEAIEFGSDDGNLGS---VYIYV----LLIVGT--LVCG--IVLGFLFKRFLRIQRLFPPVPQIK
IL-5RA   WSEWSQPIYVGNDEHKPLR---EWFVIVI--------MATICFILLILSLICKICHLWIKLFPPIPAPK
EPOR     WSAWSEPVSLLTPSDL------DPLILTLSLILVV-----ILVLLTVLALLSHRRALKQKIWPGIPSPE
TPOR     WSSWSDPTRVETATET------AWISLVTALHLVLGLSAVLGLLLLRWQFPAHYRRLRHALWPSLPDLH
         :. :*                                   :                       :*
```

FIGURE 5.1: Protein sequence alignment of the type-I cytokine receptors from the WSXWS motif (blue) to the Box-1 motif (red) that includes the predicted transmembrane domain sequence (bold). The underlined residues indicates the beginning of the sequence used in the simulations. Alignment was performed using the Multiple Sequence Alignment Tool available at www.ebi.ac.uk/Tools/msa/.

TABLE 5.1: Overview of the type-I cytokine receptor transmembrane domains systems simulated.

System	Membrane Composition (ratio)	Copies of peptide	Simulation time (ns)
mEpoR_1	POPC	1	100
mEpoR_{1C}	POPC:Cholesterol (1:1)	1	100
GHR_1	POPC	1	100
GHR_{1C}	POPC:Cholesterol (1:1)	1	100
rPRLR_1	POPC	1	100
rPRLR_{1C}	POPC:Cholesterol (1:1)	1	100
EPOR_4	POPC	4	180
EPOR_{4C}	POPC:Cholesterol (1:1)	4	180
GHR_4	POPC	4	180
GHR_{4C}	POPC:Cholesterol (1:1)	4	180
TPOR_4	POPC	4	180
TPOR_{4C}	POPC:Cholesterol (1:1)	4	180
rPRLR_4	POPC	4	180
rPRLR_{4C}	POPC:Cholesterol (1:1)	4	180
$\mathrm{CSF2R}\alpha_4$	POPC	4	180
$\mathrm{CSF2R}\alpha_{4C}$	POPC:Cholesterol (1:1)	4	180
$\mathrm{IL\text{-}4R}\alpha_4$	POPC	4	180
$\mathrm{IL\text{-}4R}\alpha_{4C}$	POPC:Cholesterol (1:1)	4	180
$\mathrm{IL\text{-}5R}\alpha_4$	POPC	4	180
$\mathrm{IL\text{-}5R}\alpha_{4C}$	POPC:Cholesterol (1:1)	4	180
$\mathrm{IL\text{-}7R}\alpha_4$	POPC	4	180
$\mathrm{IL\text{-}7R}\alpha_{4C}$	POPC:Cholesterol (1:1)	4	180

[a] POPC, 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphocholine

5.2.2 Simulation parameters

All simulations were performed using the GROMACS simulation package 4.6.5 [50] in conjunction with the GROMOS 54A7 united-atom force field [48]. Each system was simulated under periodic boundary conditions in a rectangular box. The pressure was maintained at 1 bar by weakly coupling the system to a semi-isotropic pressure bath [53] using an isothermal compressibility of 4.6×10^{-5} bar^{-1} and a coupling constant $\tau_P = 1$ ps in the XY and Z directions, corresponding to the plane and normal directions of the bilayer respectively. The temperature of the system was maintained at 298 K by independently coupling the protein, lipids and water to an external temperature bath with a coupling constant $\tau_T = 0.1$ ps using a Berendsen thermostat [53]. All bond lengths within the protein and lipids were constrained using the LINCS algorithm [66]. Water was included explicitly in the simulations using the Simple-Point Charge (SPC) model [67] and constrained using the SETTLE algorithm [68]. Explicit polar hydrogen atoms in the protein and lipids were replaced by virtual interaction sites, the positions of which were constructed at each step from the coordinates of the heavy atoms to which they are attached [71]. This allowed a 4-fs time step to be used without affecting the thermodynamic properties of the system significantly. Nonbonded interactions within the cutoff of 1.4 nm were calculated every step and the pair list was updated every three steps. A reaction-field correction was applied to the electrostatic interactions beyond the cutoff of 1.4 nm [69], using a relative dielectric permittivity constant of $\epsilon_{\mathrm{RF}} = 62$ as appropriate for SPC water [70]. All systems were energy-minimised after embedding the peptides into the membrane. The temperature of each system was then gradually increased from 50 K to 298 K in 50-K steps over 120 ps to further relax the system and obtain the starting configurations used in the simulations.

5.2.3 Analysis

5.2.3.1 Residue helicity

The GROMACS program g_helix [50] was used to examine the percentage of time each residue was in a helix over the final 50 ns of simulation using the criteria of Hirst and Brooks [84]. A residue was deemed to be in a helix if:

$$\sqrt{(\phi - \phi_c)^2 + (\psi - \psi_c)^2} < 8 \tag{5.1}$$

where ϕ_c and ψ_c are any of the 12 pairs of (ϕ, ψ) helix angles examined by Manning and Woody [85].

5.2.3.2 Helix length

Again, the GROMACS program g_helix [50] was used to obtain the total length of the helix by fitting the longest helical part to an ideal α-helix around the Z-axis. The total length was then calculated over the final 50 ns of simulation from the average distance between the Cα atoms times the total number of residues in the helix.

5.2.3.3 Helix vector angle

The overall tilt of the helix was calculated over the final 50 ns of simulation between the normal of the bilayer (Z-axis) and the vector formed from the centroids of the Cα atoms in the first and last turns in the transmembrane region of the helix. A schematic is provided in Figure 5.2A.

5.2.3.4 Local helix bending

The GROMACS program g_helixorient [50] was used to examine the local helical bending for each residue in the helix over the final 50 ns of simulation. The local residue helix bending for the *i*th residues is defined as the angle calculated between local axis for the *i*-2 and *i*-1 at each time point.

The local residue helix axis is obtained from the axis of the helix (X-axis), the Cα vector (Y-axis) and the XY cross-product (Z-axis). A schematic is provided in Figure 5.2B.

5.2.3.5 Local helix tilt

The GROMACS program g_helixorient [50] was used to examine the local residue helix tilt over the final 50 ns of simulation. Again, the local residue helix axis is obtained from the axis of the helix (X-axis), the Cα vector (Y-axis) and the XY cross-product (Z-axis). The local residue helix tilt is the angle required to rotate the XY-axes to the initial reference frame. A schematic is provided in Figure 5.2C.

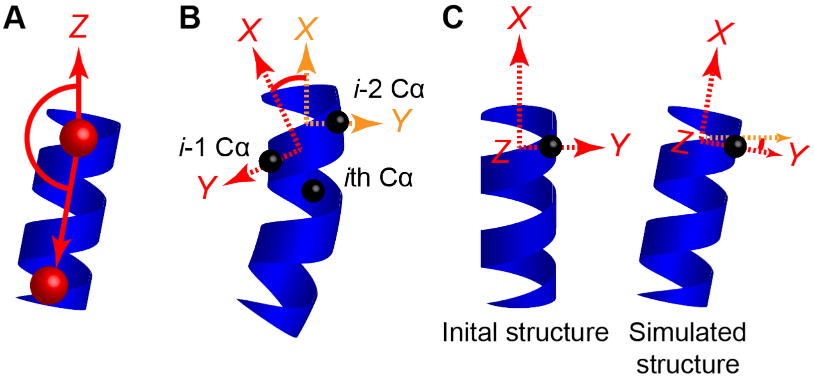


FIGURE 5.2: Schematics of the analysis performed on the transmembrane domains helices. (A) The overall helix tilt was calculated between the centroids of the Cα atoms in the first and last helix turn of the transmembrane domain. (B) The local residue bending was the angle calculated between the local axis for the preceding two residues. (C) The local residue tilt was calculated between the axis defined for each helix at each time point in the simulation with the corresponding helix in the initial structure.

5.3 Results

5.3.1 Overall behaviour of the transmembrane domains in different lipid bilayers

Initially, a single transmembrane domain from the mEpoR, rPRLR and GHR were examined in both a pure POPC or a mixed POPC and cholesterol bilayer to determine if there were obvious differences in the behaviour of the transmembrane domains in different bilayers that could be investigated further. For each system, the membrane used, as well as the starting (0 ns) and final (100 ns) configurations, is shown in Figure 5.3. For the mEpoR and rPRLR the transmembrane domains had an overall tilt in the pure POPC bilayer substantially larger than in the respective mixed POPC:cholesterol bilayer simulations. For GHR the difference in tilt between the two lipid compositions was less severe, but after visual inspection it appears some bending was present in the mixed POPC and cholesterol bilayer, whereas in the pure POPC bilayer the transmembrane domain remained virtually straight throughout the whole simulation. These simulations prompted a systematic investigation into the structural and behavioural differences of the transmembrane domains in the type-I cytokine receptor family between different membrane compositions.

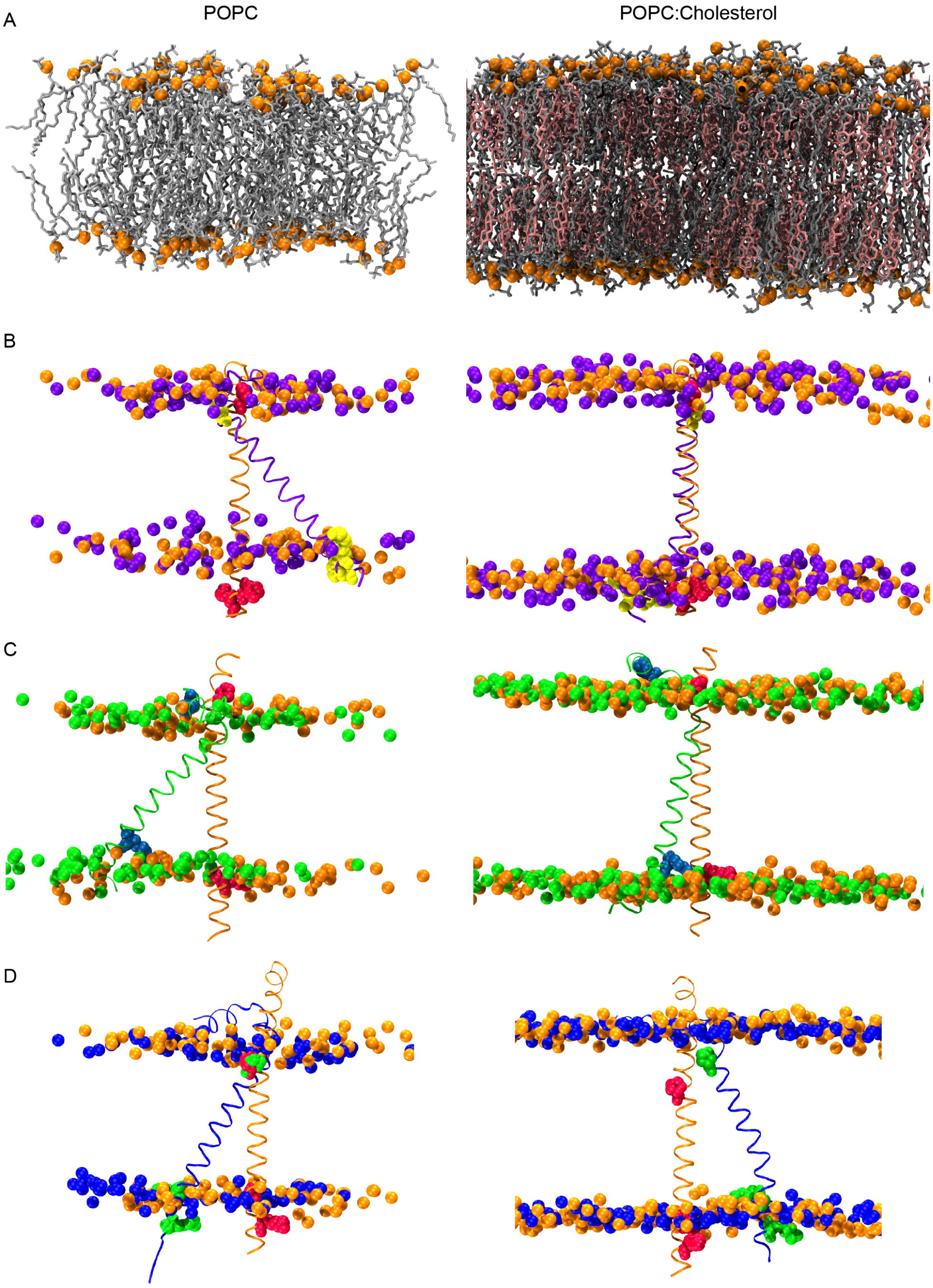


FIGURE 5.3: (A) The receptor transmembrane domains were simulated in either a pure POPC or POPC:cholesterol mixture bilayer. The coordinates of the peptide sequence were obtained either from NMR data in the case of the mEpoR (B) or built using PyMOL for the rPRLR and GHR (C-D). The orientation of the transmembrane domain from mEpoR (B) , rPRLR (C) and GHR (D) in the initial (orange helices) and final (purple, green and blue helices respectively) frames of the simulations are shown as a trace of the backbone atoms. The coordinates of the POPC phosphate atoms are represented as spheres in the initial (orange) and final (purple, green or blue) frames. The residue predicted to be the start of the transmembrane domain as well as the charge residues immediately after the C-terminal end of the transmembrane domain are represented as spheres. The head group and acyl chains of POPC (grey) and cholesterol molecules (pink) have been omitted from (B-D) for clarity.

5.3.2 Transmembrane domain helicity, helix length and overall helix tilt

A range of type-I cytokine receptor transmembrane domains were then examined by embedding four copies of the peptide into a single bilayer. Note, the simulation of four copies of the transmembrane domain embedded in a single lipid bilayer allowed us to improve the sampling statistics. The relative position of the transmembrane domains was such that there was no interaction between different copies of the transmembrane domains. The transmembrane domains were analysed for percentage of time each residue in the sequence was considered in a helical fold (Figure 5.4). Overall the residues predicted to be the transmembrane domain from these receptors corresponded to the longest continuous sequence of residues that remained predominantly helical in both bilayers. For most of the receptors the helix extended 3-5 residues (up to two turns) into the intracellular juxtamembrane domain sequence. In the case of the TPOR the helix finished at the end of the transmembrane domain sequence but another helix formed starting around 7 residues downstream (residue 496) in both bilayers (Figure 5.4E). In regard to the helicity of the N-terminal (extracellular-domain-proximal) region of the transmembrane domain, in the majority of cases, the membrane composition did not appear to affect the helicity of the residues. In the case of the TPOR the addition of cholesterol decreased the helicity of the first seven residues of the transmembrane domain (residues 468-474) (Figure 5.4E).

The length of the longest continual helix was calculated for each of the receptors in the two membranes (Figure 5.5). The average length of the helix of the EPOR, TPOR, rPRLR, CSF2Rα, IL-5Rα and IL-7Rα all decreased between 5-13% in the bilayer containing cholesterol compared with the pure POPC bilayer (Table 5.2). The TPOR had the greatest percent change (13%) decreasing from 3.9±0.4 nm to 3.4±0.2 nm, while the rPRLR had the greatest value change of 5.1±0.2 nm to 4.5±0.1 nm. The IL-4Rα and GHR were the only sequences to have an average increase in helix length in the cholesterol containing bilayer, increasing by 2% and 10% respectively.

The helix vector angle, which describes the overall tilt of the transmembrane domain, was calculated for each of the receptors from the centroids of the Cα atoms in the first and last turns in the transmembrane region (Figure 5.6). The vectors used to calculate the angle were defined so that 180° is perpendicular to the plane of the membrane. Overall the receptor transmembrane domains could be grouped into two distinct populations (Table 5.2). In the first group that comprised the EPOR, GHR, CSF2Rα, IL-4Rα, IL-5Rα and IL-7Rα the transmembrane domains were above an average angle of 160° in the pure POPC and had very similar angles in the bilayer with cholesterol, within 2-6°. The second group consists of the rPRLR and TPOR (Figure 5.6C and E respectively), for which the average helix vector angle increased by over 18° when cholesterol is present in the bilayer. The TPOR had the lowest average angles for both bilayers of 131.3°±2.0° 151.8°±1.1° in the pure POPC and mixed POPC and cholesterol bilayers.

5.3.3 Transmembrane domain local bending and tilt

The TPOR, rPRLR and GHR transmembrane domains were examined in further detail for the local bending and tilt of the helices (Figure 5.7). The TPOR displayed the largest average local bending and greatest variation in both the POPC and mixed POPC and cholesterol bilayer of 20°±17° and 13°±10

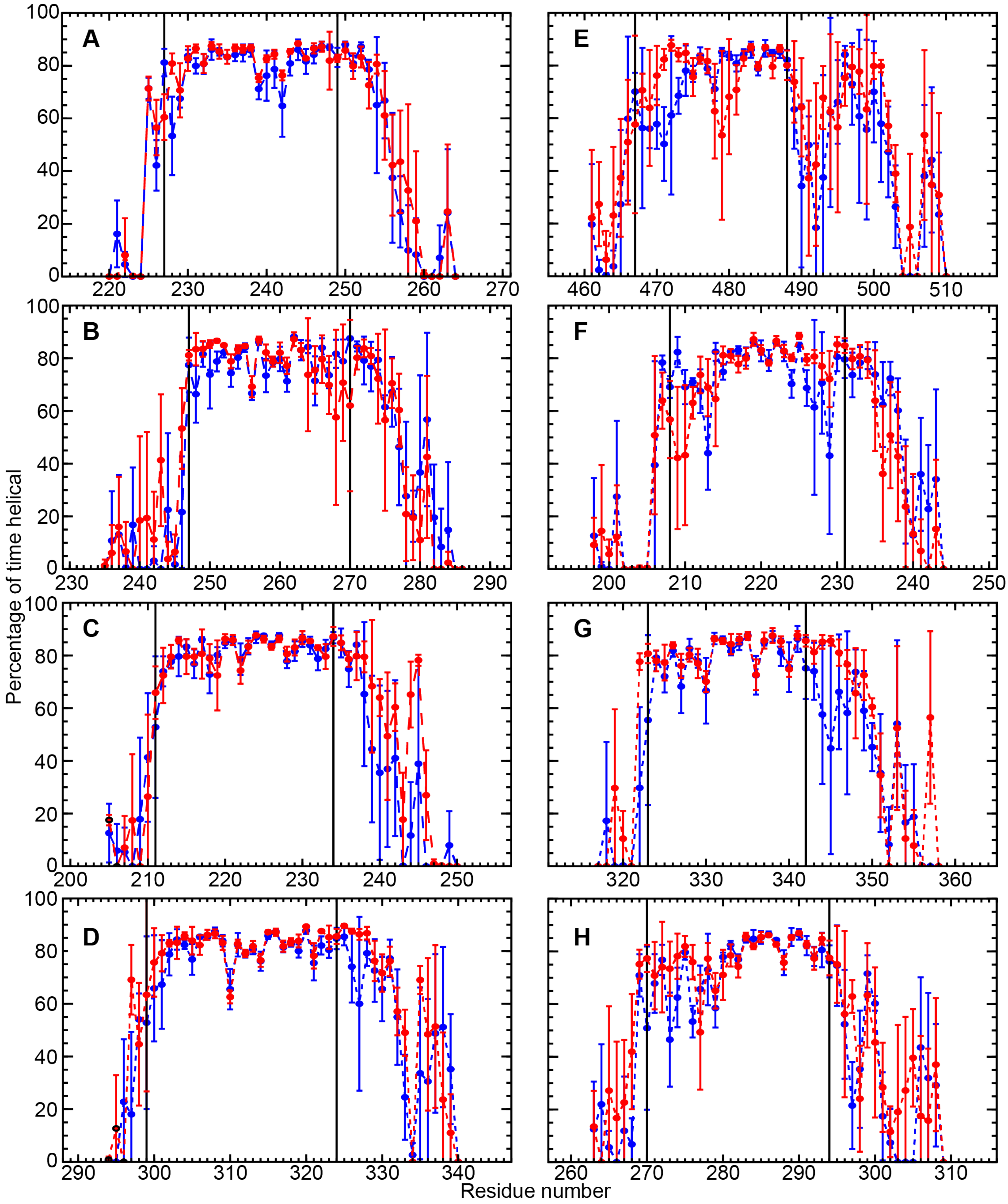


Figure 5.4: Average residue helicity in a POPC (red) or POPC:Cholesterol (blue) bilayer for the EPOR (A), GHR (B), rPRLR (C), CSF2Rα (D), TPOR (E), IL-4Rα (F), IL-5Rα (G) and IL-7Rα (H) transmembrane domains. Values are the average over all four peptide copies ± the standard deviation.

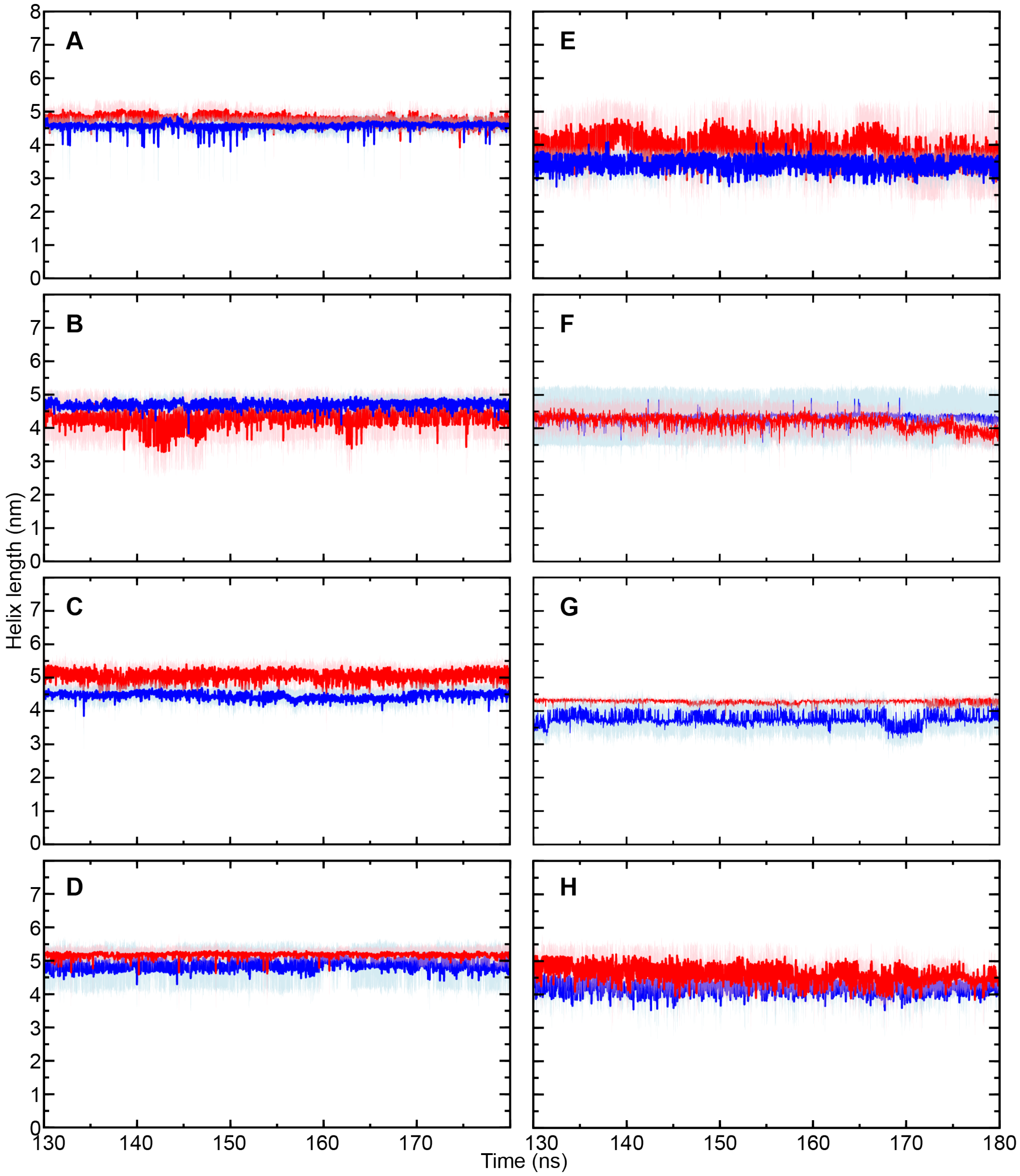


Figure 5.5: Average length of the transmembrane domain helix in a POPC (red) or POPC:Cholesterol (blue) bilayer for the EPOR (A), GHR (B), rPRLR (C), CSF2Rα (D), TPOR (E), IL-4Rα (F), IL-5Rα (G) and IL-7Rα (H). Values are the average over all four peptide copies. The pink and light blue shading is ± the standard deviation.

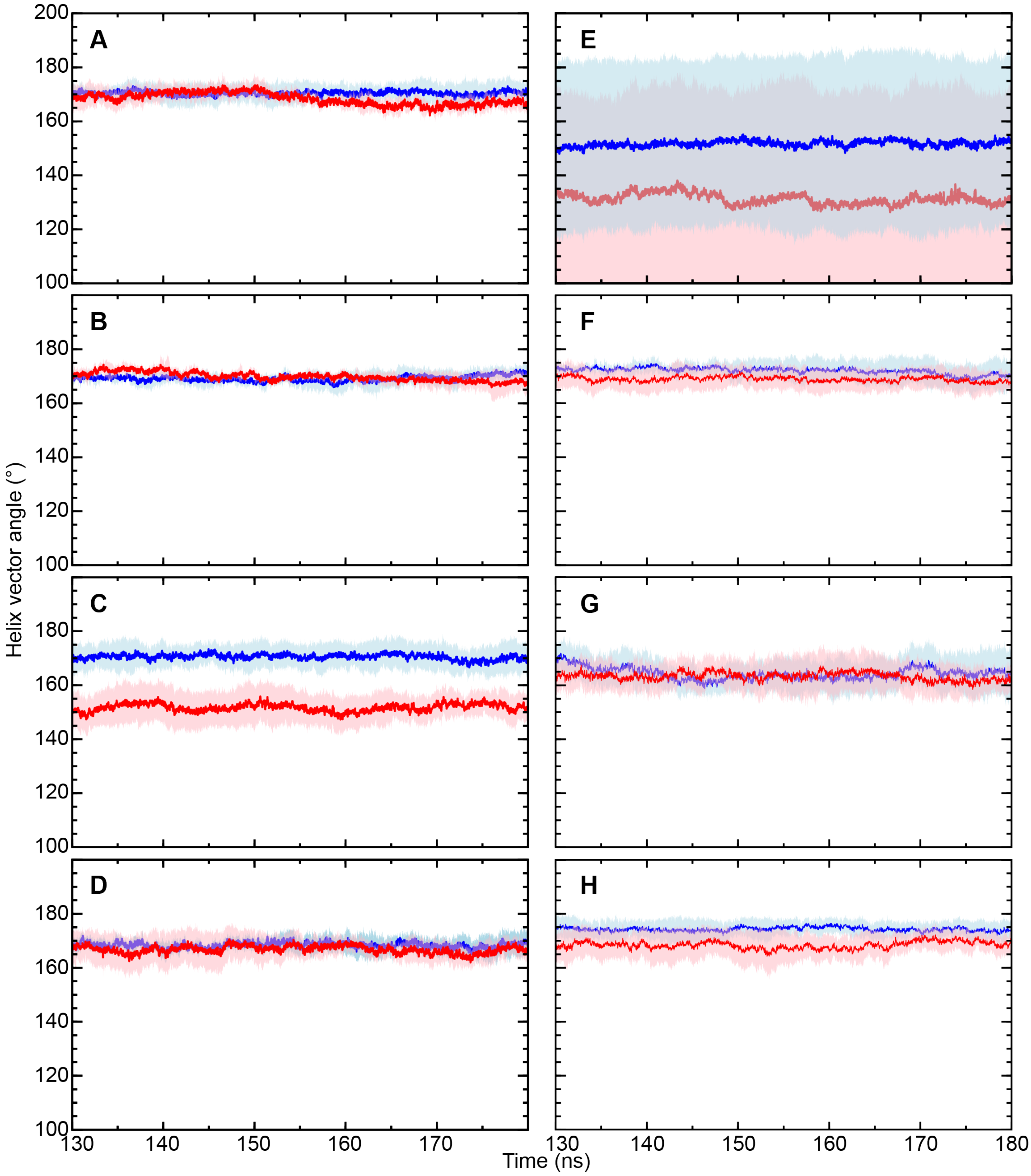


FIGURE 5.6: Average overall tilt angle in a POPC (red) or POPC:Cholesterol (blue) bilayer for the EPOR (A), GHR (B), rPRLR (C), CSF2Rα (D), TPOR (E), IL-4Rα (F), IL-5Rα (G) and IL-7Rα (H) transmembrane domains. Values are the average over all four peptide copies. The pink and light blue shading is ± the standard deviation.

TABLE 5.2: Total helix length and vector, and local helix bending and tilt. Values are mean and SD.

	Total helix		Local helix	
System	Length (nm)	Vector angle (°)	Bending (°)	Tilt (°)
EPOR$_4$	4.8±0.1	168.3±2.2	3.6±1.1	12.8±4.0
EPOR$_{4C}$	4.6±0.1	170.5±0.8	4.1±1.5	12.6±3.9
GHR$_4$	4.3±0.2	169.9±1.6	4.5±1.2	14.1±3.6
GHR$_{4C}$	4.7±0.1	169.2±1.0	4.5±1.3	16.0±3.1
TPOR$_4$	3.9±0.4	131.3±2.0	20±17	47±21
TPOR$_{4C}$	3.4±0.2	151.8±1.1	13±10	28±15
rPRLR$_4$	5.1±0.2	151.7±1.4	4.0±1.0	24.9±8.1
rPRLR$_{4C}$	4.5±0.1	170.6±1.0	3.8±1.1	11.6±2.4
CSF2Rα_4	5.2±0.1	166.8±1.4	4.0±1.1	14.8±5.2
CSF2Rα_{4C}	4.9±0.1	168.3±0.9	4.1±1.1	12.7±2.8
IL-4Rα_4	4.2±0.2	168.9±0.9	7.5±2.9	16.5±3.5
IL-4Rα_{4C}	4.3±0.1	172.1±1.2	5.1±1.6	12.0±2.7
IL-5Rα_4	4.3±0.1	163.3±1.5	3.6±1.2	17.1±5.8
IL-5Rα_{4C}	3.8±0.2	164.3±2.2	3.7±1.3	14.9±6.0
IL-7Rα_4	4.6±0.2	168.4±1.3	5.2±1.4	16.2±3.3
IL-7Rα_{4C}	4.2±0.2	174.3±0.8	5.4±1.6	17.8±5.0

(Figure 5.7A). This bending can be seen in the average structure of the TPOR to occur predominantly in the C-terminal residues of the helix as it exits the bilayers. The rPRLR and GHR had a negligible local average bending (approximately 4°) (Figures 5.7B and C). The rPRLR decreases the average local tilt from 25°±8° to 11.6°±2.4° in the presence of cholesterol, whereas the GHR remains relatively stable at approximately 15° (Table 5.2).

5.4 Discussion

Transmembrane domains provide a physical link through the plasma membrane between the interior of a cell and its environment. In some proteins, for example the G-protein coupled receptors, the transmembrane domain spans several helices that enter and exit the membrane multiple times. However, in the case of the type-I cytokine receptors each chain has a single helical transmembrane domain linking the extracellular and intracellular domains (Figure 1.1). This structural similarity between the type-I cytokine receptors allows for the extracellular domain from one receptor to be grafted onto that of another with the resulting complex capable of activating respective JAK/STAT signalling pathway. For example, a chimeric construct consisting of the EPOR extracellular domain and the gp130 transmembrane and cytosolic domains could activate STAT1 and STAT3 in the presence of EPO [86]. Given this and the fact that there are only four JAK and seven STAT proteins, it is reasonable to suggest that type-I cytokine receptors likely act via a shared mechanism. It would therefore be expected that the helix that links the extracellular and intracellular domain would have similar behaviour across the family. However, these simulations showed a range of behaviour across the family both between the family members and when comparing the same receptor in different lipid environments.

A comparison of the time each residue is considered to be in a helical fold (as a percentage of the total time) did not reveal any residues that had different structural behaviour in the receptor trans-

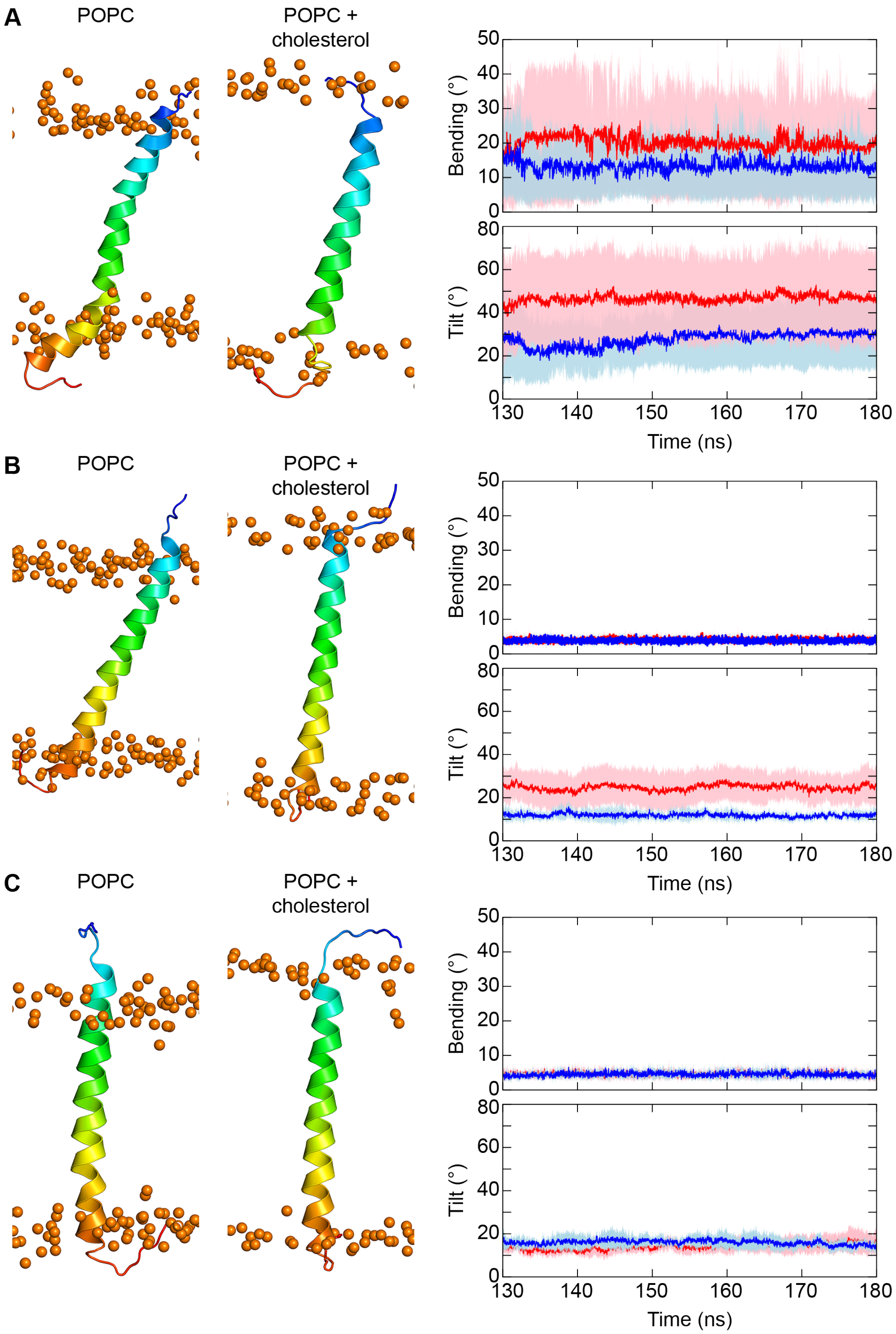


FIGURE 5.7: The local bending and tilt of the TPOR (A), rPRLR (B) and GHR (C) transmembrane domains. The average structure of the transmembrane domain from the final 50 ns of simulation are shown for each system from the POPC and POPC:cholesterol bilayers. The peptide sequence is coloured blue to red from the N- to C-termini. The phosphorus atoms of the POPC molecules are represented as orange spheres and are indicative only. The average bending is calculated over all residues in the helix for all four peptide copies for the POPC (red) and POPC:cholesterol (blue) bilayers. The average tilt is calculated over all local helices in the helix for all four peptide copies for the POPC (red) and POPC:cholesterol (blue) bilayers. The pink and light blue shading is ± the standard deviation.

membrane domain in the two lipid bilayers (Figure 5.4). Interestingly, the helix extended past the end of the proposed transmembrane domain into the intracellular juxtamembrane domain by around four residues, approximately another complete turn of the helix. This has been seen in the NMR structures of the EPOR [10, 11] and is important to note as alanine insertion mutants immediately after the final residue in the transmembrane domain sequence of the mEpoR could constitutively activate the receptor. From these data it was proposed that the orientation of the intracellular juxtamembrane domain was critical to the activation of the receptor [87]. Using the NMR structure of the mEpoR as the starting coordinates, the transmembrane domain tilted substantially more in the POPC bilayer compared to the mixed POPC and cholesterol bilayer (Figure 5.3B). This tilting was not observed in the modelled sequence of the human EPOR that was extended past the corresponding residues in the NMR sequence to include the Box 1 motif (Figure 5.6A). These extra intracellular juxtamembrane residues may have restricted the orientations that the transmembrane domain could adopt. Indeed, the tilt of the transmembrane domain in the case of the rPRLR appeared to influence the orientation of the intracellular juxtamembrane domain (Figure 5.7B). The rPRLR is intriguing at it was one of two receptors that had a large difference (18°) in the overall tilt between the two bilayers as well as the largest decrease in the helix length (Table 5.2). This tilt and change in length however did not correspond with any significant bending of the helix, or noticeable loss of residues involved in the helix. This is consistent with the NMR structure of the human PRLR transmembrane domain, which only bent as much as 6° [12]. The authors of this NMR structure also noted the PRLR transmembrane domain was embedded from residues 211 to 234 in 1,2-diheptanoyl-*sn*-glycero-3-phosphocholine (DHPC), a short chain lipid, which is in agreement with what was observed in these simulations. Both the human and rat PRLR therefore appear to have a similar structure across multiple types of bilayer, but shortened and more perpendicular to the membrane when cholesterol is present, this suggests the prolactin transmembrane domain responds to the hydrophobic mismatch by tilting rather than a structural change. This is important to know as the authors of the human PRLR transmembrane domain NMR structure have presented the model of the full-length PRLR with the transmembrane domain modelled orthogonal to the membrane [12]. In light of these simulations presented here, the full-length model of the PRLR would need to be investigated in membranes with different compositions in order to validate the model as presented or determine other potential arrangements of the transmembrane domain connected to the extracellular and intracellular domains.

The structures adopted by the TPOR transmembrane domain would be an interesting case to have validated experimentally, potentially by NMR as has been done for the EPOR and PRLR transmembrane domains [10, 11, 12]. This receptor transmembrane domain displayed large variations in the helix length, overall tilt, bending and local tilt when in the pure POPC bilayer that were somewhat dampened when in the presence of cholesterol (Table 5.2). The change in average secondary structure that occurred in the C-terminal region of this receptor is in contrast to all other receptors examined (Figure 5.7A). The GHR however was one case where there was an extension of the helix length in the presence of cholesterol corresponded to a residue at the C-terminal end of transmembrane domain becoming part of the helix (Figure 5.7C). Whether the type-I cytokine receptor transmembrane domains have cholesterol-binding motifs, such as those reported for the human nicotinic acetylcholine receptor that have been suggested to have important structural and/or functional roles [88], or a cholesterol-binding pocket formed upon dimerisation, has not been examined. These simulations may suggest, that the membrane could have a critical regulatory role in receptor activation, in particular controlling a conformational change within the transmembrane or intracellular juxtamembrane domain. This would be a novel aspect of the mechanism of activation for the type-I cytokine receptors that has not been investigated previously. Furthermore, if a conformational change was involved this would support the idea that the dimerisation of the receptor chains alone is insufficient for receptor activation. What would need to be tested is whether as a dimer these transmembrane domains had similar structural changes. This would require the use of highly-technical cryo-electron microscopy that could preserve the arrangements of the dimers in a membrane environment as has been done to study the structure of nuclear pore complex [89].

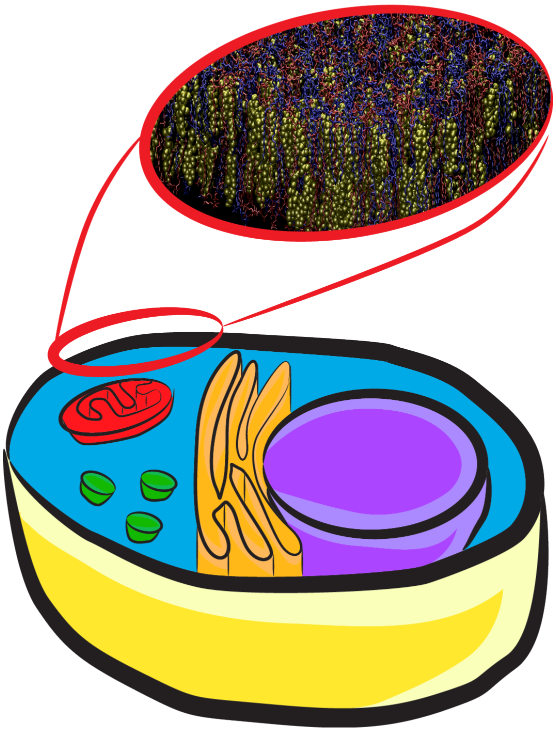
More generally these simulations demonstrate some of the difficulties in modelling proteins in a membrane environment. Large studies have been performed investigating dimer interfaces and orientation of helices based using only a pure lipid bilayer in both coarse-grained and full-atomistic molecular dynamics (MD) simulations in order to draw mechanistic conclusions [90, 91]. Indeed, particular crossed and parallel orientations of the GHR transmembrane dimer generated from simulation in a DPPC bilayer were suggested to be the active and inactive conformations respectively [18]. However, as shown previously for the epidermal growth factor receptor transmembrane domain, increasing the acyl chain length of lipids in the bilayer from 12 to 20 carbons can affect the dynamics of a transmembrane domain in simulations [92]. Indeed, changes in the tilt angle of a transmembrane domain as well as changes in the NMR structure between pure lipid and raft-like bilayers have been reported for other membrane receptors [93, 94]. The simulations presented here support the studies that suggest the membrane composition heavily influences the structure and behaviour of proteins embedded in the membrane. It would be a mistake however, in light of the results presented in Chapter 3, to interpret or propose a detailed mechanism of how membrane-embedded proteins function without considering the restrictions imposed by the extracellular and intracellular domains in these cases.

5.5 Conclusion

Here we asked whether it was important in the case of the type-I cytokine receptors to consider the membrane composition when constructing a model of the receptor embedded in the membrane. By considering a range of family members in bilayers with and without cholesterol we demonstrated variations in both the structure and behaviour of the transmembrane domains. While it would be improper to extrapolate any mechanistic details from these simulations owing to the lack of the extracellular domain, there does exist the possibility of a regulatory role for the membrane during the activation of the receptor. Further studies into whether this role is due to the presence of cholesterol itself or the generally more ordered and thicker structure of the raft-like bilayer that cholesterol promotes would be of interest in pursuing this idea. The diverse behaviour of the receptor transmembrane domains would also oppose using any single receptor as a model for the entire family.

Chapter 6

Generation and Validation of Palmitoyl Sphingomyelin Parameters and Lipid Raft-Like Bilayers



6.1 Introduction

The plasma membrane is a lipid matrix that physically separates the interior of cells from the external environment. Associated with and embedded in this structure are a wide range of membrane proteins. Some of these proteins are responsible for the transfer of chemical signals across the membrane, for example the type-I cytokine receptors. Others are transporters that facilitate the transfer of metabolites through the membrane, such as P-glycoprotein. The fluid mosaic model of the membrane proposed the lipids and membrane proteins were essentially randomly distributed throughout the membrane [95]. However, the proposal of membrane rafts within membranes challenged this idea and suggested the membrane is able to organise and compartmentalise cellular processes [40]. Early concepts of membrane rafts suggested that these domains had a regulatory role on membrane proteins including cell surface receptors [39]. Indeed, it has been suggested that the activation of some of the type-I cytokine receptors requires the co-localisation of the receptor with membrane rafts [38]. What is unclear however, is how the lipid components are able to regulate receptor behaviour. The spatial organisation and ordered nature of rafts suggests two potential mechanisms by which they could regulate receptors: (i) that they cause a spatial separation of a receptor from its secondary-messenger target molecule; or (ii) they induce a conformational change in the protein as it moves between the ordered and disordered phases of the membrane. Spatial separation of target proteins within a signalling pathway has been demonstrated in the case of the erythropoietin receptor (EPOR) and the Lyn kinase signalling pathway [38]. While the Lyn kinase was observed to be constitutively located within lipid rafts, the EPOR co-localises with lipid rafts upon stimulation with erythropoietin (EPO). What has yet to be determined is whether the type-I cytokine receptors adopt a different conformation when in a raft compared to the bulk membrane. Understanding how lipids can regulate proteins could provide new avenues to pursue in drug targeting of membrane proteins. These questions are difficult to probe experimentally but can be examined in atomic detail using molecular dynamics (MD) simulations. In order to be able to study the receptors in a raft-like membrane force field parameters are needed for the lipid molecules as well as the generation of the membrane structure.

The three major lipid families found in mammalian membranes are the glycerophospholipids, sphingolipids and sterols, e. g. phosphatidylcholines (PCs), sphingomyelins and cholesterol respectively (Figure 1.5). These lipids are found in varying ratios depending on the cell surface, e. g. the basolateral or apical surfaces [96, 97], the inner or outer leaflet [98], life cycle, cell type and organism. Membrane rafts are the result of further segregation of these lipids via lateral organisation within a membrane. The rafts are small domains (10-200 nm) that generally have a higher degree of order than the bulk membrane and are enriched in sterols and sphingolipids and can compartmentalise cellular processes [40]. Mammalian sphingolipids are based on the molecule sphingosine ((4*E*)-D-*erythro*-2-aminooctadec-4-ene-1,3-diol). *N*-acylation of sphingosine generates a ceramide. Sphingomyelin is a common sphingolipid found in membranes and is derived from a ceramide by the addition of a phosphocholine headgroup through an ester linkage. The overall structure is reminiscent of that of a phosphatidylcholine (Figure 1.5). An example of a sphingomyelin, *N*-palmitoyl-D-*erythro*-sphingomyelin (PSM), is shown in Figure 6.1. The length of the acyl chain of sphingomyelins varies with tissue expression. For example palmitoyl is predominantly found in erythrocytes [43], while stearoyl is more common in nervous tissue [44, 45]. With regard to the lateral distribution of sphingomyelin experimental analysis of the lipid composition of detergent-resistant membranes (often considered to be representative of the membrane raft *in vivo*) from rat neurons gave the molar ratio of sphingomyelins, cholesterol and glycerophospholipids of a raft as 1:3:6 compared to the cell homogenate of 1:5:32 [99]. The raft-like membrane of the human immunodeficiency virus (HIV) is also enriched in sphingomyelins by a factor of 3 compared to a T-cell leukaemia membrane [100]. The locally increased amounts of sphingomyelins and cholesterol give rise to the raft domains that are in a liquid-ordered phase (L_o) within a bulk membrane in the liquid-disordered phase (L_d) (also referred to as the liquid-crystalline phase).

The lateral segregation and generation of rafts is driven in part through the different properties of phosphatidylcholine (PC)s and sphingomyelins. Compared to the glycerophospholipids, for example 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphocholine (POPC) or 1,2-dioleoyl-*sn*-glycero-3-phosphocholine (DOPC), sphingomyelins tend to have a higher degree of acyl chain saturation. Sphingomyelins are also able to form a larger number of hydrogen bonds through the H-bond donor (hydroxyl and amide) and H-bond acceptor (amide) groups compared with the PCs that only contain H-bond acceptor (ester) groups at pH 7. These features give different gel-to-liquid crystalline transition temperatures of glycerophospholipids and sphingomyelins. For example, the transition temperatures for POPC and DOPC, both of which have unsaturated acyl chains are 269 K (-4°C) and 254 K (-19°C) respectively [44], whereas the transition temperature for PSM is between 311 K-314.5 K (38°C-41.5°C)[101, 44, 102, 103, 104]. Experimentally cholesterol is observed to preferentially co-localise with sphingomyelins over phosphatidylcholines [105]. Together these properties are suggested to account for the phase separation of the lipids in the rafts from the surrounding membrane resulting in these microdomains having tighter packed and more ordered lipids. While parameters for PC lipids have been extensively tested in MD simulations using the GROMOS 54A7 force field, parameters for sphingomyelins have not been widely used. In order to generate raft-like membranes, reliable parameters for sphingomyelins need to be generated and validated against the available experimental data.

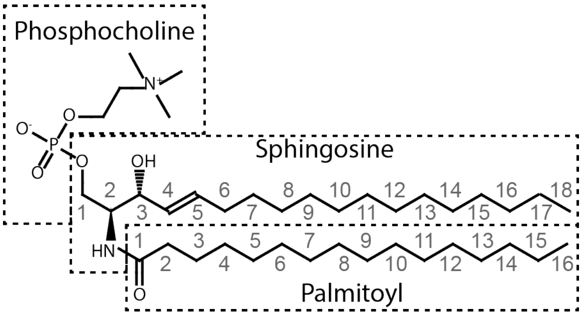


FIGURE 6.1: Molecular structure of *N*-palmitoyl-D-*erythro*-sphingomyelin (PSM). PSM consists of a sphingosine backbone linked to a phosphocholine headgroup and a palmitoyl acyl chain

Experimental data on the structural properties of PSM and other sphingomyelins are available but relatively scarce compared to the wealth of data available on glycerophospholipids. Specifically, properties such as the area per lipid, bilayer thickness and the degree of ordering of acyl chains have been investigated in a few of sphingomyelin bilayers in the gel and liquid-crystalline phases as well as in the presence or absence of cholesterol, see Tables 6.1 and 6.2. Parameters for *N*-stearoyl-D-*erythro*-sphingomyelin (SSM) using the GROMOS force field have been previously developed and examined in a pure bilayer containing 1,600 lipid molecules for 2 ns [106], with mixtures of cholesterol for 10 ns [107], or with cholesterol and DOPC for 20 ns or 200 ns [108, 109]. These parameters were derived from those of 1,2-dipalmitoyl-*sn*-glycero-3-phosphocholine (DPPC). The partial charges for the SSM were obtained based on Hartree-Fock calculations, at the HF/6-31G* level, of sphingomyelin truncated at carbon 6 on the sphingosine tail and carbon 3 on the acyl chain [106]. Two parameter sets for PSM have also been studied, the first generated from GROMACS and GROMOS building blocks [110] and the second derived from existing parameters for DPPC, POPC and GROMACS building blocks [111]. These parameters have been used to study the properties PSM as a single molecule in solution, as well as in 8 ns and 50 ns simulations of 128 PSM molecules in a bilayer, and also 100 ns simulation of a ternary mixture of POPC, PSM and cholesterol at different ratios [110, 111, 112]. Overall these studies demonstrated that the use of these parameters the bilayers were stable with properties such as the area per lipid (A_{L}), bilayer thickness ($d_{\text{p-p}}$) and chain order parameters ($|S_{\mathrm{CD}}|$) being close to experimental values (Tables 6.1). For example, the A_{L} and $d_{\text{p-p}}$ ranged between 0.52-0.6 nm^2 and 3.6-4.3 nm respectively in these simulations that were performed at temperatures between 293-325 K [110, 106, 111]. These simulations also showed a higher degree of ordering in the acyl chains when compared with PCs and greater propensity to form hydrogen bonds with cholesterol.

While in some instances these studies used relatively large membranes, for example 1,600 lipid molecules [106], and produced long timescales (3 ns) runs for the size of the systems when initially presented, these are now relatively short simulations given the computing power of today. Properties of membranes that result over milliseconds to seconds experimentally are still beyond what can be reasonably expected of today's computers and may never be captured by MD simulations. Nevertheless, improvements can be made to the force field parameters describing the molecules as well as in the tools used to generate the parameters. For example, Poger *et al.* [113] proposed a revised set a parameters for glycerophospholipids that modified the Lennard-Jones parameters between the choline atoms and the non-ester phosphate oxygens. These parameters allowed for a system of DPPC molecules to spontaneously assemble into a bilayer in a liquid-crystalline phase in solution. Automation in generating parameters is another important area of force field development and is available via tools such as the Automated Topology Builder (ATB) [65]. To obtain a set of self-consistent parameters to describe lipid molecules and to test the ability of the ATB [65] to produce reliable molecular topologies we aimed to test the quantum mechanics (QM) calculated charges for the sphingosine backbone in *N*-palmitoyl-D-*erythro*-sphingomyelin (PSM) produced by the ATB in conjunction with the phosphocholine parameters of Poger *et al.* [113]. In parallel we aim to produce a set of equilibrated bilayers containing PSM that can be used in future MD studies. In particular for use in understanding what structural and behavioural effects a raft-like bilayer imposes on a membrane protein, for example the type-I cytokine receptors.

Table 6.1: Experimental values of the area per lipid (A_L) and bilayer thickness ($d_\text{p-p}$) as a function of temperature in pure and mixed PSM bilayers.

	10°C	29°C	37°C	50°C	55°C	Ref.
A_L (nm^2)[a]	0.55	0.41	-	0.6-0.64	0.47	[101, 114, 115]
$d_\text{p-p}$ (nm)[a]	3.8	4.8-5.0	4.52-4.66	3.5-3.7	4.2-4.4	[101, 114, 115, 116]
$d_\text{p-p}$ (nm) 50:50[b]	-	4.5	-	-	4.5	[115]
$d_\text{p-p}$ (nm) 80:20[c]	-	-	4.34-4.56	-	-	[116]

[a] 100% PSM

[b] 50:50 mixture of PSM and cholesterol.

[c] 80:20 mixture of PSM and cholesterol.

Table 6.2: Experimental values of the palmitoyl chain order parameter ($|S_\text{CD}|$) for bilayers containing sphingomyelin.

Bilayer composition	$\langle\lvert S_\text{CD}^\text{plat}\rvert\rangle^*$	$\langle\lvert S_\text{CD}\rvert\rangle$	Temperature (K)	Ref.
PSM-d_{31}	0.21	-	293	[117]
PSM-d_{31}	0.26	0.22	313	[104]
PSM-d_{31}	0.25	-	318	[117]
PSM-d_{31}	0.26	0.22	321	[118]
PSM-d_{31}/Chol (2:1)[a]	0.35-0.45	-	293-333	[117]
PSM-d_{31}/POPC/Chol[b]	0.35	0.30	313	[104]
POPC-d_{31}/PSM/Chol[b]	0.29	0.23	313	[104]

* The average plateau is calculated over the carbon positions 4-6 in the palmitoyl chain, see Reference [118].

[a] Molar ratio 2:1.

[b] Molar ratio 37.5:37.5:25

6.2 Methodology

6.2.1 Systems simulated

The properties of N-palmitoyl-D-*erythro*-sphingomyelin (PSM) were examined in a pure bilayer consisting only of PSM and in mixed bilayers containing different concentrations of cholesterol (Table 6.3). The starting configuration for the pure PSM bilayer and PSM:cholesterol mixtures were obtained by progressively removing cholesterol molecules from a bilayer containing 720 PSM and 720 cholesterol molecules (PSM$_{50}$:CHOL$_{50}$). To generate the PSM$_{50}$:CHOL$_{50}$ bilayer initially, a randomly arranged patch of ten PSM and ten cholesterol molecules was generated and replicated into a 3×3 grid containing 90 PSM and 90 cholesterol molecules. The grid was further replicated three times with each replicate rotated either 90°, 180° or 270° around the normal to the grid. The four grids were merged together to create a monolayer of 360 PSM and 360 cholesterol molecules. The monolayer was duplicated and rotated 180°around an axis orthogonal to the normal of the monolayer and aligned with the first monolayer to create a bilayer that contained 720 PSM and 720 cholesterol molecules (Figure 6.2). This system was hydrated and relaxed by a steepest decent energy minimisation followed by a further 10 ns MD simulation. To create the PSM$_{60}$:CHOL$_{40}$ bilayer an equal number of cholesterol molecules from each leaflet were deleted by random selection from the final configuration of the MD simulation. This process of energy minimisation, MD relaxation and deletion of cholesterol molecules was repeated to successively obtain the PSM$_{70}$:CHOL$_{30}$, PSM$_{80}$:CHOL$_{20}$, PSM$_{90}$:CHOL$_{10}$ and pure PSM bilayers (Table 6.3). A larger pure PSM bilayer containing 1440 PSM molecules was generated by merging together the configurations of the pure PSM bilayers simulated at 293 K and 333 K (PSM_L and PSM_H bilayers respectively) after simulation of the bilayers for 360 ns and 210 ns respectively (Table 6.3).

A membrane raft-like bilayer was also investigated which consisted of an equimolar ternary mixture of PSM with POPC and cholesterol (POPC:PSM:CHOL) (Table 6.3). The initial coordinates of the POPC:PSM:CHOL bilayer were obtained by generating a randomly arranged patch of 5 POPC, 5 PSM and 5 cholesterol molecules that was replicated into a 3×3 grid. The grid was further replicated three times with each replicate rotated either 90°, 180° or 270° around the normal to the patch. The four patches were merged together to create a monolayer of 180 POPC, 180 PSM and 180 cholesterol molecules. The monolayer was duplicated and rotated 180° around an axis orthogonal to the normal of the monolayer and aligned with the first monolayer to create a bilayer that contained 360 POPC, 360 PSM and 360 cholesterol molecules (Figure 6.2).

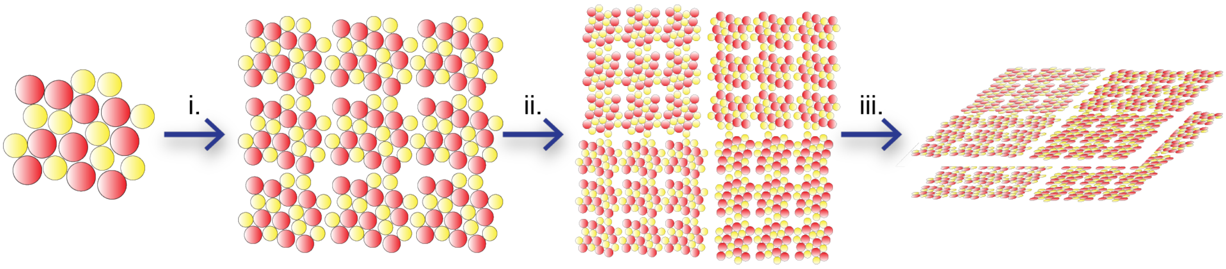


Figure 6.2: The bilayers were generated from a randomly arranged patch of lipids that were replicated into a 3×3 grid (i.). The grid was further duplicated with each duplicate rotated 90° 180°or 270° (ii.) to create a monolayer. The monolayer was duplicated, rotated and aligned to create a bilayer (iii.)

Table 6.3: Overview of the PSM containing membranes simulated.

System	Lipid species	Number of lipids	Temperature (K)	Simulation time (ns)
PSM_L	PSM	720	293	900
PSM_H	PSM	720	333	400
PSM_{90}:$CHOL_{10}$	PSM:Cholesterol	720:80	293	400
PSM_{80}:$CHOL_{20}$	PSM:Cholesterol	720:180	293	400
PSM_{70}:$CHOL_{30}$	PSM:Cholesterol	720:310	293	400
PSM_{60}:$CHOL_{40}$	PSM:Cholesterol	720:500	293	400
PSM_{50}:$CHOL_{50}$	PSM:Cholesterol	720:720	293	400
POPC:PSM:CHOL	POPC:PSM:Cholesterol	360:360:360	293	600
PSM_{293}	PSM	1440	293	25
PSM_{296}	PSM	1440	296	25
PSM_{299}	PSM	1440	299	25
PSM_{302}	PSM	1440	302	25, 25
PSM_{305}	PSM	1440	305	25, 25
PSM_{308}	PSM	1440	308	25, 25
PSM_{311}	PSM	1440	311	25, 25
PSM_{315}	PSM	1440	315	25
PSM_{318}	PSM	1440	318	25
PSM_{333}	PSM	1440	333	25

POPC, 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphocholine; PSM, *N*-palmitoyl-D-*erythro*-sphingomyelin

6.2.2 Simulation parameters

All simulations were performed using the GROMACS simulation package 4.6.7 [50] in conjunction with the GROMOS 54A7 united-atom force field [48]. The force field parameters for N-palmitoyl-D-*erythro*-sphingomyelin (PSM) were partly based on the parameters for DPPC developed by Poger *et al.* [113]. Specifically, the phosphocholine headgroup and the palmitoyl chain were derived from DPPC, whereas for the sphingosine backbone, the parameters were obtained from the ATB [65] and derived from the model molecule N-acetylhexasphingosine-O-phosphocholine (ATB molecule ID: 9232). Partial charges for the N-acetylhexasphingosine-O-phosphocholine were obtained from QM calculations performed by the ATB using GAMESS-US [119]. The molecule was initially optimised at the PM3 level of theory [120] and further optimised at the B3LYP/6-31G* level of theory [121, 122, 123] in an implicit solvent described using a polarizable continuum model [124]. All charges from the QM calculations were rounded to one decimal place. The partial charges for each atom are shown in Figure 6.3 and the complete topology file for PSM can be found in Appendix A. If needed, charge groups were manually optimised. In particular, carbon 1 in α position of phosphocholine in hexasphingosine was placed in the same charge group as the amide group in order to ensure compatibility with the charge groups used to describe PC in the GROMOS 54A7 force field. The resulting excess positive charge was distributed evenly over the atoms in the charge group. Parameters for POPC and cholesterol were also obtained from the ATB; molecule IDs: 1506 [125] and 1731 respectively.

Each system was simulated under periodic conditions in a rectangular triclinic box. The pressure was maintained at 1 bar by weakly coupling the system to a semi-isotropic pressure bath [53] using an isothermal compressibility of 4.6×10^{-5} bar^{-1} and a coupling constant $\tau_P = 1$ ps in the XY and Z directions (corresponding to the lateral and normal directions with respect to the bilayer). The temperature of the system was maintained by independently coupling the lipids and water to an external temperature bath with a temperature (T) (Table 6.3) and a coupling constant $\tau_T = 0.1$ ps using a Berendsen thermostat [53]. All bond lengths within the lipids were constrained using the LINCS algorithm [66]. Water was included explicitly in the simulations using the Simple-Point Charge (SPC) model [67] and constrained using the SETTLE algorithm [68]. The explicit hydrogen atom in the amide group of PSM was replaced by a virtual interaction site, the position of which was constructed at each step from the coordinates of the nitrogen atom to which it is attached as described by Feenstra *et al.*[71]. This allowed a 4-fs time step to be used without affecting the thermodynamic properties of the system significantly. Nonbonded interactions within the cutoff of 1.4 nm were calculated every step and the pair list was updated every five steps. A reaction-field correction was applied to the electrostatic interactions beyond the cutoff of 1.4 nm [69], using a relative dielectric permittivity constant $\epsilon_{\mathrm{RF}} = 62$ as appropriate for SPC water [70]. All systems, excluding PSM$_{\mathrm{H}}$, were energy-minimised. The temperature of each system was then gradually increased from 50 K to 293 K in 50-K steps over 120 ps to further relax the system and obtain the starting configurations used in the simulations.

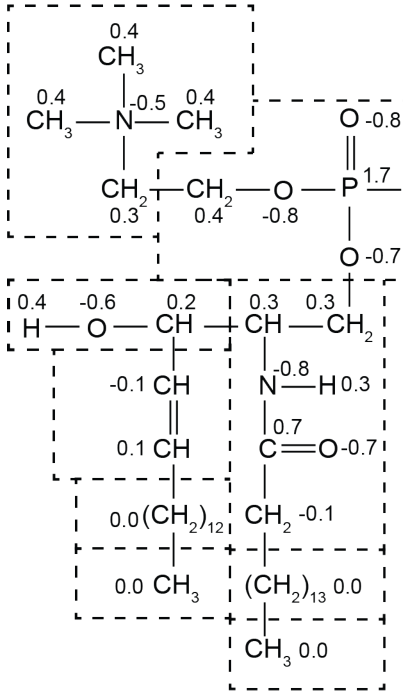


Figure 6.3: The charge groups and the partial charges for each atom in PSM. The charges and atom types for the phosphocholine groups were taken from Poger *et al.* [113]. The partial charges and atom types for the sphingosine head groups were obtained from the QM calculations performed by the ATB [65]. The remaining charges and atom types were obtained from the GROMOS54A7 force field.

6.2.3 Analysis

6.2.3.1 Area and volume per lipid

The area per lipid (A_L) was calculated in two ways. Firstly, for the pure PSM bilayers the A_L was taken as the lateral area of the simulation box divided by the number of lipids per leaflet. Secondly, the program GridMAT-MD was used for the bilayer mixtures [126]. In this method each lipid molecule was represented as a reference point and assigned to the upper or lower leaflet. Each leaflet was divided into a 200 × 200 grid in the lateral dimension. The resulting grid was then divided into polygons with each polygon associated to a single lipid reference point. The number of grid cells within each polygon was then evaluated for each lipid reference point [126]. The atoms taken for the reference points were the phosphorus atoms in POPC and PSM and the oxygen of the hydroxl group in cholesterol.

The volume per lipid (V_L) was calculated using the equation:

$$V_\mathrm{L} = \frac{V - n_w V_w}{n_\mathrm{L}} \tag{6.1}$$

where V is the volume of the simulation box, n_L and n_w are the number of lipid and water molecules respectively. V_w is the volume per water molecule for SPC. At 293 K $V_w = 3.06 \times 10^{-2}\,\mathrm{nm}^3$ and at 333 K $V_w = 3.16 \times 10^{-2}\,\mathrm{nm}^3$ obtained from independent simulations of water. The partial specific volume is calculated by dividing the volume per lipid by the mass of one PSM molecule and is equal to $1.167 \times 10^{-21}\,\mathrm{nm}^3 \cdot \mathrm{g}^{-1}$.

6.2.3.2 Bilayer thickness ($d_{\text{p-p}}$)

The bilayer thickness was calculated between the phosphorus atoms in the two leaflets from either the calculated electron density or from a grid-based nearest neighbour search [126]. The electron density profiles were calculated by dividing the simulation box into slices along the bilayer normal (Z-axis). The partial density for each atom in each slice was weighted by the number of electrons and averaged over the whole simulation. The distance was then calculated between the two maxima corresponding to the phosphorus atoms. The grid-based method assigned a reference point for each lipid to either the upper or lower bilayer leaflet. A 20×20 grid was aligned on each leaflet and for each grid point the nearest reference point in the opposite bilayer in the XY directions was used to calculate the thickness between the two leaflets for that grid point [126].

6.2.3.3 Lipid bond order parameter

The degree of order within a membrane can be probed experimentally using ^{2}H-NMR spectroscopy. Specifically, the carbon-deuterium bond order parameter S_{CD} provides a measure of the relative orientation of individual carbon-deuterium (C-D) bonds with respect to the overall magnetic field which is aligned with the bilayer normal. The order parameter S^i_{CD} of a methylene at position i is defined as:

$$S^i_{\text{CD}} = \frac{1}{2}\langle 3\cos^2\theta_i - 1\rangle \tag{6.2}$$

where θ_i is the angle between a C-D vector of the ith methylene in an acyl chain and the normal to the bilayer (Z-axis) (Figure 6.4). The angular brackets indicate an ensemble average. As the GROMOS force field uses an united-atom representation in which aliphatic hydrogens are not treated explicitly but incorporated into the carbons to which they are bound, the positions of the hydrogen atoms bound to a methylene carbon C$_i$ were constructed based on the positions of the neighbouring carbons (C$_{i-1}$ and C$_{i+1}$) assuming tetrahedral geometry.

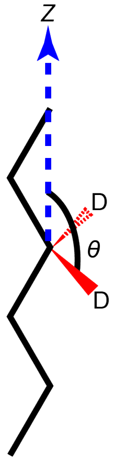


FIGURE 6.4: The carbon-deuterium bond order parameter is calculated from the angle formed between the the overall magnetic field which is aligned with bilayer normal (Z-axis) (blue) and the C-D bond (red) of the ith carbon atom in the acyl chain (black).

6.2.3.4 2-Dimensional radial distributions

The 2-dimensional radial distribution function between two atoms A and B is defined as:

$$\begin{aligned} g_{AB}(r) &= \frac{\langle \rho_B(r) \rangle}{\langle \rho_B \rangle_{local}} \\ &= \frac{1}{\langle \rho_B \rangle_{local}} \frac{1}{N_A} \sum_{i \in A}^{N_A} \sum_{j \in B}^{N_B} \frac{\delta(r_{ij} - r)}{4\pi r^2} \end{aligned} \tag{6.3}$$

where $\langle \rho_B(r) \rangle$ is the particle density of atom B at a distance r around atoms A, and $\langle \rho_B \rangle_{local}$ is the particle density of atom B averaged over all circles around atoms A with radius r_{max} in the XY plane, with r_{max} being half of the lateral box length.

6.2.3.5 Membrane bifurcation

The transitioning of a membrane between the gel and liquid-disordered (L_d) phases is a temperature dependent property at atmospheric pressure. Bifurcation was monitored by calculating the area per lipid (A_L) for a membrane composed of a 50/50 mixture of PSM_L and PSM_H membrane at a range of temperatures from 293 K to 333 K (Table 6.3). Using the numerical analysis package SciPy a 3-dimensional surface was then interpolated over all data points for time, temperature and A_L. The surface was used to produce curves of A_L versus temperature for multiple time points. The transition temperature was calculated by finding the average intersect between all curves in the range 0-12.5 ns with all curves in the range 12.5-25 ns.

6.3 Results

6.3.1 Area and volume per lipid of PSM

The pure PSM bilayers at low and high temperatures (systems PSM_L and PSM_H respectively) were analysed for equilibration by convergence in the area and volume per lipid (Figure 6.5). Over the final 100 ns the two bilayers gave an area and volume per lipid of $0.58\,nm^2 \pm 0.00\,nm^2$ and $1.16\,nm^3 \pm 0.00\,nm^2$ at 293 K, and $0.63\,nm^2 \pm 0.00\,nm^2$ and $1.20\,nm^3 \pm 0.00\,nm^2$ at 333 K. These volumes per lipid equate to partial specific volumes of 0.995 mL/g and 1.029 mL/g at 293 K and 333 K respectively. Upon introduction of cholesterol into the PSM bilayer, the A_L of PSM decreased to $0.55\,nm^2$ in the PSM_{90}:$CHOL_{10}$ bilayer (Figure 6.6A). Each subsequent addition of cholesterol decreased the A_L with the PSM_{50}:$CHOL_{50}$ bilayer having the smallest value of $0.43\,nm^2$. Cholesterol followed an inverse trend by having a higher A_L of $0.39\,nm^2$ in the PSM_{50}:$CHOL_{50}$ bilayer and a lower value of $0.36\,nm^2$ in the PSM_{90}:$CHOL_{10}$ (Figure 6.6B). The A_L values for PSM, POPC and cholesterol in the ternary mixture are $0.48\,nm^2$, $0.43\,nm^2$ and $0.36\,nm^2$ respectively (Figure 6.6C). The individual values for each bilayer are summarised in Table 6.4.

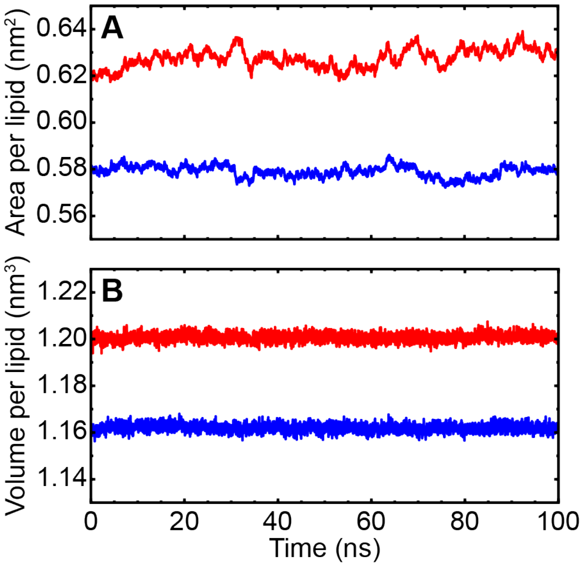


FIGURE 6.5: The area (A_L) (A) and volume (V_L) (B) per lipid from the pure PSM bilayers PSM_L at 293 K (blue) and PSM_H at 333 K (red). Note, 0 ns corresponds to the time at 800 ns and 300 ns in the PSM_L and PSM_H simulations respectively.

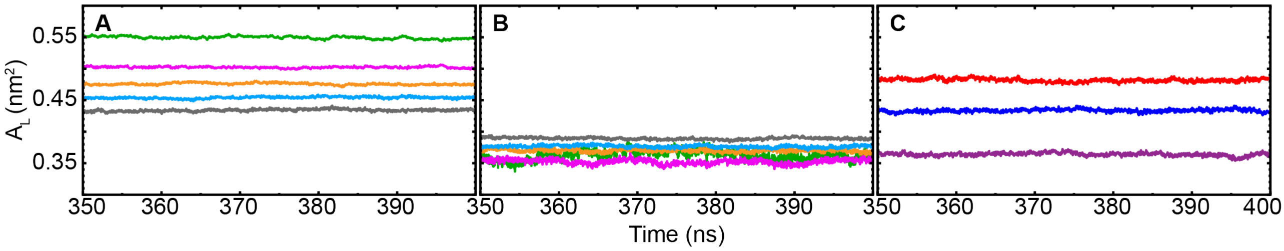


FIGURE 6.6: The area per lipid (A_L) for mixed bilayers containing PSM. The A_L was calculated for both PSM (A) and cholesterol (B) for the bilayers PSM_{90}:$CHOL_{10}$ (green), PSM_{80}:$CHOL_{20}$ (violet), PSM_{70}:$CHOL_{30}$ (orange), PSM_{60}:$CHOL_{40}$ (cyan) and PSM_{50}:$CHOL_{50}$ (grey). (C) The A_L values for POPC (blue), PSM (red) and cholesterol (purple) in the ternary mixture (POPC:PSM:CHOL).

6.3.2 Bilayer thickness of membranes containing PSM

The electron density profiles were calculated for all membranes and used to calculate the bilayer thickness between the peaks corresponding to the phosphate atoms ($d_{\text{p-p}}$). Increasing the temperature from 293 K to 333 K resulted in a decrease of $d_{\text{p-p}}$ from 3.8 nm to 3.6 nm in the pure PSM bilayers (Figure 6.7A). The presence of cholesterol increases the thickness of the bilayer to a maximum of 4.2 nm, which is reached after addition of at least 20% cholesterol (Figure 6.7B). The ternary mixture bilayer (POPC:PSM:CHOL) had the largest thickness of 4.4 nm (Figure 6.7C).

It is noted that the electron density profile for the PSM_L bilayer is asymmetrical (Figure 6.7A). Visual inspection of the final configurations for the PSM_L and bilayer reveals patches of the bilayer that are thicker than the surrounding bilayer (Figures 6.8A and B), whereas the PSM_H has no such domains of lipids (Figure 6.8C). These visual elements are reflected in the membrane thickness for these membranes calculated using the grid-based approach [126]. In the PSM_L the patches of bilayer could be as thick as 4.8 nm but also had valleys in the membrane with a thickness of 2.6 nm and gave an average thickness of 3.7 nm (Figure 6.8D). The PSM_H bilayer had a narrower range of values between 3.2 nm to 4.0 nm with an average thickness of 3.6 nm (Figure 6.8E). These values are summarised in Table 6.4.

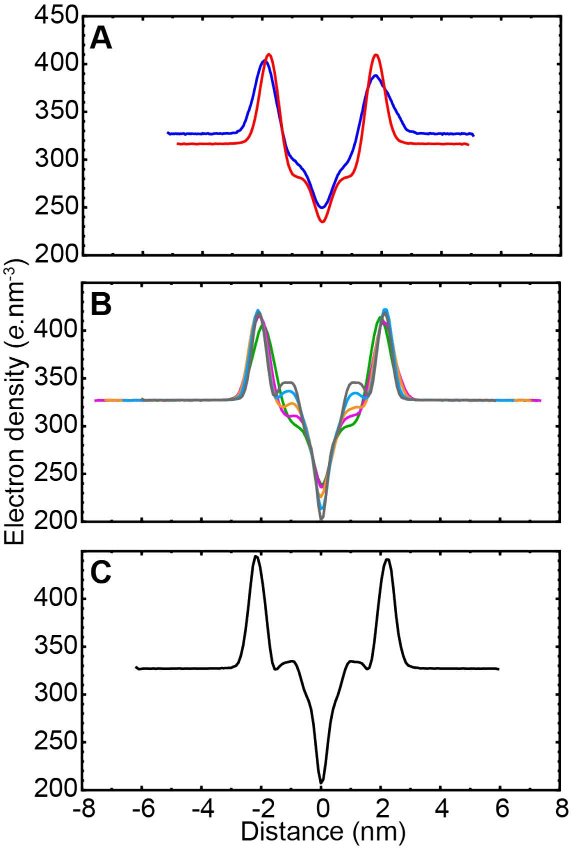


FIGURE 6.7: Electron density profiles across the hydrated bilayers containing PSM. (A) Pure PSM bilayers at 293 K (red) (PSM_L) and 333 K (blue) (PSM_H). (B) PSM bilayers containing increasing amounts of cholesterol: PSM_{90}:$CHOL_{10}$ (green); PSM_{80}:$CHOL_{20}$ (violet); PSM_{70}:$CHOL_{30}$ (orange); PSM_{60}:$CHOL_{40}$ (cyan); PSM_{50}:$CHOL_{50}$ (grey). (C) POPC:PSM:CHOL system.

6.3.3 Influence of temperature and lipid composition on the ordering of acyl chains in PSM

The $|S_{\mathrm{CD}}|$ profiles of the acyl tail in sphingosine and N-palmitoyl chains for the pure PSM bilayers and with increasing amounts of cholesterol are shown in Figure 6.9. The carbon numbering in Figure 6.9 corresponds to the numbering shown in Figure 6.1, that is from carbon 4 in sphingosine and carbon 2 in palmitoyl. The average $|S_{\mathrm{CD}}|$ order parameter for each carbon atom in each lipid is calculated independently for the sphingosine and palmitoyl chains and each averaged over 100 ns. In the pure PSM bilayers a similar profile is obtained in the sphingosine chain from carbon 4 to carbon 7 at 293 K and 333 K. However, from carbon 8 there is a continual decrease in $|S_{\mathrm{CD}}|$ at 333 K while a plateau is maintained from carbon 6 to carbon 13 at 293 K. The introduction of 10% cholesterol results in an increase in the order parameter of carbons 6-14 in the sphingosine chain and 4-13 in the palmitoyl chain, the effect is not maintained over the final 2-3 carbons with similar values to those obtained in the PSM_L bilayer being observed. Amounts of cholesterol over 20% led to an increase of the $|S_{\mathrm{CD}}|$ over the length of both chains. At all concentrations of cholesterol a plateau occurs from carbons 6-13 in the sphingosine chain (S-S_{CD}^{plat}). Experimentally the plateau in the palmitoyl chain (P-S_{CD}^{plat}) occurs over carbons 4-6 [118]. The average values for both plateaus (S-S_{CD}^{plat} and P-S_{CD}^{plat}) are presented in Table 6.5. In the ternary mixture (POPC:PSM:Chol) the $|S_{\mathrm{CD}}|$ profiles of the sphingosine chains are similar to what was observed for the PSM and cholesterol bilayer mixtures (Figures 6.9C and D). S-S_{CD}^{plat} and

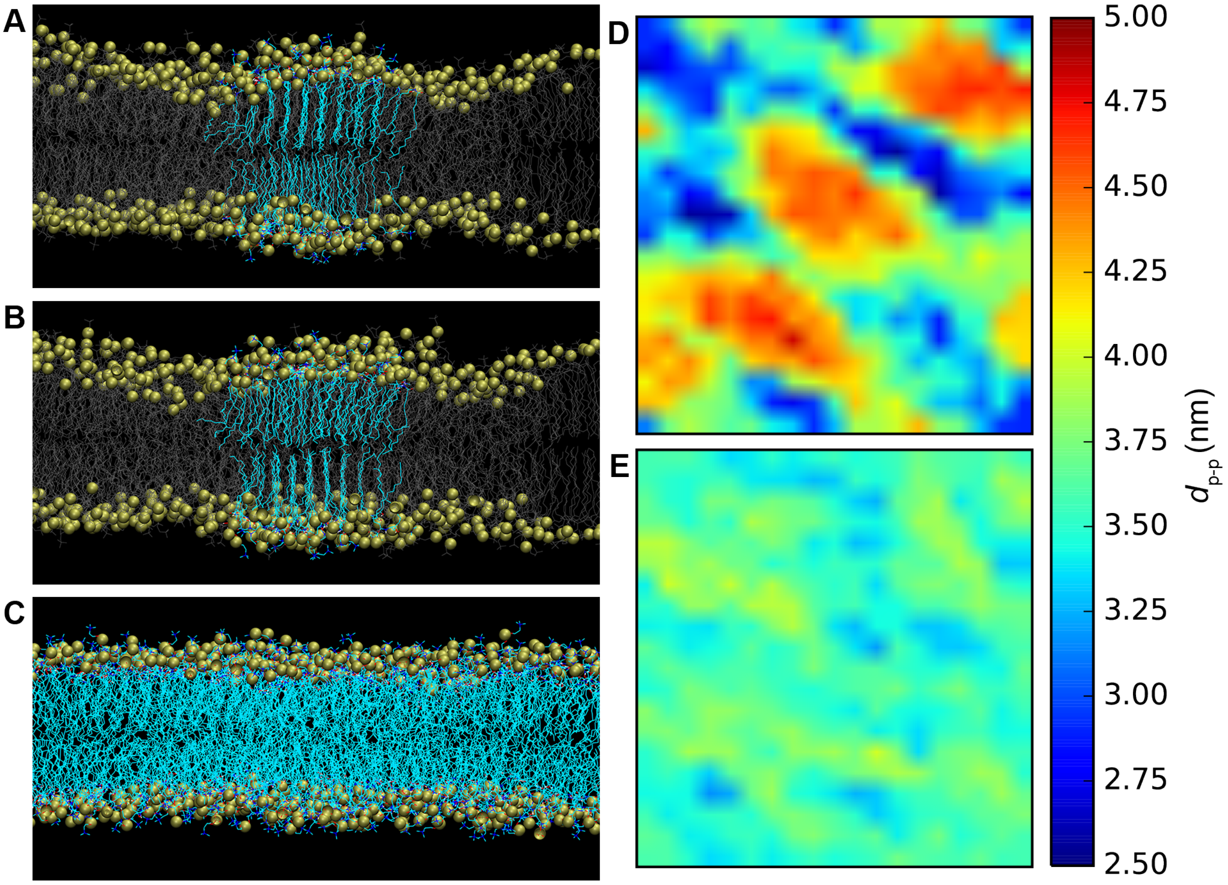


FIGURE 6.8: (A-C) Final configurations of the pure PSM bilayers at 293 K (PSM_{L}) and 333 K (PSM_{H}). (A) and (B) show the same patch of membrane at 293 K from different angles. The phosphate atoms of the lipids have been drawn as gold spheres and the carbon tails as cyan or grey lines. In (A) and (B) the lipids in a thicker region of the bilayer have been coloured cyan. (D) and (E) heat maps of the bilayer thickness calculated using GridMAT-MD [126]. Plots are the thickness averaged over the final 5 ns of simulation in the PSM_{L} (D) and PSM_{H} (E) systems. In panels (D) and (E) the X and Y-axes correspond to the plane of the membrane (XY-coordinates), with the corresponding thickness in the Z-direction coloured according to the colour bar scale.

P-S_{CD}^{plat} are 0.40 and 0.37 respectively (Table 6.5), higher than the values in the PSM_{70}:CLR_{30} mixture that would correspond to a similar ratio to cholesterol (Table 6.5). The acyl chains in POPC give a profile trace that is characteristic for the palmitoyl (Figure 6.9D) [127] and oleoyl (Figure 6.9E) [125] chains. The S_{CD}^{plat} in the POPC palmitoyl chain is higher at 0.40±0.02 than the P-S_{CD}^{plat} in PSM of 0.37. However, by the end of both the POPC and PSM palmitoyl chains there is no distinguishable difference between the $|S_{\text{CD}}|$ (Figure 6.9D).

It was noted above that the electron density profile for the PSM_{L} bilayer was asymmetric and there were patches of thicker membrane. To establish if this membrane is still evolving in regard to the order of the lipid tails, the $|S_{\text{CD}}|$ order parameter was calculated over 100 ns intervals for the carbon 6 and carbon 16 atoms of the sphingosine chain. These were selected because of their proximity to the start and end of the carbon chain respectively. Figure 6.10 shows the evolution of the distribution of the $|S_{\text{CD}}|$ for these atoms over the entire length of the simulation. The average and median values for carbon 6 stabilise within the first 200-300 ns, however growth of the upper quartile is observed over each time interval (Figure 6.10A). The distribution of the $|S_{\text{CD}}|$ for carbon 16 shows a continual increase in the average and median values, as well as continual growth of the upper quartile and reduction in the number of outliers (Figure 6.10B) over the length of the simulation.

Table 6.4: Area per lipid and bilayer thickness of membranes containing PSM.

	Area per lipid (nm^2)			Bilayer thickness (nm)	
System	PSM	Cholesterol	POPC	Electron density	GridMAT-MD [126]
PSM$_L$	0.58 ± 0.00	-	-	3.8	3.7 (2.6-4.8)
PSM$_H$	0.63 ± 0.00	-	-	3.6	3.6 (3.2-4.0)
PSM$_{90}$:CHOL$_{10}$	0.55 ± 0.00	0.36 ± 0.01	-	4.0	-
PSM$_{80}$:CHOL$_{20}$	0.50 ± 0.00	0.35 ± 0.00	-	4.2	-
PSM$_{70}$:CHOL$_{30}$	0.48 ± 0.00	0.37 ± 0.00	-	4.2	-
PSM$_{60}$:CHOL$_{40}$	0.45 ± 0.00	0.38 ± 0.00	-	4.2	-
PSM$_{50}$:CHOL$_{50}$	0.43 ± 0.00	0.39 ± 0.00	-	4.2	-
POPC:PSM:CHOL	0.48 ± 0.00	0.36 ± 0.00	0.43 ± 0.00	4.4	-

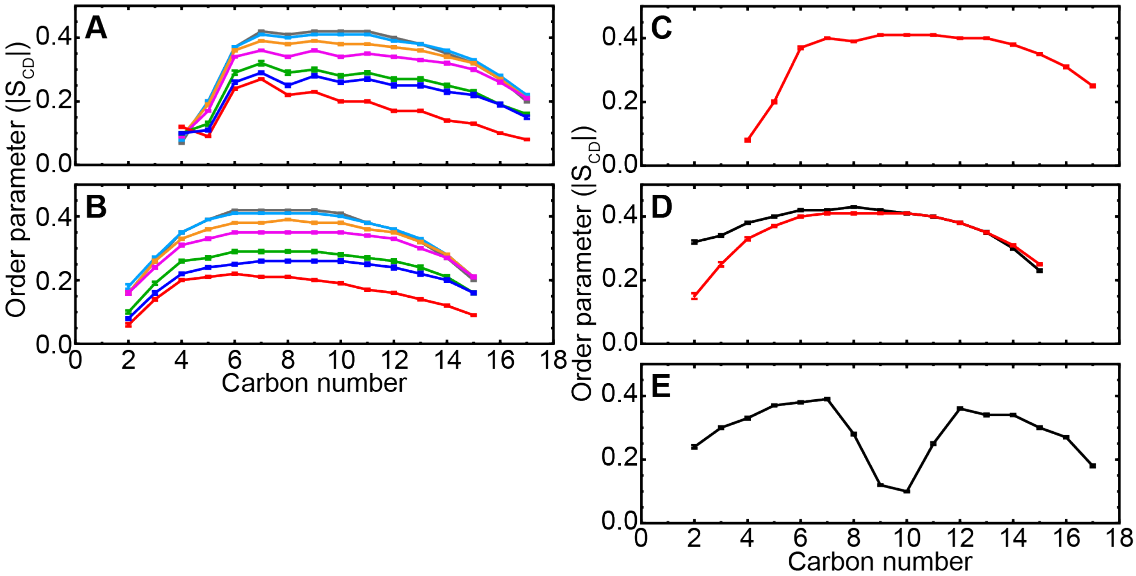


Figure 6.9: Carbon deuterium bond order parameter ($|S_{\mathrm{CD}}|$) profiles for bilayer containing PSM. (A) sphingosine and (B) palmitoyl chains from PSM in pure bilayers and with increasing amounts of cholesterol. PSM$_L$ (blue); PSM$_H$ (red); PSM$_{90}$:CHOL$_{10}$ (green); PSM$_{80}$:CHOL$_{20}$ (violet); PSM$_{70}$:CHOL$_{30}$ (orange); PSM$_{60}$:CHOL$_{40}$ (cyan); PSM$_{50}$:CHOL$_{50}$ (grey). (C-E) the profiles for POPC and PSM in the POPC:PSM:CHOL system. (C) sphingosine, (D) palmitoyl from POPC (black) and PSM (red) and (E) oleoyl chains. The order parameters were calculated for each 0.5 ns interval and averaged over the final 100 ns. Error bars are ±1 standard deviation.

6.3.4 Analysis of the distribution of lipids around cholesterol

The interaction of cholesterol with either PSM or POPC was examined by calculating the 2-D radial distribution function (g(r)) of the amide oxygen in PSM and the *sn*-1 or *sn*-2 carbonyl oxygens in POPC around the hydroxyl hydrogen from cholesterol as illustrated in Figure 6.11A and B. In bilayers containing only mixtures of PSM and cholesterol there is a single peak in the 2-D radial distribution with a height of between 6-7 in all membranes at around 0.15 nm (Figure 6.11C). This corresponds to a cumulative number of oxygen atoms around the hydroxyl hydrogen of between 0.30-0.52 at a distance of 0.25 nm (Table 6.6). In the ternary mixture, again there is only a single peak for both the PSM and either of the two POPC oxygens at 0.15 nm (Figure 6.11B). In this ternary bilayer PSM has a greater probability, with a distribution height over 9, than either of the two POPC oxygens with distribution heights of 4.5 and 2 for the oleoyl and palmitoyl chains respectively. These values correspond to a cumulative number of the oxygen atoms around the hydroxyl hydrogen of 0.29, 0.16 and 0.07 respectively at a distance of 0.25 nm (Table 6.6).

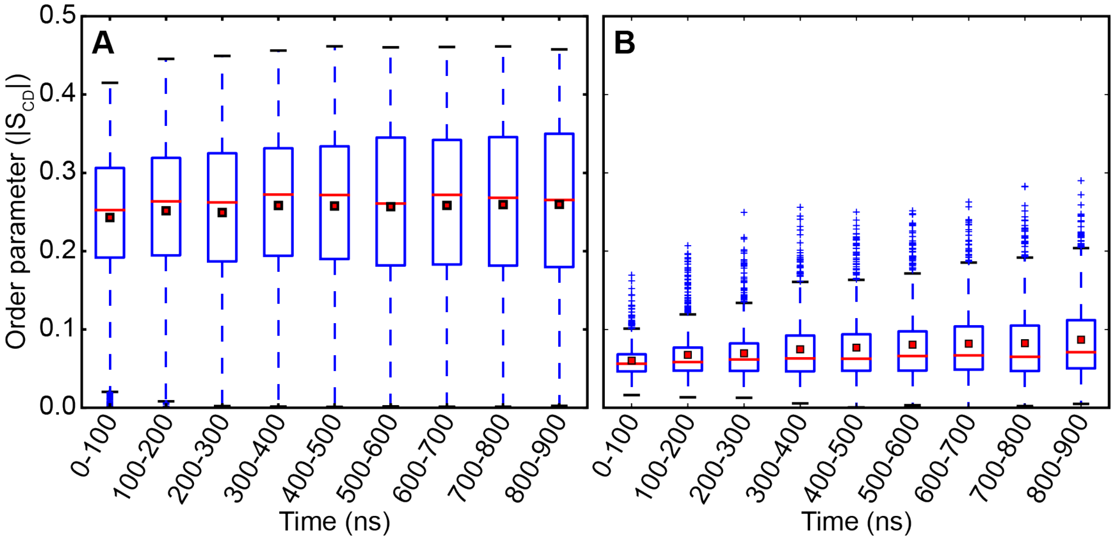


Figure 6.10: The order parameter for carbon 6 (A) and carbon 16 (B) in the sphingosine chain of PSM at 293 K (PSM$_L$). The order parameter was calculated over a 100 ns interval for the individual molecules in the system. Plotted are the average (red square), median (red line), upper and lower quartiles (box), and outliers (blue crosses).

Table 6.5: Average carbon-deuterium bond order parameter ($|S_{CD}|$) of the carbon tail plateau in lipid bilayers containing PSM.

	PSM		POPC
System	Sphingosine[a]	Palmitoyl[b]	Palmitoyl[b]
PSM$_L$	0.26±0.03	0.24±0.01	-
PSM$_H$	0.21±0.03	0.21±0.01	-
PSM$_{90}$:CHOL$_{10}$	0.29±0.02	0.27±0.01	-
PSM$_{80}$:CHOL$_{20}$	0.35±0.01	0.33±0.02	-
PSM$_{70}$:CHOL$_{30}$	0.38±0.01	0.36±0.02	-
PSM$_{60}$:CHOL$_{40}$	0.40±0.01	0.38±0.02	-
PSM$_{50}$:CHOL$_{50}$	0.40±0.02	0.39±0.03	-
POPC:PSM:CHOL	0.40±0.01	0.37±0.03	0.40±0.02

[a] Plateau calculated over carbon positions 6-16.
[b] Plateau calculated over carbon positions 4-6, see Reference [118].

6.3.5 Melting temperature of a pure PSM bilayer

The bifurcation of the pure PSM bilayer was examined by monitoring the change in the area per lipid (A_L) of a bilayer composed of a 50/50 mixture of PSM molecules that had been equilibrated at 293 K and 333 K (Figure 6.12A) (above and below the experimental melting temperature respectively). This bilayer was simulated at the range of temperatures listed in Table 6.3. The bifurcation of the A_L as a function of time and temperature is shown in Figure 6.12B. As a general statement, these simulations show a trend of increasing A_L with increasing temperature. At temperatures of 311 K and below the A_L at later time points (red points) are lower than the earlier time points (blue points). In simulations performed at and above 315 K the later time points give higher A_L values than than the earlier time points. To analyse all time points from the independent simulations performed at the different temperatures a surface was interpolated over all data points. The surface was used to generate the lines describing the change in A_L as a function of temperature for each time point (Figure 6.12B).

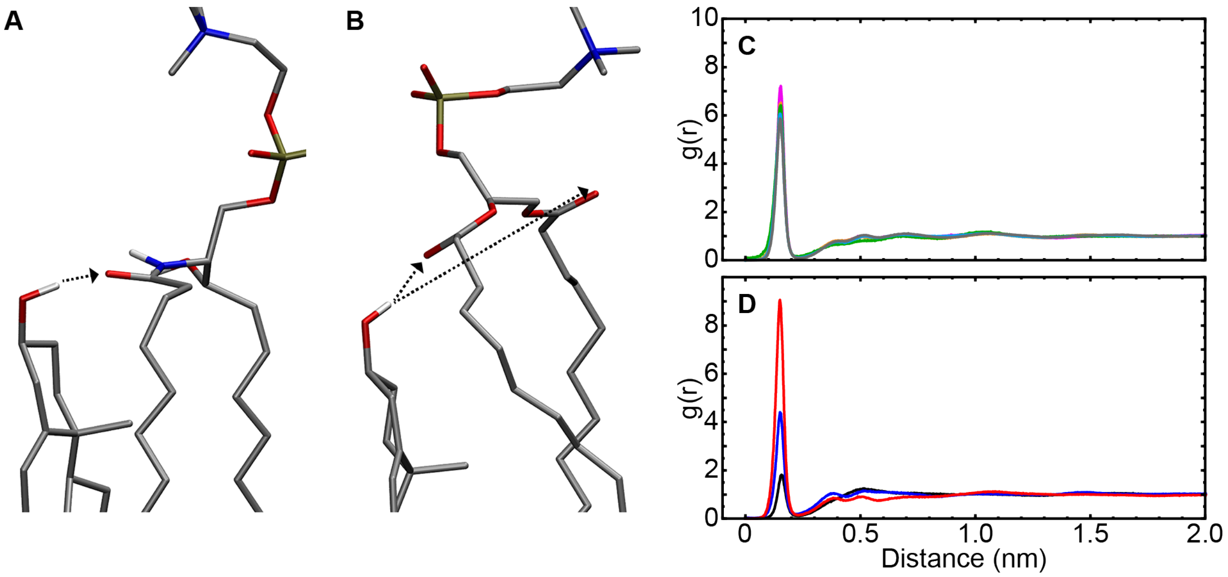


Figure 6.11: (A) and (B) the 2-dimensional radial distribution function (g(r)) was calculated between the hydroxyl hydrogen of cholesterol and the amide oxygen in PSM (A) and the *sn*-1 or *sn*-2 carbonyl oxygens in POPC (B). (C) g(r) for bilayers containing increasing amounts of cholesterol: PSM_L (blue); PSM_H (red); PSM_{90}:$CHOL_{10}$ (green); PSM_{80}:$CHOL_{20}$ (violet); PSM_{70}:$CHOL_{30}$ (orange); PSM_{60}:$CHOL_{40}$ (cyan); PSM_{50}:$CHOL_{50}$ (grey). (D) g(r) of the ternary mixture for PSM (red) and the oleoyl (blue) or palmitoyl (black) carbonyl oxygens in POPC.

Table 6.6: Cumulative number of cholesterol hydroxyl hydrogens within 0.25 nm of the PSM and POPC oxygen atoms.

		POPC carbonyl oxygen	
System	PSM amide oxygen	*sn*-1 (palmitoyl)	*sn*-2 (oleoyl)
PSM_{90}:$CHOL_{10}$	0.51	-	-
PSM_{80}:$CHOL_{20}$	0.52	-	-
PSM_{70}:$CHOL_{30}$	0.44	-	-
PSM_{60}:$CHOL_{40}$	0.37	-	-
PSM_{50}:$CHOL_{50}$	0.30	-	-
POPC:PSM:CHOL	0.29	0.07	0.16

The intersect of all the lines for 0-12.5 ns with all the lines for 12.5-25 ns were averaged to calculate a predicted melting temperature of 311.9 K ± 2.8 K (Figure 6.12B).

6.4 Discussion

The biological importance of sphingolipids cannot be understated. Not only are they a key lipid in the formation of rafts, but the ceramide derivatives are also used in cell signalling [128]. While structurally similar to PCs the sphingolipids have an amide group, and in general have more saturated carbon tails [44]. These structural differences are suggested to account for the generally higher gel-to-liquid crystalline transition temperatures of sphingomyelins compared to PCs [44] and the preferential packing of cholesterol with sphingomyelins seen in simulations [109, 129]. It is remarkable that cells have been able to exploit these physical properties of lipids in their membranes to create rafts that are able to regulate membrane protein function. Indeed, while the study of lipids and membrane structure is of itself an interesting and expansive area, an overarching goal of this work is to provide models of raft-like membranes to be able to study membrane proteins with MD simulations in a physiologically

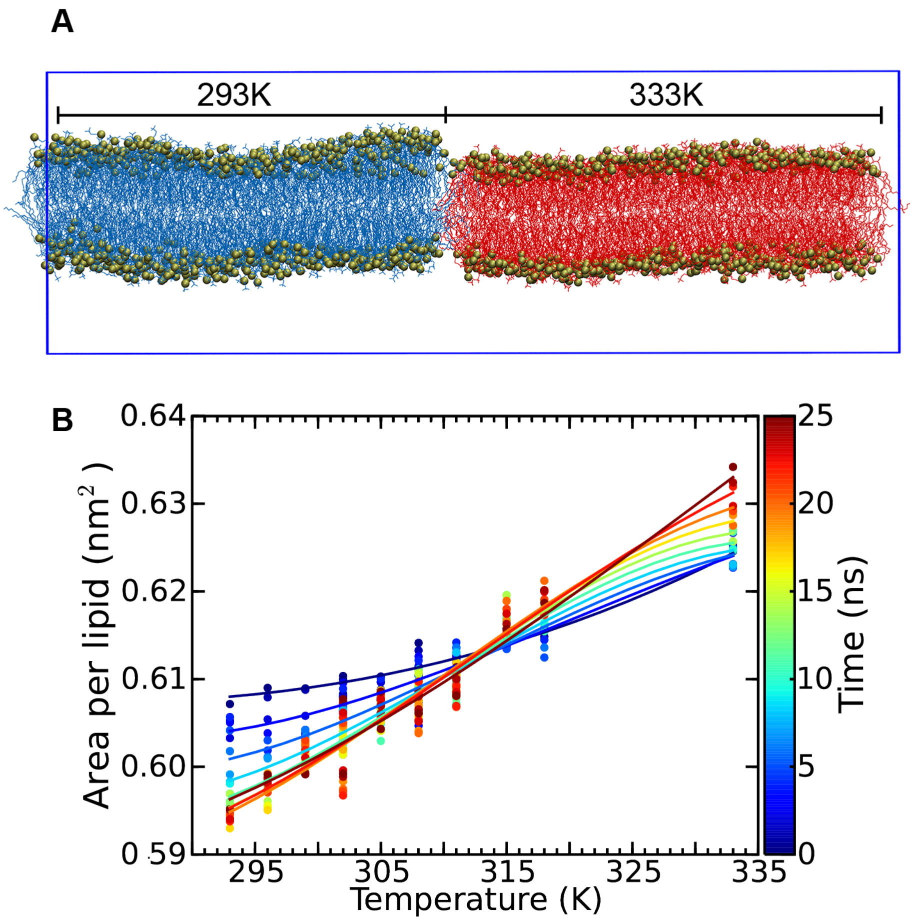


FIGURE 6.12: (A) The initial configuration of the pure PSM used to determine the melting temperature. A bilayer simulated at 293 K and another simulated at 333 K were merged together and placed in a single rectangular simulation box. (B) Bifurcation of the pure PSM bilayer was monitored by calculating the area per lipid (A_L) at a range of temperatures from 293 K to 333 K. The transition temperature was calculated from curves describing a 3-dimensional surface fitted to all the data points. The dashed vertical lines indicate the average intersect of the curves ± one standard deviation. For clarity only the average A_L value over 1 ns intervals is shown as points, and the lines correspond to the surface divided at 2 ns intervals.

relevant environment. These membranes therefore need the development of a set of self-consistent parameters for a range of lipids that have been validated against experimental data. To generate parameters for PSM that were self-consistent with existing parameters for POPC and cholesterol, we generated the parameters using the automated protocol of the ATB [65]. A common goal of force field development is to have parameters that are transferable between molecules containing similar functional groups. Here we tested the parameters developed for the phosphocholine headgroup for PCs with a sphingosine backbone.

Two common properties of lipid bilayers that can be inferred from experimental data such as X-ray diffraction data are the area per lipid (A_L) and bilayer thickness ($d_\text{p-p}$). For a pure PSM bilayer below the experimental melting temperature estimates range between 0.41-0.55 nm^2 and 3.8-4.6 nm. Above the experimental melting temperature these estimates range between 0.38-0.50 nm^2 and 3.5-4.4 nm (Table 6.1). The A_L for the PSM$_\mathrm{L}$ and PSM$_\mathrm{H}$ bilayers obtained in this work were, 0.58 nm^2 and 0.63 nm^2 respectively. While these fall outside the range of values estimated experimentally they reproduce the general trend of a lower A_L and thicker bilayer below the melting temperature and a higher A_L and thinner bilayer above the transition temperature. The large variation in values proposed based on experimental data make it difficult to know how or if these values should compare to the simulations. It also must be kept in mind that while the $d_{\text{p-p}}$ can be derived from X-ray data using a model of the lipid, the A_{L} is a derived property requiring knowledge of the partial specific volume of the lipid. In their work Maulik and Shipley [115] assumed values of the partial specific volume for PSM of 0.945 mL/g and 1.012 mL/g at 298 K and 328 K taken from DPPC. These are lower than the partial specific volumes obtained for the PSM_{L} and PSM_{H} simulations of 0.995 mL/g and 1.029 mL/g respectively. Further complicating the comparison of these values is the organisation of the bilayer leaflets. In their work Maulik and Shipley [115] report the PSM bilayers to be in a hexagonal phase, while the simulations presented here were in a lamellar phase. Care must therefore be taken not to over interpret differences between the simulations and the experimental values. It should also be noted that in a previous MD simulation of a pure PSM bilayer below the gel-to-liquid crystalline transition temperature yielded an A_{L} of 0.49 nm^2 [108], closer but still significantly higher than the 0.41 nm^2 as reported by Maulik and Shipley [115]. Together with our simulations this may suggest the A_{L} in the gel phase for PSM may not be as low as it has been reported experimentally.

The order parameters of the deuterium-carbon bonds in the lipid tails can provide information on the bilayer phase. Overall the trend in the order parameters for PSM are in agreement with what would be expected experimentally, with lower values and more disordered chains at high temperatures, and increased ordered with increasing concentrations of cholesterol. The profile of the order parameters for the palmitoyl chain in PSM however differ from what has been observed experimentally (see Figure 6.13). In particular the carbons at positions 2-4 have a lower value and do not form a plateau that has been observed by NMR [118]. In contrast the palmitoyl chain in POPC, in the ternary mixture (Figure 6.9D) has a closer profile to the experimental data. This disorder may be due to the difference in the charges placed on the carbons. While in POPC all the carbons in the palmitoyl chain do not have a charge, in PSM the QM calculations led to the placement a partial negative charge on carbon 2 (Figure 6.3). However, QM calculations performed by Chiu *et al.* [106] did not place a charge on this atom, yet in subsequent simulations using these parameters for PSM a similar curved profile was obtained over these atoms in the palmitoyl chain [108, 109]. This would suggest that there may be an inappropriate interaction of the parameters in the amide group causing the decreased order in the palmitoyl chain.

The preferential interaction of cholesterol with sphingomyelins over PCs is an often reported property suggested to play a major role in the segregation of membrane rafts from the bulk membrane. The interaction of cholesterol with PSM and POPC was examined by calculating the 2-D radial distribution of the hydroxyl hydrogens around the oxygen atoms in the lipid backbone. Interestingly as the cholesterol content increased the less cholesterol was found to be interacting with the amide oxygen in PSM (Table 6.6). A possible explanation for this behaviour may be less competition of cholesterol molecules for PSM which allows for longer interaction times between the atoms. This may also explain why the POPC:PSM:CHOL system has a similar cumulative number of cholesterol hydroxyl hydrogens around the PSM (0.29) to the PSM_{50}:CHOL_{50} (Table 6.6), but has a similar amount of cholesterol as the PSM_{70}:CHOL_{30}, as the POPC oxygens provide extra competing oxygens. Deeper analysis into the H-bonding network between, for example between the amide groups in PSM as well

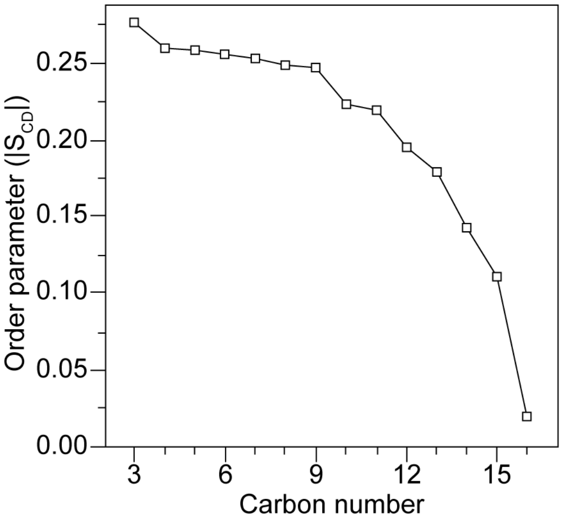


FIGURE 6.13: The NMR order parameter profile for PSM with a perdeutrerated acyl chain (PSM-d_{31}). Reproduced from fig. 4A from Mehnert *et al.* [118].

as the length of time for which the H-bonds exist will be needed to understand how these groups can promote the lateral separation of the different lipids.

The temperature at which the lipids undergo a phase transition is a property that has not been previously addressed in simulations of PSM-containing bilayers. The thermodynamic temperature of a MD simulation is dependent on the force field. To determine if states sampled by the molecule correspond to the reference temperature set one can determine the temperature at which the system undergoes a phase transition. For this we used a method that has previously been applied to PC lipids and estimated the transition by merging bilayers in different phases and simulating the resulting membrane at a range of temperatures [130]. While the previous method analysed tail dihedrals in a neighbour search algorithm to determine the fraction of lipids in the gel phase, our analysis simply used the change in A_{L} as an indicator of whether the two phase bilayer was moving towards a gel or liquid crystalline phase. By interpolating a surface over all data points we were able to construct curves that describe the A_{L} as a function of temperature. We then used the average intersect point of these curves to provide an estimate of the transition temperature. This method gave a result of 311.9 ± 2.8 K which is in good agreement of the experimental range of 311 K-314.5 K (most experiments give the exact value as 314.5 K) [101, 44, 102, 103, 104]. While the system setup here involved over five times as many lipids as what was used to determine the transition temperatures of PCs, this method shows that at reasonable accurate estimate of the transition temperature can be obtained using only 25 ns simulations as opposed to 100 ns simulations used in simulations of PCs [130]. Furthermore, a relatively simple description of the system, in this case the A_{L} is enough, and calculation of more complex properties such as order parameters or dihedral angles appears unnecessary.

In comparison to PSM described by other force fields, specifcally CHARMM [131, 132] and SLIPIDS [133], the performance of the parameters for the measured values A_{L}, $d_{\text{p-p}}$ and $|S_{\mathrm{CD}}|$ compared to the experimental values is somewhat poorer. In particular the CHARMM parameters produce values for the A_{L} between 0.52-0.57 nm^2 and a $d_{\text{p-p}}$ thickness of 4.0-4.4 nm at temperatures of 310-321 K and the Slipids parameters gave values of 0.54 nm^2 and 4.2 nm at 323K. The higher A_{L} and thinner $d_{\text{p-p}}$ reported here for PSM may indicate that the density of PSM is incorrect. This may be a product of the setup where cholesterol was deleted from the bilayer producing holes the PSM filled, but the bilayer did not compress to an ideal value. This is probably reflected in the $|S_{\text{CD}}|$, where the CHARMM and Slipids bilayers produced lipids with a higher degree of order, in particular around carbons 4-6 in the sphingosine chain, and more closely trace the order parameter profile produced from NMR experiments. It is interesting to note however, that the parameters described here produce a similar order parameter profile to the previous simulations performed with the GROMOS force field. Suggesting this may be a artefact from the united atom force field. The density of PSM could be addressed in two ways to acheive an A_{L} closer to experimental values; firstly the pure PSM system could be rebuilt at a lower A_{L}, or secondly cycle the bilayer through rounds of compression and relaxation by increasing and decreasing the reference pressure.

6.5 Conclusions

In this study we proposed new parameters for *N*-palmitoyl-D-*erythro*-sphingomyelin (PSM) generated in part from an automated tool but in part manually curated. Comparison to available experimental data suggests these parameters can reproduce physical properties of PSM, in particular the estimated transition temperature (311.9±2.8 K) was very close to the experimental value of 314.5 K. These simulations provide a new set of PSM containing bilayers that have self-consistent parameters with other lipid molecules developed for the GROMOS force field. Based on these results we have confidence in the use of these parameters and bilayer systems we have generated for future studies of membrane proteins in raft-like membranes.

6.6 Disclaimer

During the final write up of the results for this section an error was discovered in the parameters for sphingomyelin. The improper dihedral for the carbon-carbon double bond was set to 0° instead of the correct 180°. This resulted in the *cis* configuration of the double bond rather than the required *trans* configuration. Initial tests suggest this error does not alter the properties of sphingomyelin or the mixed bilayers. The values for the reported properties are being recalculated (Figure 6.14). The correct *trans* configuration is used for all simulations containing protein in the remainder of the thesis.

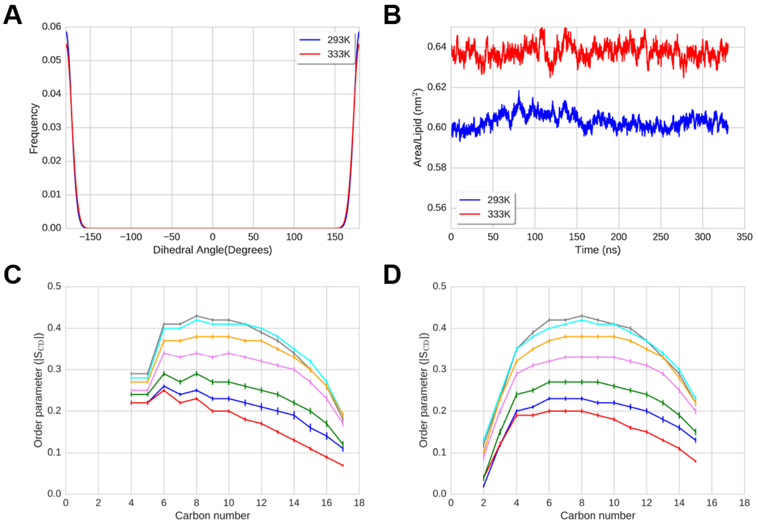
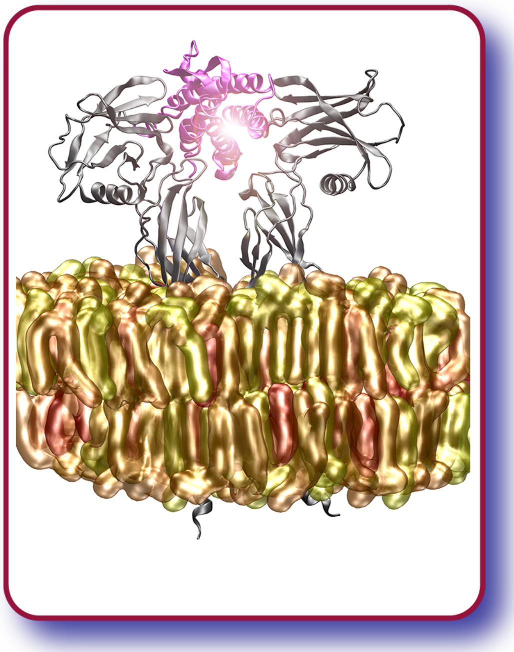


FIGURE 6.14: The recalculated properties of PSM with the *trans* double bond. (A) The distribution of angles for the double bond dihedral at 293 K and 333 K. (B) The evolution of the A_L values at 293 K and 333 K. (C, D) The order parameter values for the sphingosine and palmitoyl chains respectively over the final 100 ns of simulation: PSM$_L$ (blue); PSM$_H$ (red); PSM$_{90}$:CHOL$_{10}$ (green); PSM$_{80}$:CHOL$_{20}$ (violet); PSM$_{70}$:CHOL$_{30}$ (orange); PSM$_{60}$:CHOL$_{40}$ (cyan); PSM$_{50}$:CHOL$_{50}$ (grey). Values are the average ± the standard deviation.

Chapter 7

What are Active Conformations of a Type-I Cytokine Receptor Transmembrane Domain Dimer?



7.1 Introduction

In Chapters 2 and 3 we investigated whether the X-ray crystal structures of the erythropoietin receptor (EPOR) extracellular domain alone, that is without a transmembrane domain, was sufficient to determine active or inactive conformations of the dimer complex. Our reinterpretation of the available crystal structures has cast doubt on the physiological significance of the differences between these structures, and the simulations confirmed that these were not unique conformations of the receptor extracellular domain dimer. Indeed this work highlighted the need to consider the receptor in a more physiologically relevant environment, that is with the transmembrane domain embedded in a membrane environment. We subsequently explored, in Chapter 5, the effects of lipid composition on the transmembrane region embedded in bilayer for a range of the type-I cytokine receptors. The simulations indicated that lipid composition was an essential factor to consider in the activation of these receptors and therefore raft-like bilayer containing *N*-palmitoyl-D-*erythro*-sphingomyelin (PSM) and cholesterol were built and equilibrated in Chapter 6. The current Chapter builds upon this work: in particular we examine dimers of transmembrane domains from a range of type-I cytokine receptors. While computational techniques have been employed previously to predict and study possible dimer interfaces in the transmembrane domains for the EPOR and GHR, neither the effect of the membrane composition nor restrictions imposed by an extracellular domain have been taken into consideration [21, 82, 18]. As evident from our work in (Chapters 2-5), these factors will affect the structures obtained and draw into question the physiological significance of any result obtained without them. We therefore sought to improve upon our understanding of how type-I cytokine receptor transmembrane domain dimers interact in a membrane environment.

Currently no experimental structure of a dimer of a type-I cytokine receptor transmembrane domain exists. Furthermore, the structural details on how the extracellular domains and transmembrane domains are linked, i. e. whether the extracellular juxtamembrane region has secondary structure or is an unstructured loop, are not known. We therefore have focused on model systems in which dimers of the transmembrane domains were naturally conformationally restricted. Previously, the GHR transmembrane dimer was modelled in a pure 1,2-dipalmitoyl-*sn*-glycero-3-phosphocholine (DPPC) bilayer with a disulfide bond in the N-terminal region (proximal to the extracellular domain), associated with activation of the receptor, to obtain a range of dimer conformations. We also found in the literature a report of another disulfide-linked dimer associated with an oncogenic mutation in the interleukin-7 receptor subunit α (IL-7Rα) due to the insertion of the amino acid sequence CPT in the the N-terminal region of the transmembrane domain. Several chimeric constructs of a dimerisation motif from a transcription factor attached to the transmembrane domain of a receptor have also been studied experimentally. In particular these are the Put3 and mouse EpoR [21], and c-Jun and human GHR chimeras [18]. The Put3-EpoR chimeras were designed to induce a relative rotation between the two transmembrane domains by progressively deleting residues from the N-terminus of the transmembrane domain sequence. The Jun-GHR chimeras were suggested to affect the proxity of the N-termini of the transmembrane domains and thereby orientate the intracellular juxtamembrane domains.

Transcription-factor-transmembrane-domain chimeric constructs have been invaluable for understanding receptor activation. They demonstrate that dimerisation alone is insufficient for receptor activation, and that the orientation of the transmembrane domain also affects the activation of the downstream signalling domains. These constructs provide a relatively simple system in which the transmembrane domains of a type-I cytokine receptor are restricted by the presence of an extracellular region, but can change the activation state with the insertion or deletion of a single amino acid (Table 7.1). They are also informative in that both the transcription factor dimerisation domains and the transmembrane domains are helical, so it is likely that the full construct (up until the end of the transmembrane domain) will be helical. While these constructs have provided new insights into the activation of the receptors, it stills remains to be answered whether these chimeric constructs are appropriate substitutions for the receptor extracellular domain. That is, are the structure of the chimeras similar to that of the wild-type full-length receptor, or do they provide an alternative mechanism to activate the kinase domains? In this chapter, we examine the structures of several type-I cytokine receptor transmembrane domain dimers and ask whether the conformations of the constructs known to be constitutively active can be distinguished from inactive dimers.

7.2 Methodology

7.2.1 Systems simulated

Three sets of dimeric receptor constructs of alternative receptor transmembrane domains have been examined: (i) an oncogenic mutation of the IL-7Rα (referred to as ``P2 clinical mutant") that allows for a disulfide-linked dimer and constitutive activation of the receptor [134], (ii) chimeric constructs of the yeast Put3 transcription factor attached to the mouse EpoR transmembrane domain [21], and (iii) a chimeric construct of the Jun transcription factor attached to the human GHR transmembrane domain [18]. Sequences were taken from UniProt database entries P14753 (mEpoR), P10912 (hGHR) and P16871 (IL-7Rα) with the amino acid numbering ignoring the signal peptide sequence. In all cases the transmembrane domains were modelled as ideal α-helices using PyMOL [64]. The IL-7Rα dimer was created by duplicating the monomer, and rotating and translating the second transmembrane domain into proximity with the first transmembrane domain so that a disulfide bond could be defined between the N-terminal cysteines. The Put3-EPOR chimeras were built by extending the Put3 dimer taken from the crystal structure (PDB entry 1ZME [135]) with the mouse EpoR transmembrane domain sequence as an ideal α-helix. Similarly, the NMR structure of the c-Jun dimer from the (PDB entry 1JUN [136]) was extended with the human GHR transmembrane domain sequence as an ideal α-helix. The initial configurations of these constructrs are illustrated in Figure 7.1. The amino acid sequences of the peptides and constructs used are presented in Figure 7.2. All systems were simulated with four copies of the dimer embedded in a raft-like bilayer containing 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphocholine (POPC), PSM and cholesterol (Table 7.1). A second system of the IL-7Rα dimer was simulated in a pure POPC bilayer (Table 7.1).

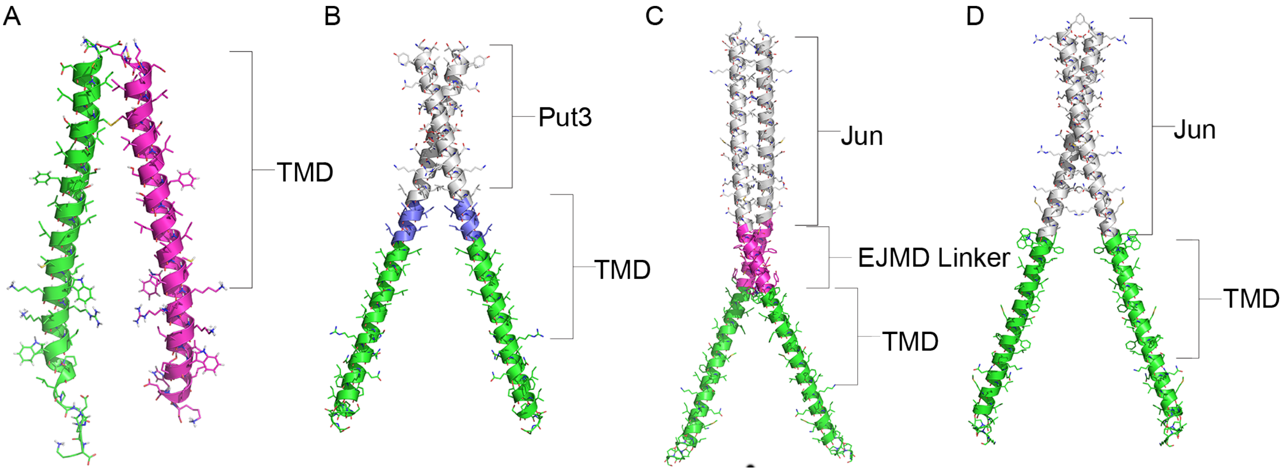


Figure 7.1: Illustration of the dimer systems simulated. (A) The **il-7ra!** P2 clinical mutant linked through a disulfide bridge. (B) The Put3-mEpoR dimer. (C) and (D) the Jun-GHR dimers with and without the extracellular juxtamembrane domain linker.

Table 7.1: Overview of the type-I cytokine receptor transmembrane domains systems simulated.

System	Membrane Composition (ratio)[a]	Simulation time (ns)	Activity[b]	Reference
IL-7RA_P2-CM$_{popc}$	POPC	300	n/d[c]	
IL-7RA_P2-CM	POPC:PSM:Chol (1:1:1)	260	+	[134]
Put3-EpoR-0	POPC:PSM:Chol (1:1:1)	280	+/-	[21]
Put3-EpoR-1	POPC:PSM:Chol (1:1:1)	270	-	[21]
Put3-EpoR-2	POPC:PSM:Chol (1:1:1)	260	-	[21]
Put3-EpoR-3	POPC:PSM:Chol (1:1:1)	260	+	[21]
Put3-EpoR-4	POPC:PSM:Chol (1:1:1)	270	-	[21]
Put3-EpoR-5	POPC:PSM:Chol (1:1:1)	260	-	[21]
Put3-EpoR-6	POPC:PSM:Chol (1:1:1)	260	+	[21]
Jun-L-GHR	POPC:PSM:Chol (1:1:1)	340	+/-	[18]
Jun-GHR	POPC:PSM:Chol (1:1:1)	400	+	[18]

[a] POPC, 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphocholine; PSM, *N*-palmitoyl-D-*erythro*-sphingomyelin; Chol, cholesterol

[b] The activity of the construct is relative to the respective constructs of a receptor and not between the different receptors.

[c] Not determined.

7.2.2 Simulation parameters

Simulations of the Put3-EpoR dimers were performed using the GROMACS simulation package 4.6.5 [50], while the simulations of the IL-7Rα and Jun-GHR dimers were performed using the GROMACS simulation package 5.0.4 [137]. All simulations were run in conjunction with the GROMOS 54A7 united-atom force field [48]. Each system was simulated under periodic boundary conditions in a rectangular box. The pressure was maintained at 1 bar by weakly coupling the system to a semi-isotropic pressure bath [53] using an isothermal compressibility of 4.6×10^{-5} bar^{-1} and a coupling constant $\tau_P = 1$ ps in the lateral (XY) and normal (Z) directions of the bilayer. The temperature of the system was maintained at 298 K by independently coupling the protein, lipids and water to an external temperature bath with a coupling constant $\tau_T = 0.1$ ps using a Berendsen thermostat [53].

```
A
 IL-7Rα P2-CM GEMDPILLTCPTISILSFFSVALLVILACVLWKKRIKPIVWPSLPDHK

B
 Put3-EpoR-0  STKYLQQLQKDLNDKTEENNRLKALLLLILTLSLILVLISLLLTVLALLSHRRTLQ
 Put3-EpoR-1  STKYLQQLQKDLNDKTEENNRLKALLL-ILTLSLILVLISLLLTVLALLSHRRTLQ
 Put3-EpoR-2  STKYLQQLQKDLNDKTEENNRLKALLL--LTLSLILVLISLLLTVLALLSHRRTLQ
 Put3-EpoR-3  STKYLQQLQKDLNDKTEENNRLKALLL---TLSLILVLISLLLTVLALLSHRRTLQ
 Put3-EpoR-4  STKYLQQLQKDLNDKTEENNRLKALLL----LSLILVLISLLLTVLALLSHRRTLQ
 Put3-EpoR-5  STKYLQQLQKDLNDKTEENNRLKALLL-----SLILVLISLLLTVLALLSHRRTLQ
 Put3-EpoR-6  STKYLQQLQKDLNDKTEENNRLKALLL------LILVLISLLLTVLALLSHRRTLQ

 Put3-EpoR-0  QKIWPGIPSPES
 Put3-EpoR-1  QKIWPGIPSPES
 Put3-EpoR-2  QKIWPGIPSPES
 Put3-EpoR-3  QKIWPGIPSPES
 Put3-EpoR-4  QKIWPGIPSPES
 Put3-EpoR-5  QKIWPGIPSPES
 Put3-EpoR-6  QKIWPGIPSPES

C
Jun-Linker-hGHR IARLEEKVKTLKAQNSELASTANMLREQVAQLKQKVMNAALPQMSQFTCEEDFYFP
       Jun-hGHR IARLEEKVKTLKAQNSELASTANMLREQVAQLKQKVMNAA------------FYFP

Jun-Linker-hGHR WLLIIIFGIFGLTVMLFVFLFSKQQRIKMLILPPVPVPK
       Jun-hGHR WLLIIIFGIFGLTVMLFVFLFSKQQRIKMLILPPVPVPK
```

FIGURE 7.2: Amino acid sequences of the constructs used to study the transmembrane domain dimer in a raft-like membrane. The human IL-7Rα P2 clinical mutant (A) was modelled as a dimer through a cysteine-linked disulfide bridge. The mouse EpoR (B) and human GHR (A) transmembrane domains were dimerised using the Put3 or Jun transcription factors respectively. The transcription factor and transmembrane domain sequences are highlighted in grey and blue respectively. The insertion mutation in the IL-7Rα sequence is underlined.

All bond lengths within the protein and lipids were constrained using the LINCS algorithm [66]. Water was included explicitly in the simulations using the Simple-Point Charge (SPC) model [67] and constrained using the SETTLE algorithm [68]. Explicit polar hydrogen atoms in the protein and lipids were replaced by virtual interaction sites, the positions of which were constructed at each step from the coordinates of the heavy atoms to which they are attached [71]. This allowed a 4-fs time step to be used without affecting the thermodynamic properties of the system significantly. Nonbonded interactions within the cutoff of 1.4 nm were calculated every step and the pair list was updated every five steps. A reaction-field correction was applied to the electrostatic interactions beyond the cutoff of 1.4 nm [69], using a relative dielectric permittivity constant of $\epsilon_{\text{RF}} = 62$ as appropriate for SPC water [70]. All systems were energy-minimised after embedding the peptides into the membrane. The temperature of each system was then gradually increased from 50 K to 298 K in 50-K steps over 120 ps to further relax the system and obtain the starting configurations used in the simulations.

7.2.3 Analysis

7.2.3.1 Residue helicity

The GROMACS program g_helix [50] was used to examine the percentage of time each residue was in a helix over the final 50 ns of simulation using the criteria of Hirst and Brooks [84]. A residue was deemed to be in a helix if:

$$\sqrt{(\phi - \phi_c)^2 + (\psi - \psi_c)^2} < 8 \tag{7.1}$$

where ϕ_c and ψ_c are any of the 12 pairs of (ϕ, ψ) helix dihedral angles examined by Manning and Woody [85].

7.2.3.2 Helix length

Again, the GROMACS program g_helix [50] was used to obtain the total length of the helix by fitting the longest helical part to an ideal α-helix around the Z-axis. The total length was then calculated over the final 50 ns of simulation from the average distance between the Cα atoms times the total number of residues in the helix.

7.2.3.3 Helix vector angle

Like in Chapter 5 the overall tilt of the helix was calculated between the normal of the bilayer (Z-axis) and the vector formed from the centroids of the Cα atoms in the first and last turns in the transmembrane region of the helix. A schematic is provided in Figure 5.2A.

7.2.3.4 Calculation of contacts between the transmembrane domains

The GROMACS program g_mindist [50] was used to calculate the number of contacts between the transmembrane domains in a dimer. An atom in the first transmembrane domain of a dimer was deemed to be in contact with an atom in the second transmembrane domain if the distance between any of the atoms was less than 0.4 nm. Atoms were only counted as being in contact with the second transmembrane domain once.

7.2.3.5 Relative orientation of the transmembrane domains

The relative orientation of the transmembrane domains within the mEpoR and hGHR transmembrane domains dimers were calculated from the positions of the leucine residues in the C-terminal region of the transmembrane domain. For mEpoR these were residues L245, L246, L247, L250, L252 and L253, and for the hGHR these were residues L264 and L286. In both cases the angle was calculated from the average vector generated from the vectors from the Cα to the Cβ atoms for each residue in the respective transmembrane domain. The vectors were defined such that an angle of 180° would indicate that the leucine residues in the two transmembrane domains were orientated towards each other in the dimer.

7.2.3.6 Secondary structure

The GROMACS program do_dssp [50] in conjunction with DSSP version 2.0.4 [138, 139] was used to assign secondary structure to the receptor protein sequence from the intrabackbone hydrogen-bonds.

7.3 Results

7.3.1 IL-7Rα P2 clinical mutant transmembrane domain dimer

The IL-7Rα P2 clinical mutant transmembrane domain (Figure 7.2A) was built as an ideal α-helix and dimerised by linking the N-terminal cysteine residues with a disulfide bond. This provided a relatively simple, yet novel dimer of an activated type-I cytokine receptor transmembrane domain. This was used to investigate whether the effects of the membrane composition on the structure and behaviour of the transmembrane domains observed in Chapter 5 also occur with a dimer construct. To this end, the same properties examined for the transmembrane domains in Chapter 5 were initially calculated again for the IL-7Rα P2 clinical mutant. The overall trace of the average helicity was similar between the two bilayer compositions (Figures 7.3A and B). However, a noticeable decrease in the helicity was observed between residues 237-241 (numbering based off the first residue after the signal peptide in the UniProt entry being residue 1) in the raft-like membrane (Figure 7.3B). This did not substantially affect the length or angle of the transmembrane domain with both systems averaging around 4 nm and 165° respectively (Figures 7.3C-F). The two systems showed differences between the contacts in the two halves of the transmembrane domains. The dimers in the pure POPC bilayer formed a number of contacts in both halves of the transmembrane domain ranging between 0-30 and 0-10 in the top and bottom halves respectively (Figures 7.3G and I). The number of contacts in the dimers in the raft-like bilayer ranged between 10-30 and 0-1 for the top and bottom halves (Figures 7.3H and J). Embedding the transmembrane domain dimer into a raft-like bilayer resulted in a slight increase in the average distance between the C-termini from between 1.5-2.0 nm, in the POPC bilayer, to 1.7-2.2 nm (Figures 7.3K and L). Visualisation of the average dimer structures for the transmembrane domains in the pure

POPC and raft-like bilayers showed only the S249 residues are aligned in the POPC bilayer (Figure 7.4A) whereas both serine residues, S246 and S249, orientated towards the corresponding residue in the raft-like bilayer (Figure 7.4B). Overall the dimer in the raft-like bilayer is more clamped in the top half and more splayed in the bottom half.

7.3.2 Put3-EpoR chimeras

The Put3-EpoR chimeras of Seubert *et al.* [21] range in activity from inactive to constitutively active (Table 7.1). The constructs differed by the length of the transmembrane domain, specifically the transmembrane domain was shortened at the N-terminus by progressively deleting residues. These chimeras were modelled by extending the crystal structure of the Put3 transcription factor (PDB entry 1ZME [135]) with the sequences shown in Figure 7.2B. Figure 7.5A illustrates the key features of these constructs highlighting the Put3 and mEpoR sequences as well as the transmembrane domain and the residues that were sequentially deleted. As can be seen in Figure 7.5B the conformation of the dimers were restricted by the Put3 dimer. The overall helicity of these constructs was well maintained in all simulations with only a slight loss of helicity at the termini (Figures 7.5C-I).

The dimerisation of these constructs was maintained through the Put3 transcription factor. In fact very few contacts were observed between the residues in the mEpoR transmembrane domains (Figure 7.6). The Put3-EpoR-2 chimera displayed the largest number of contacts typically ranging between 5-10 per dimer (Figure 7.6G). These contacts in the Put3-EpoR-2 chimera were almost entirely from interactions of the N-terminal leucine and threonine residues. The C-termini in the Put3-EpoR chimeras never made contacts with each other (Figure 7.5B). Typically the C-termini were separated by around 2.5 nm with the Put3-EpoR-(0,1) chimeras having a slightly larger separation of over 2.8 nm (Figures 7.6B, E, H, K, N, Q and T). The relative orientation of the transmembrane domain adopted by the chimeras was bimodal averaging around 40° for Put3-EpoR-(0,1,4,5) and 80° for Put3-EpoR-(2,3,6) (Figures 7.6C, F, I, L, O, R and U).

7.3.3 Jun-GHR chimeras

Two Jun and GHR transmembrane domain chimeras were investigated that were linked either via 12 amino acids from the extracellular juxtamembrane sequence or directly to the Jun sequence, referred to as Jun-L-GHR and Jun-GHR respectively (Figure 7.7A). Both of these constructs are constitutively active, with the Jun-GHR being the more potent of the two [18]. Overall, both of these chimeras retained helicity throughout the Jun and hGHR transmembrane sequences, however some of the residues in the extracellular juxtamembrane domain linker lost helicity (Figures 7.7B and C). The C-termini of the transmembrane domains were separated by 1.5 nm and 4.5 nm in the Jun-L-GHR and Jun-GHR dimers respectively (Figures 7.7D and E). The overall tilt of the helix through the membrane was similar for the two chimeras at 155°, although the Jun-L-GHR did display a greater variation (Figures 7.7F and G). The relative orientation of the transmembrane domains calculated from the leucine face was 35° and 115° for the Jun-L-GHR and Jun-GHR dimers respectively (Figures 7.7H and I).

As the GHR sequences were observed to interact in the Jun-L-GHR construct these dimers were

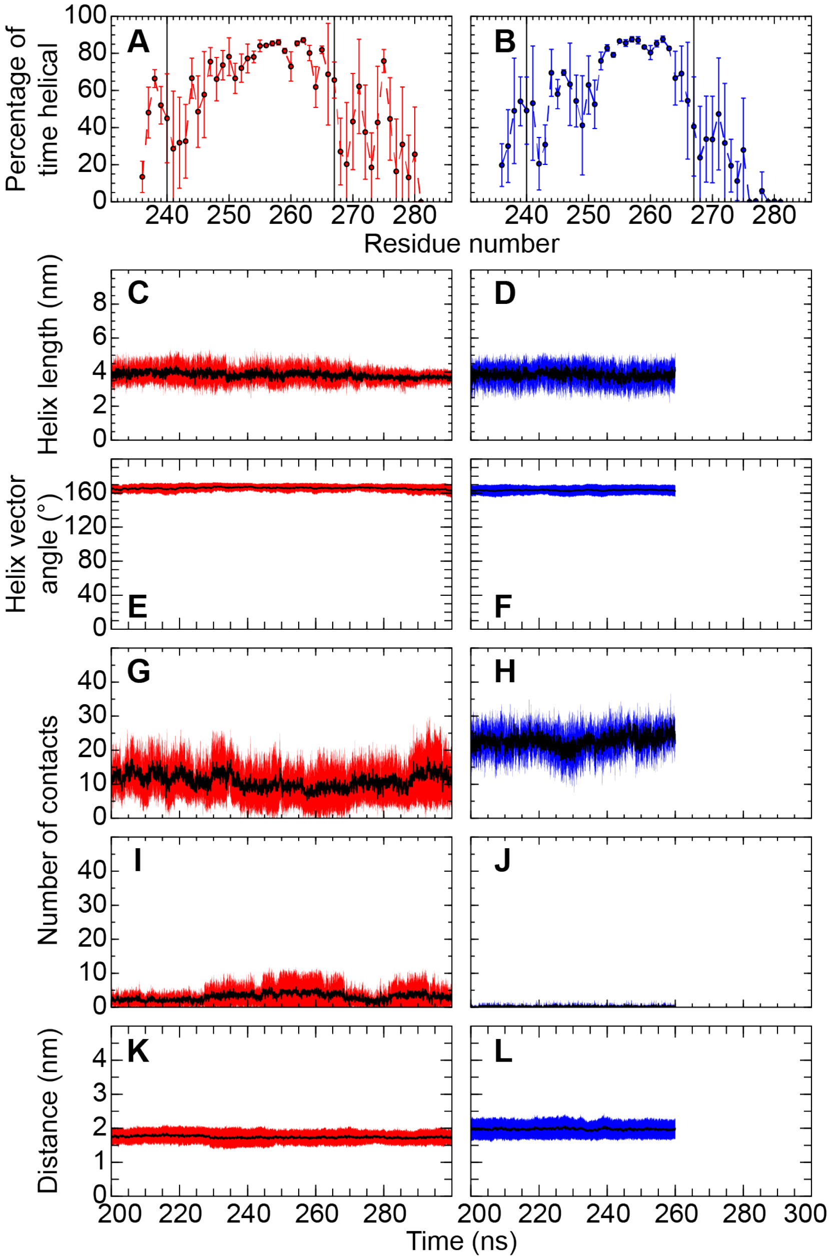


Figure 7.3: The results for the IL-7Rα P2 clinical mutant dimer in a pure POPC (red) or raft-like bilayer (blue). (A-B) The percentage of time the residues within the eight chains from the four dimers were calculated as being helical. (C-D) The length of the longest continual helix. (E-F) The overall tilt angle of the helices in the dimers. (G-H) and (I-J) are the number of contacts between the dimers for the top and bottom halves of the transmembrane domain respectively. (K-L) The distance between the C-terminal end of the transmembrane domains in the dimers. Values are the average (black line) ± the standard deviation (red or blue shading) over the eight chains (A-F) or for the four dimers (G-L) in each system.

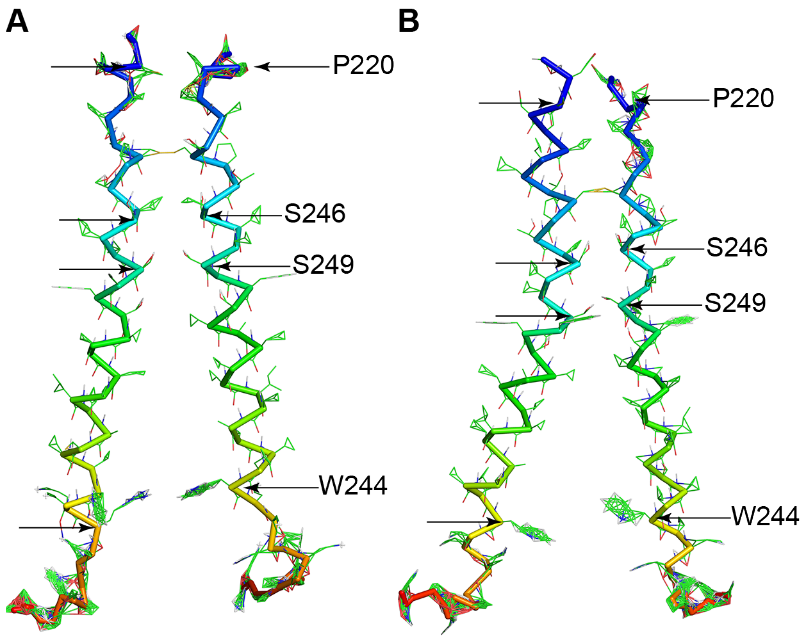


FIGURE 7.4: The average structures of the IL-7Rα P2 clinical mutant transmembrane domain dimers in (A) a pure POPC bilayer and (B) a raft-like bilayer. Peptide sequence has been coloured blue to red.

examined further in terms of their relative orientation and distances between the residues. As can been seen in Figure 7.8A, the GHR sequence formed contacts between the two dimers in the extra-cellular juxtamembrane linker sequence through to the residues in the GxxG motif. The dimers were arranged such that only one GxxG motif (red trace in 7.8A) was interacting with the second chain (blue trace). This orientation of the dimers resulted in corresponding residues pairs (e. g. L250 from each transmembrane domain), suggested to be orientated towards each other in the inactive dimer from cysteine-mutagenesis cross-linking studies, to be rotated by 180° (Figure 7.8B). This resulted in residues that displayed minimal cross-linking (e. g. I252) having a smaller average distance between the chains than residues that displayed strong cross-linking (I254) (Figure 7.8C). These conformations also prevented the cysteine residues C241 in the linker region, known to form a disulfide bond during receptor activation, from coming within 0.4 nm in three of the four dimers (Figure 7.8C). Despite loosing the initially modelled helicity, the linker region was observed to partially fold during the simulations. Indeed, the algorithm DSSP [138] calculated two α-helices in this region that were spaced by residues that could form a bend (Figure 7.8D). Furthermore, the algorithm suggests there may be a further bend or coil between the extracellular juxtamembrane domain and beginning of the transmembrane domain.

7.4 Discussion

Undoubtedly the binding of a ligand to the extracellular region of a cell-surface receptor is associated with the mechanical transmission of the signal via a transmembrane region to the intracellular region of the cell. In the case of the G-protein coupled receptors (GPCRs) it is suggested that the binding of an agonist would stabilise the transmembrane regions of the receptor in an active conformation [140].

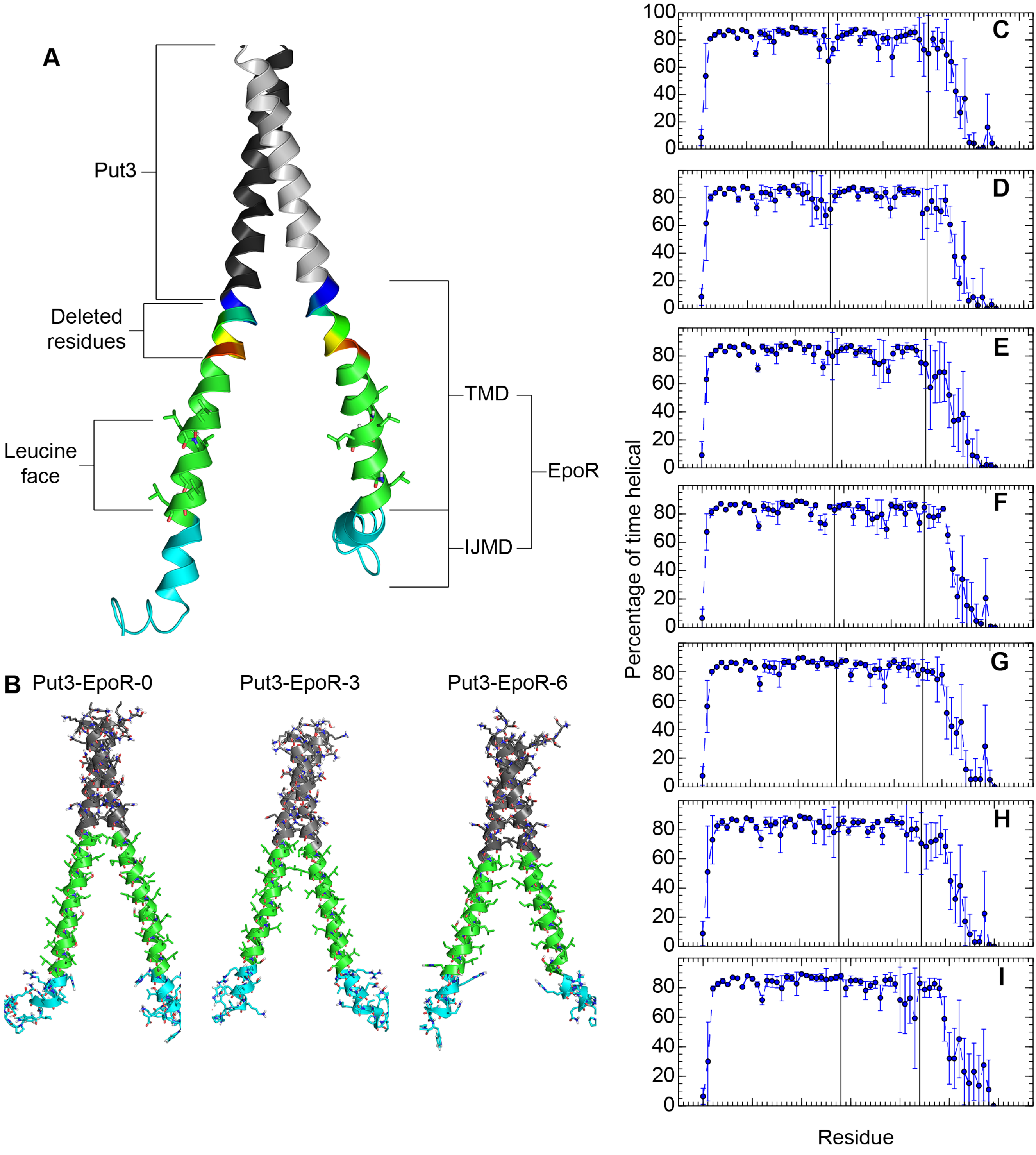


FIGURE 7.5: The Put3 and mEpoR chimera. (A) The Put3 leucine-zipper was extended as an α-helix with the mouse EpoR transmembrane domain sequence as shown in Figure 7.2B. Indicated on the figure are start and end of the transmembrane domain (TMD) and intracellular juxtamembrane domain (IJMD), the residues that were progressively deleted, and the C-terminal leucine residues of the transmembrane domain. (B) The final configuration of a dimer construct from the Put3-EpoR-0, Put3-EpoR-3 and Put3-EpoR-6. (C-I) The percentage of time the residues within the Put3-EpoR-(0-6) chimeric constructs were calculated as being helical. The black lines indicate the start and end of the EpoR transmembrane domain sequence. Values are the average (black line) ± the standard deviation over the eight chains in the four dimers in each system.

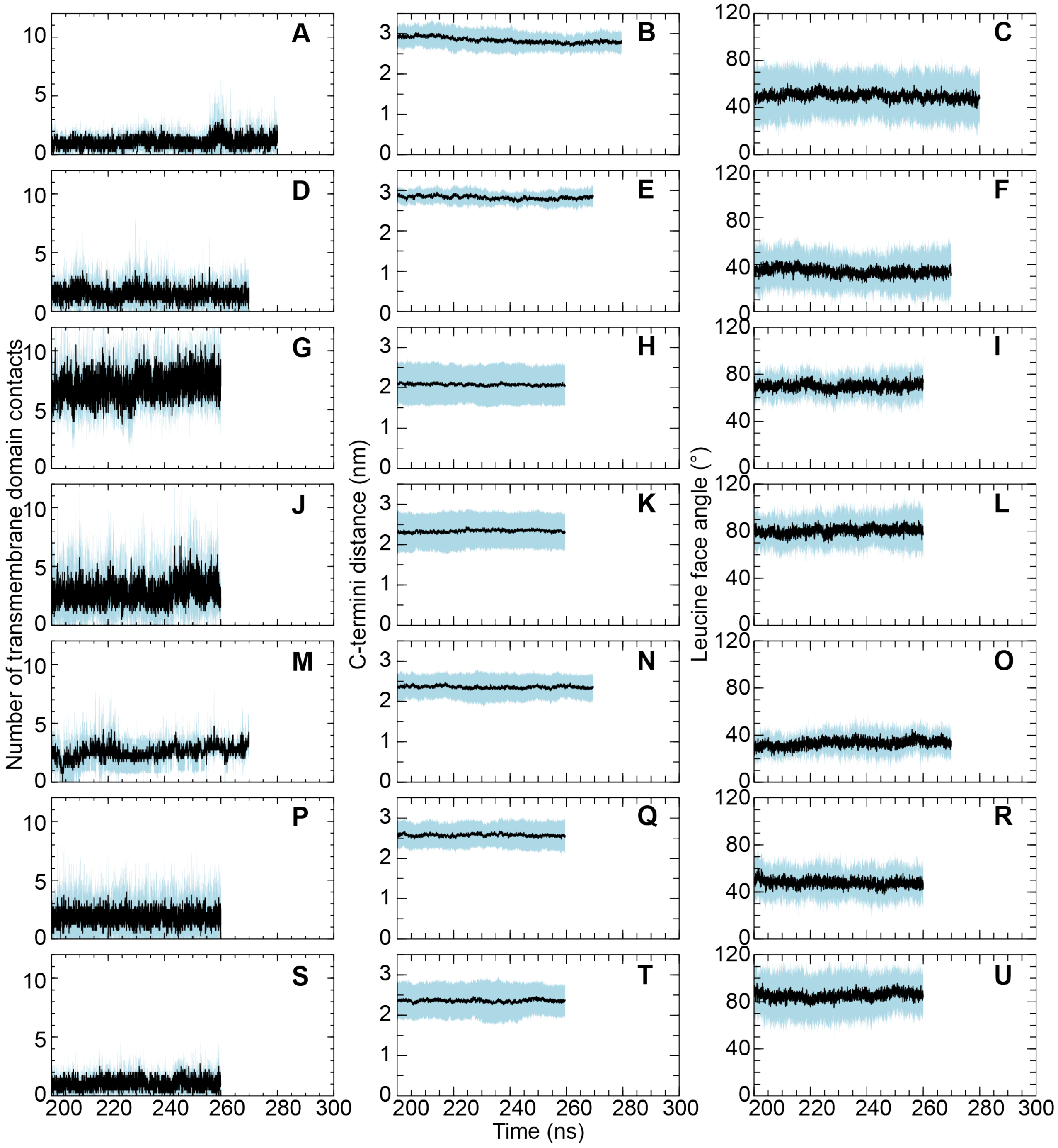


FIGURE 7.6: The number of contacts between the first two turns of the EpoR transmembrane domain, the distance between the C-terminus of the transmembrane domain and the angle between the C-terminal leucine face on the transmembrane domain for: Put3-EpoR-0 (A-C); Put3-EpoR-1 (D-F); Put3-EpoR-2 (G-I); Put3-EpoR-3 (J-L); Put3-EpoR-4 (M-O); Put3-EpoR-5 (P-R); Put3-EpoR-6 (S-U). Values are the average (black line) ± the standard deviation (blue shading) over the the four dimers in each system.

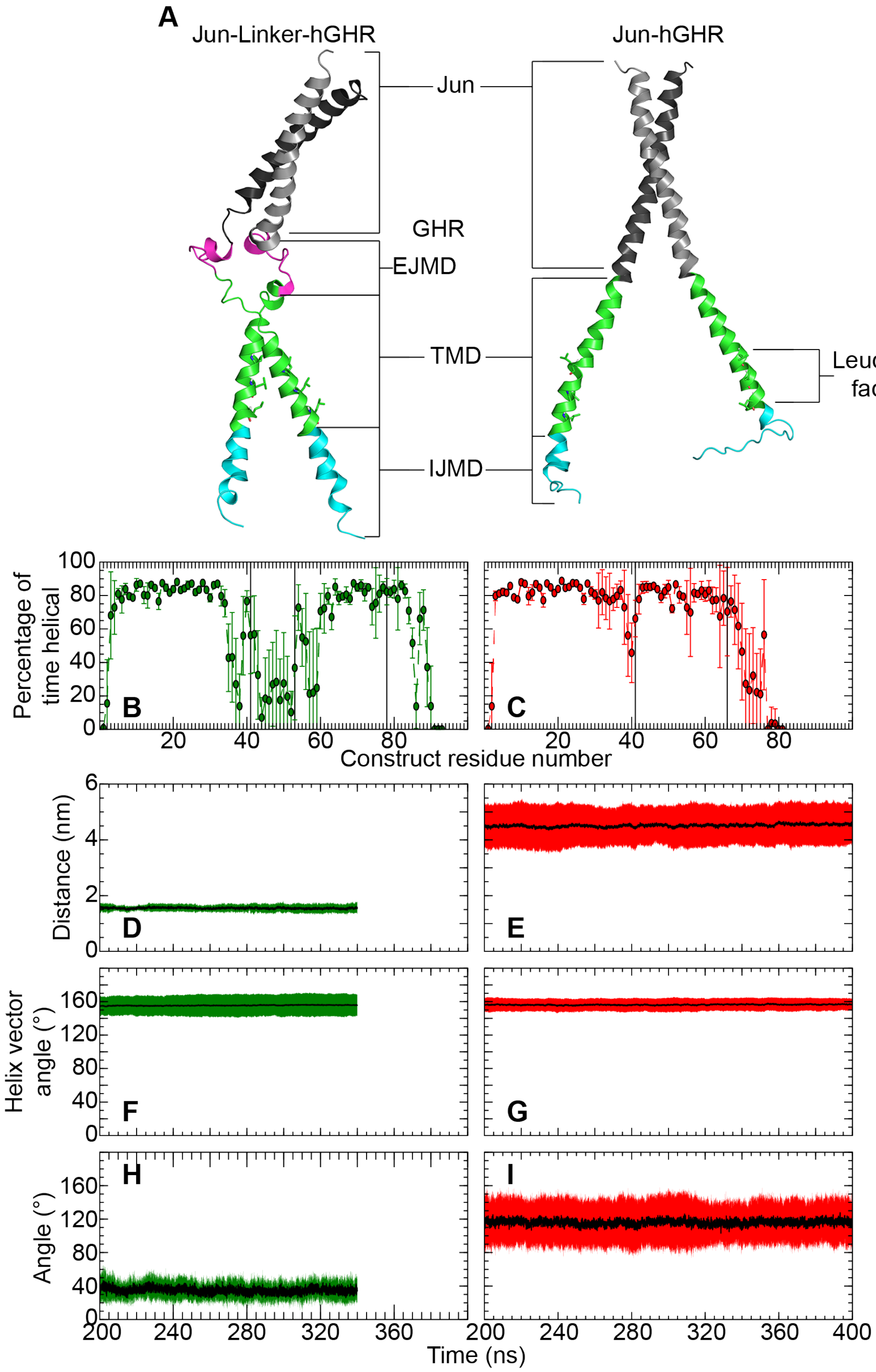


FIGURE 7.7: (A) The Jun leucine-zipper was extended as an α-helix with either the human GHR extracellular juxtamembrane domain as a linker with the transmembrane domain (Jun-Linker-hGHR) or only the transmembrane domain (Jun-hGHR). The peptide sequence used are shown in Figure 7.2C. Indicated on the figure are the Jun dimer (Jun) (grey), extracellular juxtamembrane domain (EJMD) (magenta), the start and end of the transmembrane domain (TMD) (green), intracellular juxtamembrane domain (IJMD) (cyan), and the C-terminal leucine residues of the transmembrane domain. (B-C) The percentage of time the residues within the Jun-L-GHR and Jun-GHR chimeric constructs were calculated as being helical. The black lines indicate the start and end of the mEpoR extracellular juxtamembrane and transmembrane domain sequences. Values are the average ± the standard deviation over the eight chains in the four dimers in each system. (D-E) The distance between the C-terminal residues in the transmembrane domain. (F-G) The overall tilt of the transmembrane domain. (H-I) The relative orientation of the transmembrane domain calculated from the leucine face. Values are the average (black line) ± the standard deviation (green or red shading) over the eight chains in the four dimers in each system.

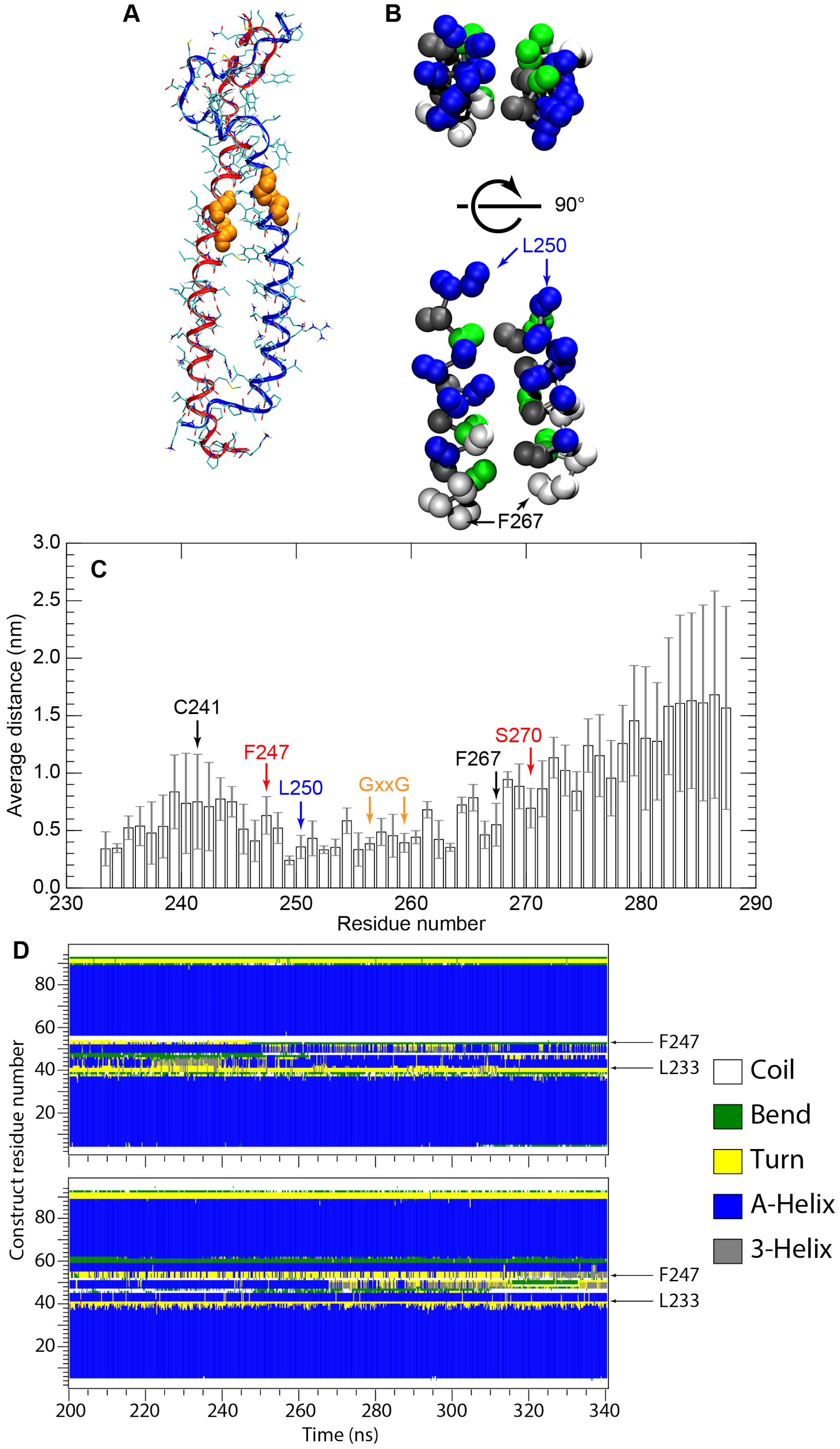


Figure 7.8: Representative structure of the hGHR sequence from the Jun-L-GHR chimera showing (A) the interactions occurring in the N-terminal region and (B) the relative orientation of the residues in the transmembrane domain. In (A) the orange spheres are the glycine residues of the GxxG dimerisation motif. In (B) only the Cα and Cβ atoms from residues L250-F267 are shown. Residues have been coloured according to the level of cross-linking reported from cysteine-scanning mutagenesis; blue-strong, green-weak, and grey-minimal [18]. (C) The average distance of each of the residues with the second chain in the dimer. (D) A representative secondary structure plot of the two protein chains within a single dimer.

However, unlike the GPCRs that have seven transmembrane helices, the type-I cytokine receptors only have one single-pass transmembrane domain per chain (Figure 1.9). Furthermore, the type-I cytokine receptors require two receptor chains in order to signal. There is also the suggestion that, at least for the EPOR, growth hormone receptor (GHR) and IL-7Rα, these receptors can exist as preformed homo- (GHR, EPOR) or heterodimers (IL-7Rα with the γ-common chain) on the cell surface [27, 17, 28]. It would appear therefore, that binding of an agonistic ligand would not only stabilise an active conformation, but also drive a change in the relative orientation of the two receptor chains within the dimer. Indeed, studies that investigated mutations in the EPOR and GHR transmembrane and juxtamembrane domains suggest the activation of the receptor can be modulated by changing the relative orientation of the receptor chains [87, 21, 17, 18]. The suggestion that the membrane environment is essential in the activation of the receptors adds further complexity. Not only have the IL-7Rα and EPOR been observed to colocalise with rafts upon stimulation with the respective cytokines, but removal of cholesterol or sphingomyelins from the membrane can prevent the signalling of the IL-7Rα, EPOR and GHR [81, 38, 141]. Using this information we built upon our previous work to investigate several dimers of the type-I cytokine receptor transmembrane domains that had a restricted conformational space in order to shed light on potential active and inactive conformations of the transmembrane domain.

The IL-7Rα P2 clinical mutant was the simplest transmembrane dimer that we investigated. The transmembrane domain was built as an ideal α-helix and the dimer restricted by the presence of a disulfide bond at the N-terminus (the end proximal to the extracellular domain). In the case of the IL-7Rα dimer there was noticeably different behaviour in the upper-half of the transmembrane domain depending on the membrane composition. In particular, more contacts were maintained, which resulted in two transmembrane domains being in closer contact within this region with the serine residues (S246 and S249) aligned between the transmembrane domains in the raft-like bilayer. This gave the IL-7Rα dimer in the raft-like bilayer a splayed appearance at the C-termini. However, this did not translate into a significantly larger distance between the two termini of the transmembrane domains, with an average difference of 0.5 nm. It is not possible to conclude whether the conformation observed corresponded to the activated state of the transmembrane domain as the extracellular and intracellular domains were not included in the simulations. Further work may be required to examine the effect of removing the disulfide bond in the dimer embedded in the pure POPC, as in both cases the system simulated may represent an `active' dimer. They do indicate however, that raft-like bilayer does affect the conformation of the transmembrane domain dimer.

The lack of structural data on the transmembrane dimers and the extracellular juxtamembrane domain sequence means that any modelling of the coupling between the extracellular and transmembrane domains would need to be extensively validated. We therefore investigated the Put3-mEpoR and Jun-hGHR chimeras, for which experimental data is available, for information on how the transmembrane domains could potentially be coupled to the extracellular domains for these receptors. A noticeable feature of these constructs was that the transcription factor clearly had a large influence in the conformation of the receptor transmembrane domains. For all the Put3-EpoR chimeras and the Jun-GHR chimera (lacking the linker), the constructs were a continuous helix stretching several residues into the intracelluar juxtamembrane domain. The conformations that were generated were very stable in the simulations with the distance between the splayed C-termini, and relative orientation of the transmembrane domains remaining constant despite relatively few contacts between the receptor sequences. In contrast the Jun-L-GHR construct (containing the linker), was partially folded in parts of the linker region. The suggestion from the simulations that this region is largely structured draws into question the model of the human GHR presented by Brooks *et al.* [18] where the extracellular domain is linked by an extended loop to a dimer of the transmembrane domains. The secondary structure analysis indicates this linker region may fold as a helix-turn/bend-helix motif that could be stablised by contacts with the extracellular domain. In this case, this could suggest that the GHR chain is well structured from the extracellular domain down through to the intracellular juxtamembrane domain. This model would support a mechanism whereby a change in the orientation in the extracellular region could transfer as a mechanical signal to change the relative orientation of the intracellular domains. From a purely mechanical perspective this would be sensible as it would be easier to transfer a mechanical force through a rigid/structured body rather than through a loop or unstructured region.

It was not possible to directly compare the dimer conformations obtained with the Put3-EpoR chimeras with the previously suggested dimer interfaces generated from MD simulations performed *in vacuo* [21] given the lack of contacts between the mEpoR sequences. However, in both the Put3 and Jun chimeras the relative orientation of the transmembrane domains could be influenced by the removal of residues in the N-terminal or extracellular juxtamembrane regions. Interestingly, in the case of the Put3-EpoR chimeras, not every deletion resulted in a change in the relative orientation of the transmembrane domains which has been assumed in the interpretation of the experimental data using these constructs. In the case of the Jun-hGHR chimeras the structure of the constructs followed the trend deduced experimentally from FRET analysis where the more active construct Jun-GHR was suggested to have a greater distance between the C-termini than the Jun-L-GHR construct [18]. In addition the contacts between the GHR chains in the Jun-L-GHR chimera did not correspond with what would be expected for the inactive dimer nor the active dimer from cysteine mutagenesis cross-linking studies, suggesting an intermediary alignment of the transmembrane domain which may reflect its partial activity seen experimentally [18]. It is possible that the results reflect the choice of starting structures. In order to identify potential artefacts due to the structure of the transcription factors, simulations in which the transcription factor chains are aligned in parallel, as in the PDB entry 1AJY, will need to be performed in the future [142].

In concert with the simulations of the transmembrane dimers, several attempts were made at connecting the extracellular and transmembrane domains for the EPOR and GHR (Figure 7.9). These include relatively simple constructs consisting of a single D2 sub-domain connected to the transmembrane domain through the extracellular juxtamembrane region modelled as an α-helix or a loop (Figure 7.9A), as well as larger complexes of the two receptor chains bound to a cytokine (Figure 7.9B and C). Not only were these systems computationally expensive to run, the setup and interpretation of the structures obtained were difficult. For example, given that the X-ray crystal structures observed are unlikely to be the exact conformation observed on the cell surface, it must be questioned whether it is appropriate to use the dimer as seen in the crystal structures. Furthermore, directly building the

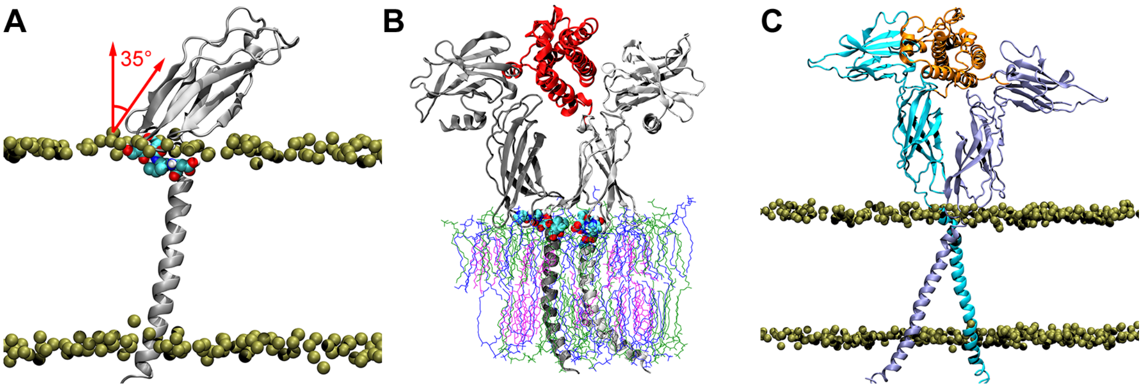


Figure 7.9: The extracellular domain of the EPOR was modelled linked to the transmembrane domain via a loop (A) or an α-helix (B). The extracellular domain residues 221-225 are shown as spheres. (C) The extracellular domain of the GHR was modelled linked to transmembrane domain via an α-helix. In (A) the angle was calculated between the plane defining the D2 domain and the normal to the membrane (Z-axis). In (A) and (C) the phosphorus atom of the lipid headgroup is represented as gold sphere. In (B) the lipids are coloured blue-POPC, green-PSM and magenta-cholesterol.

transmembrane domains as a continuation of a given crystal structure led to nonphysical structures within the transmembrane domains. This raises the question of whether the orientation of the transmembrane domain should be adjusted to match the extracellular domain or *vice versa*. This problem is exemplified by the heavy tilt of the D2 domain of 35° from normal of the bilayer observed in Figure 7.9A . Given the limitation in resources and time, further production and investigation of these systems was halted. The setup of these extracellular-transmembrane domain systems is critical given the starting conditions heavily influence the outcomes of the simulations. How these models can be reconciled with experimental and structural data, and whether they are physiologically relevant, will need to be considered in any future work. It has become apparent from these simulations that the interactions between the D2 sub-domain and the membrane may be crucial. Indeed, the simulations ruled out consistently that these receptors float proud above the membrane. Instead the simulations suggest the extracellular domain sits on, if not partially in, the plasma membrane.

7.5 Conclusion

This work aimed to understand how the transmembrane domain in a dimeric conformation could influence the activation of the receptor. In part this work supports the model in which the transmembrane domains could move apart during the activation of the GHR and that the orientation of the dimer is also important. Alternatively, the simulations of the chimeric constructs could support a mechanism of activation of the receptors that would consist in widely separating the transmembrane domains. Given the similarities in the structures of the Put3-EpoR chimeras as well as those between the IL-7Rα dimers it is not possible to assign particular conformations to active or inactive structures. Future work will address the orientation of the intracellular region that may need to be extended past the Box 1 motif to improve the resolution of how this region behaves. Overall, these simulations indicate that care needs to be taken when interpreting different models of a system and that mechanisms need to be drawn from known structural data, for example the chimeric constructs were not observed to have a parallel arrangement of the transmembrane domains as quite often represented in the literature. Finally, the chimeric constructs questions the utility in interpreting the fine details in studies that examine only a part of a system, for example a single transmembrane domain, which neglect contributions to the structure from the extra and intracellular regions.

Chapter 8

Conclusions and perspectives

The ability of a cell to sense its environment is a hallmark of life. Indeed, even viruses that are in general considered non-living biological entities, still require the ability to recognise their target host cell through surface receptor molecules. How signals are transferred through cell membranes is a fundamental question in biology, and a problem that has been solved by evolution in numerous ways including two-component regulatory systems, voltage-gated ion channels, receptor tyrosine kinases, G-protein-coupled receptors and cytokine receptors. The cytokine receptors are a medically important group of receptors. In particular, the type-I cytokine receptor family regulate a diverse range of functions including inflammation and immune responses, growth and development, and tissue maintenance. With continued research on these receptors, exciting new roles in health and disease are being discovered. For example, the EPOR was initially only reported in the bone marrow but in recent years has been found in neuronal tissue and suggested to have a role in protecting the brain during ischaemia and reperfusion [72]. These new roles greatly expand the potential of the EPOR as a drug target, for example in the treatment of stroke injury. However, understanding how receptor ligands and drugs modulate the activity of a receptor is critical as agonism for one disease (e. g. anaemia) would be detrimental in others (e. g. blood cancers). This thesis was dedicated to improving our understanding of how the type-I cytokine receptors transfer signals across the plasma membrane. In particular, how the event of a ligand binding to the extracellular domain of these receptors is mechanically coupled to the activation of the intracellular region.

The type-I cytokine receptors are a widely studied family of receptors. This is reflected in the numerous X-ray crystal structures of the extracellular domains both with and without ligands that are publicly available. When they were first solved, these structures provided novel insights into how a single cytokine activated two receptors by binding in between the two receptor chains [33]. The conformations of the receptor complex within different crystals have widely been assumed to be physiologically relevant. In the case of the EPOR, differences between alternative structures were used to propose a detailed mechanism of action, essentially linearly interpolating between the structures to infer dynamic behaviour. In Chapter 2 we reviewed the available experimental data in the literature for the EPOR and provided a reinterpretation of the crystal structures used to propose the cross-action scissor-like mechanism of activation. We found that this mechanism of action is not supported with currently available literature and the motions assumed by the model were incompatible with the structures originally used to propose the mechanism. This prompted us to ask in Chapter 3 whether it is appropriate to use the X-ray crystal structures, that only consider a portion of the receptor, as conformations that represent active and inactive conformations. It became apparent from the simulations that the conformations observed crystallographically could not be maintained in solution. This is a significant finding not only for the type-I cytokine receptors, but more widely to the structural biology community. It highlights the importance of also considering the environment in which the structure is observed or used to measure an activity of the protein. In cases where only a single, small, globular, stable protein like BPTI is being considered, conclusions on structural elements may be drawn with confidence from these X-ray crystal structures. However, in crystal structures involving truncated proteins in larger complexes where the influence of the local environment on the conformation of the complexes is unknown, the potential physiological relevance of the conformations diminishes. Critical questions must therefore be asked; such as: does it make physiological sense to have a preformed receptor dimer that occludes the ligand-binding sites lying flat on the surface of the cell [29]? Or, when proposing an alternative mechanism of activation that separates the receptor chains over 13 nm, is it appropriate to interpret a difference of 1--2 nm between an agonist or antagonist-bound crystal structure as physiologically significant [9]? Far from criticising the models of the X-ray crystal structures themselves, or belittling the effort to solve the crystal structures, this work serves to raise awareness of how to maximise the utility of X-ray crystal structures when addressing mechanistic questions. Indeed, in Chapter 4 we investigated how reliable were the models of the X-ray crystal structures containing the EPOR were by replicating the crystal lattice. If anything the simulations demonstrated the models themselves were reasonable despite the unfavorable scores in the PDB validation reports for these structures. Unfortunately the structure factors for these crystal structures are unavailable and so it was not possible to see whether the MD simulations could have aided in the refinement of the models. In the future it would be beneficial to see reports on new X-ray crystal structures with dynamic information produced from MD simulations to advance structural biology from the mindset of interpreting a single model as the physiologically relevant form, to appreciating that these models are one amongst many possible arrangements or conformations.

To validate our claims that it is essential these receptor proteins be considered in a membrane environment, in Chapter 5 we investigated the effect of raft-like bilayers on the transmembrane domains compared to a bulk membrane-like bilayer. These results showed a surprising variation in the behaviour across the family, ranging from a negligible effects in the case of the GHR, to large changes in the tilt of the transmembrane domain in the case of the rPRLR. In the case of the TPOR the differences observed could almost be considered a conformational change. If these changes in the transmembrane domain were to happen in the case of the full-length receptors where the N- and C-termini of the transmembrane domains are more restricted, then not only is the presence of the membrane essential to consider when proposing a model of activation but also the membrane composition. Furthermore, the effect of the raft-like domain in the case of the rPRLR and TPO could suggest a novel point of regulation of these receptors that has not previously been appreciated. While it is known rafts are required in the activation of the receptors, it has not been previously reported that the rafts could influence the structure of these receptors. This is where interdisciplinary studies will be necessary to fully elucidate how these receptors signal. While NMR spectroscopy techniques have provided some insight into the structure of the transmembrane domains in a pure phosphocholine micelle, the structure of the transmembrane domain in the presence of cholesterol is not known. Also lacking is a structure of the extracellular domain coupled to the transmembrane domain. Cryo-electron microscopy (Cryo-EM) is one experimental technique that could solve this problem and provide structural data of the receptors in a membrane environment. We attempted in Chapter 7 to generate conformations of a transmembrane domain dimer for several receptors, the intention eventually being these could be coupled to the extracellular domains. While we tried to restrict the search for active orientations of the mEpoR and GHR by using the chimeric constructs, the orientation of the transmembrane domains in these constructs was not compatible with the extracellular domains from the crystal structures, and could represent an alternative mechanism to activate these receptors. It was also difficult to reconcile experimental data, such as the cross-linking studies with the GHR, of the full-length receptor with the chimeric constructs. While a somewhat unexpected result, it reinforces the point made earlier with the simulations of the extracellular domain in solution, in that full consideration should be given to what the model actually is before extrapolating mechanistic details, especially when structural data is absent. This point illustrates a difficulty in molecular modelling in terms of which data can be trusted and relatively weighing the different pieces of information available that can be used in building a system. Small errors can have a profound impact on the results obtained. Indeed, modelling the attachment of the extracellular and transmembrane domains proved to be difficult and computationally expensive. Several simulations were performed but it was not possible to validate that the conformations obtained were physiologically meaningful. In particular, it was not clear on which domain to orientate, either the transmembrane or extracellular domain of the protein, or how to enforce the dimer, whether through the transmembrane domains or maintain the configuration of the crystal structure. This may be a problem unsuitable for MD simulations and require an initial structure produced experimentally from, for example, Cryo-EM before further simulations of such a large complex are investigated.

Validation remains an important aspect of molecular modelling and to know the reliability and the limits of a model are essential. In Chapter 6 we produced parameters for the molecule PSM and built several bilayers containing this lipid. These bilayers were used to examine several properties of PSM and compare them with experimental values. Although while writing the thesis an error was found (*cis* rather than *trans* of the carbon-carbon double bond in the sphingosine backbone) in the parameters of PSM, given how well the bilayers with the *cis* configuration compared well to experimental values, especially the predicted melting temperature, the correct *trans* configuration of the double bond is not expected to vary significantly from these results and is currently being run. These *cis* bond containing bilayers present us with an opportunity to explore what are the effects of the double-bond configuration. Bilayers are particularly suitable for study with MD simulations as atomic detail can be seen that cannot be obtained experimentally due to the averaging that occurs over time and through space from measurements. Of particular interest in the PSM bilayers is whether there is an extensive H-bond network that corresponds with the packing or ordering of the lipids. Furthermore, what influence does cholesterol have in promoting or disrupting an H-bond network between the PSM molecules? In regard to cholesterol, it would be interesting to pursue whether the receptor transmembrane domains had a cholesterol-binding motif. If so, does it exist in the single transmembrane domain sequence or does receptor-ligand binding promote a dimer conformation of the transmembrane domains that preferentially interacts with cholesterol resulting in the co-localisation of the receptors with rafts. These questions would require the use of complementary computational techniques such as molecular docking and bioinformatic analysis of the transmembrane domain sequences. If such cholesterol-binding sites could be found it would open up new avenues in which to target drugs that could modulate the activity of these receptors.

The success and utility of MD simulations in the future depends on three points: (i) reliability of predictions, (ii) speed of generating the predictions, and (iii) integration of experimental and theoretical data. The reliability of MD simulations depends almost entirely on the force field used. While proteins are on the whole generally well described by current force fields, the description of small heteromolecular molecules that are substrates, co-factors or potential drugs are less robust and are an area of focused research in order to improve these shortcomings [143]. The improvement to the speed of the algorithms running the calculations over the past years has exploded, especially with the development and implementation of code that exploits GPU-acceleration [144, 145, 137]. Indeed, these improvements dramatically reduce the cost in both computing hardware and time needed to achieve microsecond simulations. These savings make MD simulations more accessible to many researchers and will render hardware specifically designed to run MD simulations comparatively expensive and obsolete. The most critical point however is the continual need to successfully integrate experimental and theoretical data not only for MD, but also the advancement of structural biology as a whole. Not only do the theoretical scientists need to be able trust the structures produced by the experimentalists, the experimentalists in turn need to be able to trust the computational predictions. This requires computational techniques being promoted as a tool that can help establish models, rather than as a supplement to purely structural interpretations. It would be a shame to see larger, more detailed protein complexes solved by new experimental techniques again be treated as snapshots of the system with non-physical linear interpolations proposed as dynamical properties. We will able to see further if we can stand on each others shoulders.

Bibliography

[1] M. S. P. Corbett, D. Poger, and A. E. Mark, ``Revisiting the scissor-like mechanism of activation for the erythropoietin receptor," *FEBS Letters*, vol. 590, no. 18, pp. 3083--3088, 2016.

[2] L. J. Witts, ``Some aspects of the pathology of anaemia," *British Medical Journal*, vol. 2, no. 5249, pp. 404--410, 1961.

[3] S. B. Krantz, O. Gallien-Lartigue, and E. Goldwasser, ``The effect of erythropoietin upon heme synthesis by marrow cells *in vitro*," *The Journal of Biological Chemistry*, vol. 238, no. 12, pp. 4085--4090, 1963.

[4] T. Taga, M. Hibi, Y. Hirata, K. Yamasaki, K. Yasukawa, T. Matsuda, T. Hirano, and T. Kishimoto, ``Interleukin-6 triggers the association of its receptor with a possible signal transducer, gp130," *Cell*, vol. 58, no. 3, pp. 573--581, 1989.

[5] O. Livnah, E. A. Stura, D. L. Johnson, S. A. Middleton, L. S. Mulcahy, N. C. Wrighton, W. J. Dower, L. K. Jolliffe, and I. A. Wilson, ``Functional mimicry of a protein hormone by a peptide agonist: The EPO receptor complex at 2.8 Å," *Science*, vol. 273, no. 5274, pp. 464--471, 1996.

[6] O. Livnah, D. L. Johnson, E. A. Stura, F. X. Farrell, F. P. Barbone, Y. You, K. D. Liu, M. A. Goldsmith, W. He, C. D. Krause, S. Pestka, L. K. Jolliffe, and I. A. Wilson, ``An antagonist peptide-EPO receptor complex suggests that receptor dimerization is not sufficient for activation," *Nature Structure Biology*, vol. 5, no. 11, pp. 993--1004, 1998.

[7] R. S. Syed, S. W. Reid, C. Li, J. C. Cheetham, K. H. Aoki, B. Liu, H. Zhan, T. D. Osslund, A. J. Chirino, J. Zhang, J. Finer-Moore, S. Elliott, K. Sitney, B. A. Katz, D. J. Matthews, J. J. Wendoloski, J. Egrie, and R. M. Stroud, ``Efficiency of signalling through cytokine receptors depends critically on receptor orientation," *Nature*, vol. 395, no. 6701, pp. 511--516, 1998.

[8] O. Livnah, E. A. Stura, S. A. Middleton, D. L. Johnson, L. K. Jollife, and I. A. Wilson, ``Crystallographic evidence for preformed dimers of erythropoietin receptor before ligand activation," *Science*, vol. 283, no. 5404, pp. 987--990, 1999.

[9] I. Moraga, G. Wernig, S. Wilmes, V. Gryshkova, C. Richter, W.-J. Hong, R. Sinha, F. Guo, H. Fabionar, T. Wehrman, P. Krutzik, S. Demharter, I. Plo, I. Weissman, P. Minary, R. Majeti, S. Constantinescu, J. Piehler, and K. Garcia, ``Tuning cytokine receptor signaling by re-orienting dimer geometry with surrogate ligands," *Cell*, vol. 160, no. 6, pp. 1196--1208, 2015.

[10] Q. Li, Y. L. Wong, Q. Huang, and C. Kang, ``Structural insights into the transmembrane domain and the juxtamembrane region of the erythropoietin receptor in micelles," *Biophysical Journal*, vol. 107, no. 10, pp. 2325--2336, 2014.

[11] Q. Li, Y. L. Wong, M. Y. Lee, Y. Li, and C. Kang, ``Solution structure of the transmembrane domain of the mouse erythropoietin receptor in detergent micelles," *Scientific Reports*, vol. 5, p. 13586, 2015.

[12] K. Bugge, E. Papaleo, G. W. Haxholm, J. T. S. Hopper, C. V. Robinson, J. G. Olsen, K. Lindorff-Larsen, and B. B. Kragelund, ``A combined computational and structural model of the full-length human prolactin receptor," *Nature Communications*, vol. 7, p. 11578, 2016.

[13] M. J. Waters and A. J. Brooks, ``JAK2 activation by growth hormone and other cytokines," *Biochemical Journal*, vol. 466, no. 1, pp. 1--11, 2015.

[14] B. C. Cunningham, M. Ultsch, A. M. de Vos, M. G. Mulkerrin, K. R. Clauser, and J. A. Wells, ``Dimerization of the extracellular domain of the human growth hormone receptor by a single hormone molecule," *Science*, vol. 254, no. 5033, pp. 821--825, 1991.

[15] M. Ultsch, A. M. de Vos, and A. A. Kossiakoff, ``Crystals of the complex between human growth hormone and the extracellular domain of its receptor," *Journal of Molecular Biology*, vol. 222, no. 4, pp. 865--868, 1991.

[16] D. J. Matthews, R. S. Topping, R. T. Cass, and L. B. Giebel, ``A sequential dimerization mechanism for erythropoietin receptor activation," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 93, no. 18, pp. 9471--9476, 1996.

[17] R. J. Brown, J. J. Adams, R. A. Pelekanos, Y. Wan, W. J. McKinstry, K. Palethorpe, R. M. Seeber, T. A. Monks, K. A. Eidne, M. W. Parker, and M. J. Waters, ``Model for growth hormone receptor activation based on subunit rotation within a receptor dimer," *Nature Structural & Molecular Biology*, vol. 12, no. 9, pp. 814--821, 2005.

[18] A. J. Brooks, W. Dai, M. L. O'Mara, D. Abankwa, Y. Chhabra, R. A. Pelekanos, O. Gardon, K. A. Tunny, K. M. Blucher, C. J. Morton, M. W. Parker, E. Sierecki, Y. Gambin, G. A. Gomez, K. Alexandrov, I. A. Wilson, M. Doxastakis, A. E. Mark, and M. J. Waters, ``Mechanism of activation of protein kinase JAK2 by the growth hormone receptor," *Science*, vol. 344, no. 6185, p. 1249783, 2014.

[19] J. S. Philo, K. H. Aoki, T. Arakawa, L. O. Narhi, and J. Wen, ``Dimerization of the extracellular domain of the erythropoietin (EPO) receptor by EPO: One high-affinity and one low-affinity interaction," *Biochemistry*, vol. 35, no. 5, pp. 1681--1691, 1996.

[20] I. Remy, I. A. Wilson, and S. W. Michnick, ``Erythropoietin receptor activation by a ligand-induced conformation change," *Science*, vol. 283, no. 5404, pp. 990--993, 1999.

[21] N. Seubert, Y. Royer, J. Staerk, K. F. Kubatzky, V. Moucadel, S. Krishnakumar, S. O. Smith, and S. N. Constantinescu, ``Active and inactive orientations of the transmembrane and cytosolic domains of the erythropoietin receptor dimer," *Molecular Cell*, vol. 12, no. 5, pp. 1239--1250, 2003.

[22] X. Pang and H.-X. Zhou, ``A common model for cytokine receptor activation: Combined scissor-like rotation and self-rotation of receptor dimer induced by class I cytokine," *PLoS Computational Biology*, vol. 8, no. 3, p. e1002427, 2012.

[23] G. Fuh, B. C. Cunningham, R. Fukunaga, S. Nagata, D. V. Goeddel, and J. A. Wells, ``Rational design of potent antagonists to the human growth hormone receptor," *Science*, vol. 256, pp. 1677--1680, 1992.

[24] S. Elliott, T. Lorenzini, D. Yanagihara, D. Chang, and G. Elliott, ``Activation of the erythropoietin (EPO) receptor by bivalent anti-EPO receptor antibodies," *Journal of Biological Chemistry*, vol. 271, no. 40, pp. 24691--24697, 1996.

[25] S. S. Watowich, D. J. Hilton, and H. F. Lodish, ``Activation and inhibition of erythropoietin receptor function: Role of receptor dimerization," *Molecular and Cellular Biology*, vol. 14, no. 6, pp. 3535--3549, 1994.

[26] O. Miura and J. N. Ihle, ``Dimer- and oligomerization of the erythropoietin receptor by disulphide bond formation and significance of the region near the WSXWS motif in intracellular transport," *Archives of Biochemistry and Biophysics*, vol. 306, no. 1, pp. 200--208, 1993.

[27] S. N. Constantinescu, T. Keren, M. Socolovsky, H. Nam, Y. I. Henis, and H. F. Lodish, ``Ligand-independent oligomerization of cell-surface erythropoietin receptor is mediated by the transmembrane domain," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 98, no. 8, pp. 4379--4384, 2001.

[28] T. Rose, A.-H. Pillet, V. Lavergne, B. Tamarit, P. Lenormand, J.-C. Rousselle, A. Namane, and J. Thèze, ``Interleukin-7 compartmentalizes its receptor signaling complex to initiate CD4 T lymphocyte response," *Journal of Biological Chemistry*, vol. 285, no. 20, pp. 14898--14908, 2010.

[29] C. A. McElroy, P. J. Holland, P. Zhao, J.-M. Lim, L. Wells, E. Eisenstein, and S. T. R. Walsh, ``Structural reorganization of the interleukin-7 signaling complex," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 109, no. 7, pp. 2503--2508, 2012.

[30] P. Hensley, M. L. Doyle, D. G. Myszka, R. W. Woody, M. R. Brigham-Burke, C. L. Erickson-Miller, C. A. Griffin, C. S. Jones, D. E. McNulty, S. P. O'Brien, B. Y. Amegadzie, L. MacKenzie, M. Ryan, and P. R. Young, ``Evaluating energetics of erythropoietin ligand binding to homodimerized receptor extracellular domains," in *Energetics of Biological Macromolecules, Part C* (M. L. Johnson and G. K. Ackers, eds.), vol. 323 of *Methods in Enzymology*, pp. 177--207, Academic Press, 2000.

[31] D. Poger and A. E. Mark, ``Turning the growth hormone receptor on: Evidence that hormone binding induces subunit rotation," *Proteins: Structure, Function and Bioinformatics*, vol. 78, pp. 1163--1174, 2010.

[32] F. S. Groothuizen, D. Poger, and A. E. Mark, ``Activating the prolactin receptor: Effect of the ligand on the conformation of the extracellular domain," *Journal of Chemical Theory and Computation*, vol. 6, no. 10, pp. 3274--3283, 2010.

[33] A. M. de Vos, M. Ultsch, and A. A. Kossiakoff, ``Human growth hormone and extracellular domain of its receptor: Crystal structure of the complex," *Science*, vol. 255, no. 5042, p. 306, 1992.

[34] P. A. Elkins, H. W. Christinger, Y. Sandowski, E. Sakal, A. Gertler, A. M. de Vos, and A. A. Kossiakoff, ``Ternary complex between placental lactogen and the extracellular domain of the prolactin receptor," *Nature Structural & Molecular Biology*, vol. 7, no. 9, pp. 808--815, 2000.

[35] J. van Agthoven, C. Zhang, E. Tallet, B. Raynal, S. Hoos, B. Baron, P. England, V. Goffin, and I. Broutin, ``Structural characterization of the stem-stem dimerization interface between prolactin receptor chains complexed with the natural hormone," *Journal of Molecular Biology*, vol. 404, no. 1, pp. 112--126, 2010.

[36] K. F. Kubatzky, W. Ruan, R. Gurezka, J. Cohen, R. Ketteler, S. S. Watowich, D. Neumann, D. Langosch, and U. Klingmüller, ``Self assembly of the transmembrane domain promotes signal transduction through the erythropoietin receptor," *Current Biology*, vol. 11, no. 2, pp. 110--115, 2001.

[37] A. Z. Ebie and K. G. Fleming, ``Dimerization of the erythropoietin receptor transmembrane domain in micelles," *Journal of Molecular Biology*, vol. 366, no. 2, pp. 517--524, 2007.

[38] L. McGraw, Kathy, G. M. Fuhler, J. O. Johnson, J. A. Clark, G. C. Caceres, L. Sokol, and A. F. List, ``Erythropoietin receptor signaling is membrane raft dependent," *PLoS One*, vol. 7, no. 4, p. e34477, 2012.

[39] K. Simons and E. Ikonen, ``Functional rafts in cell membranes," *Nature*, vol. 387, no. 6633, pp. 569--572, 1997.

[40] L. J. Pike, ``Rafts defined: a report on the keystone symposium on lipid rafts and cell function," *Journal of Lipid Research*, vol. 47, no. 7, pp. 1597--1598, 2006.

[41] H. Li, L. M. Ayer, J. Lytton, and J. P. Deans, ``Store-operated cation entry mediated by CD20 in membrane rafts," *Journal of Biological Chemistry*, vol. 278, no. 43, pp. 42427--42434, 2003.

[42] P. W. Janes, S. C. Ley, A. I. Magee, and P. S. Kabouridis, ``The role of lipid rafts in T cell antigen receptor (TCR) signalling," *Seminars in Immunology*, vol. 12, no. 1, pp. 23--34, 2000.

[43] D. J. Hanahan, R. M. Watts, and D. Pappajohn, ``Some chemical characteristics of the lipids of human and bovine erythrocytes and plasma," *Journal of Lipid Research*, vol. 1, no. 5, pp. 421--432, 1960.

[44] D. Marsh, *Handbook of lipid bilayers*. Boca Raton, FL: CRC Press, 1990.

[45] M. Valsecchi, L. Mauri, R. Casellato, S. Prioni, N. Loberto, A. Prinetti, V. Chigorno, and S. Sonnino, ``Ceramide and sphingomyelin species of fibroblasts and neurons in culture," *Journal of Lipid Research*, vol. 48, no. 2, pp. 417--424, 2007.

[46] A. R. Leach, *Molecular Modelling. Principles and applications*. Edinburgh Gate: Pearson Education Limited, 2 ed., 2001.

[47] H. J. C. Berendsen, *Simulation the Physical World*. The Edinburgh Building: Cambridge University Press, 2007.

[48] N. Schmid, A. P. Eichenberger, A. Choutko, S. Riniker, M. Winger, A. E. Mark, and W. F. van Gunsteren, ``Definition and testing of the GROMOS force-field versions 54A7 and 54B7," *European Biophysics Journal*, vol. 40, no. 7, pp. 843--856, 2011.

[49] D. van der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, and H. J. C. Berendsen, ``GROMACS: Fast, flexible, and free," *Journal of Computational Chemistry*, vol. 26, no. 16, pp. 1701--1718, 2005.

[50] B. Hess, C. Kutzner, D. van der Spoel, and E. Lindahl, ``GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation," *Journal of Chemical Theory and Computation*, vol. 4, no. 3, pp. 435--447, 2008.

[51] R. W. Hockney, ``The potential calculation and some applications," *Methods in Computational Physics*, vol. 9, pp. 136--211, 1970.

[52] L. Verlet, ``Computer `experiments' on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules," *Physical Review*, vol. 159, pp. 98--103, 1967.

[53] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, and J. R. Haak, ``Molecular dynamics with coupling to an external bath," *The Journal of Chemical Physics*, vol. 81, no. 8, pp. 3684--3690, 1984.

[54] J. Kyte, *Structure in protein chemistry*, ch. 4, pp. 117--146. Garland Publishing, Inc., 1995.

[55] E. Eyal, S. Gerzon, V. Potapov, M. Edelman, and V. Sobolev, ``The limit of accuracy of protein modeling: Influence of crystal packing on protein structure," *Journal of Molecular Biology*, vol. 351, pp. 431--442, 2005.

[56] H. Yamada, T. Nagae, and N. Watanabe, ``High-pressure protein crystallography of hen egg-white lysozyme," *Acta Crystallographica Section D: Biological Crystallography*, vol. 71, no. 4, pp. 742--753, 2015.

[57] Y. Zhang, L. Wang, S. Dey, M. Alnaeeli, S. Suresh, H. Rogers, R. Teng, and C. T. Noguchi, ``Erythropoietin action in stress response, tissue maintenance and metabolism," *International Journal of Molecular Sciences*, vol. 15, no. 6, pp. 10296--10333, 2014.

[58] N. Suzuki, H. Y. Mukai, and M. Yamamoto, ``*In vivo* regulation of erythropoiesis by chemically inducible dimerization of the erythropoietin receptor intracellular domain," *PLoS ONE*, vol. 10, no. 3, pp. 1--19, 2015.

[59] E. C. Lowe, A. Baslé, M. Czjzek, S. J. Firbank, and D. N. Bolam, ``A scissor blade-like closing mechanism implicated in transmembrane signaling in a bacteroides hybrid two-component system," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 109, no. 19, pp. 7298--7303, 2012.

[60] M. S. Lee, C. R. Glassman, N. R. Deshpande, H. B. Badgandi, H. L. Parrish, C. Uttamapinant, P. S. Stawski, A. Y. Ting, and M. S. Kuhns, ``A mechanical switch couples t cell receptor triggering to the cytoplasmic juxtamembrane regions of CD3$\zeta\zeta$," *Immunity*, vol. 43, no. 2, pp. 227--239, 2015.

[61] M. Atanasova and A. Whitty, ``Understanding cytokine and growth factor receptor activation mechanisms," *Critical Reviews in Biochemistry and Molecular Biology*, vol. 47, no. 6, pp. 502--530, 2012.

[62] M. J. Waters, A. J. Brooks, and Y. Chhabra, ``A new mechanism for growth hormone receptor activation of JAK2, and implications for related cytokine receptors," *JAK-STAT*, vol. 3, no. 2, p. e29569, 2014.

[63] X. Lu, A. W. Gross, and H. F. Lodish, ``Active conformation of the erythropoietin receptor: Random and cysteine-scanning mutagenesis of the extracellular juxtamembrane and transmembrane domains," *Journal of Biological Chemistry*, vol. 281, no. 11, pp. 7002--7011, 2006.

[64] ``The PyMOL molecular graphics system, version 1.6.0.0." The PyMOL Molecular Graphics System, Version 1.6.0.0 Schrödinger, LLC., June 2013.

[65] A. K. Malde, L. Zuo, M. Breeze, M. Stroet, D. Poger, P. C. Nair, C. Oostenbrink, and A. E. Mark, ``An automated force field topology builder (ATB) and repository: version 1.0.," *Journal of Chemical Theory and Computation*, vol. 7, no. 12, pp. 4026--4037, 2011.

[66] B. Hess, H. Bekker, H. J. C. Berendsen, and J. G. E. M. Fraaije, ``LINCS: A linear constraint solver for molecular simulations," *Jornal of Computational Chemistry*, vol. 18, no. 12, pp. 1463--1472, 1997.

[67] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, and J. Hermans, ``Interaction models for water in relation to protein hydration," in *Intermolecular forces* (B. Pullman, ed.), pp. 331--342, Dordrecht, The Netherlands: Reidel, 1981.

[68] S. Miyamoto and P. A. Kollman, ``SETTLE: An analytical version of the SHAKE and RATTLE algorithm for rigid water models," *Journal of Computational Chemistry*, vol. 13, no. 8, pp. 952--962, 1992.

[69] I. G. Tironi, R. Sperb, P. E. Smith, and W. F. van Gunsteren, ``A generalized reaction field method for molecular dynamics simulations," *The Journal of Chemical Physics*, vol. 102, no. 13, pp. 5451--5459, 1995.

[70] T. N. Heinz, W. F. van Gunsteren, and P. H. Hünenberger, ``Comparison of four methods to compute the dielectric permittivity of liquids from molecular dynamics simulations," *The Journal of Chemical Physics*, vol. 115, no. 3, pp. 1125--1136, 2001.

[71] K. A. Feenstra, B. Hess, and H. J. C. Berendsen, ``Improving efficiency of large time-scale molecular dynamics simulations of hydrogen-rich systems," *Journal of Computational Chemistry*, vol. 20, no. 8, pp. 786--798, 1999.

[72] C. Ott, H. Martens, I. Hassouna, B. Oliveira, C. Erck, M.-P. Zafeiriou, U.-K. Peteri, D. Hesse, S. Gerhart, B. Altas, T. Kolbow, H. Stadler, H. Kawabe, W.-H. Zimmermann, K.-A. Nave, W. Schulz-Schaeffer, O. Jahn, and H. Ehrenreich, ``Widespread expression of erythropoietin receptor in brain and its induction by injury," *Molecular Medicine*, vol. 21, no. 1, pp. 803--815, 2015.

[73] Z. Liu, V. S. Stoll, P. J. DeVries, C. G. Jakob, N. Xie, R. L. Simmer, S. E. Lacy, D. A. Egan, J. E. Harlan, R. R. Lesniewski, and E. B. Reilly, ``A potent erythropoietin-mimicking human antibody interacts through a novel binding site," *Blood*, vol. 110, no. 7, pp. 2408--2413, 2007.

[74] B. Rupp, *Biomolecullar crystallography. Principles, practice and application to structural biology.* New York NY: Garland Science, Taylor & Fancis Group, LLC, 1 ed., 2001.

[75] I. A. Wilson and L. K. Jolliffe, ``The structure, organization, activation and plasticity of the erythropoietin receptor," *Current Opinion in Structural Biology*, vol. 9, no. 6, pp. 696--704, 1999.

[76] M. Marquart, J. Walter, J. Deisenhofer, W. Bode, and R. Huber, ``The geometry of the reactive site and of the peptide groups in trypsin, trypsinogen and its complexes with inhibitors," *Acta Crystallographica Section B*, vol. 39, no. 4, pp. 480--490, 1983.

[77] T. Darden, D. York, and L. Pedersen, ``Particle mesh Ewald: An $N \cdot \log N$ method for Ewald sums in large systems," *Journal of Chemical Physics*, vol. 98, no. 12, pp. 10089--10092, 1993.

[78] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen, ``A smooth particle mesh Ewald method," *Journal of Chemical Physics*, vol. 103, no. 19, pp. 8577--8593, 1995.

[79] R. Walser, P. H. Hünenberger, and W. F. van Gunsteren, ``Comparison of different schemes to treat long-range electrostatic interactions in molecular dynamics simulations of a protein crystal," *Proteins: Structure, Function, and Genetics*, vol. 43, no. 4, pp. 509--519, 2001.

[80] R. Walser, P. H. Hünenberger, and W. F. van Gunsteren, ``Molecular dynamics simulations of a double unit cell in a protein crystal: Volume relaxation at constant pressure and correlation of motions between the two unit cells," *Proteins: Structure, Function, and Genetics*, vol. 48, no. 2, pp. 327--340, 2002.

[81] N. Yang, Y. Huang, J. Jiang, and S. J. Frank, ``Caveolar and lipid raft localization of the growth hormone receptor and its signaling elements: Impact on growth hormone signaling," *Journal of Biological Chemistry*, vol. 279, no. 20, pp. 20898--20905, 2004.

[82] V. Becker, D. Sengupta, R. Ketteler, G. M. Ullmann, J. C. Smith, and U. Klingmüller, ``Packing density of the erythropoietin receptor transmembrane domain correlates with amplification of biological responses," *Biochemistry*, vol. 47, no. 45, pp. 11771--11782, 2008.

[83] M. G. Wolf, M. Hoefling, C. Aponte-Santamaría, H. Grubmüller, and G. Groenhof, ``g_membed: Efficient insertion of a membrane protein into an equilibrated lipid bilayer with minimal perturbation," *Journal of Computational Chemistry*, vol. 31, no. 11, pp. 2169--2174, 2010.

[84] J. D. Hirst and C. L. Brooks III, ``Helicity, circular dichroism and molecular dynamics of proteins," *Journal of Molecular Biology*, vol. 243, no. 2, pp. 173--178, 1994.

[85] M. C. Manning and R. W. Woody, ``Theoretical CD studies of polypeptide helices: Examination of important electronic and geometric factors," *Biopolymers*, vol. 31, no. 5, pp. 569--586, 1991.

[86] U. Hemmann, C. Gerhartz, B. Heesel, J. Sasse, G. Kurapkat, J. Grötzinger, A. Wollmer, Z. Zhong, J. E. Darnell, Jr., L. Graeve, P. C. Heinrich, and F. Horn, ``Differential activation of acute phase response factor/Stat3 and Stat1 via the cytoplasmic domain of the interleukin 6 signal transducer gp130: II. Src homology SH2 domains define the specificity of STAT factor activation," *Journal of Biological Chemistry*, vol. 271, no. 22, pp. 12999--13007, 1996.

[87] S. N. Constantinescu, L. J. Huang, H. Nam, and H. F. Lodish, ``The erythropoietin receptor cytosolic juxtamembrane domain contains an essential, precisely oriented, hydrophobic motif," *Molecular Cell*, vol. 7, no. 2, pp. 377--385, 2001.

[88] C. J. Baier, J. Fantini, and F. J. Barrantes, ``Disclosure of cholesterol recognition motifs in transmembrane domains of the human nicotinic acetylcholine receptor," *Scientific Reports*, vol. 1, p. 69, 2011.

[89] J. Mahamid, S. Pfeffer, M. Schaffer, E. Villa, R. Danev, L. Kuhn Cuellar, F. Förster, A. A. Hyman, J. M. Plitzko, and W. Baumeister, ``Visualizing the molecular sociology at the hela cell nuclear periphery," *Science*, vol. 351, no. 6276, pp. 969--972, 2016.

[90] L. Zhang, A. Polyansky, and M. Buck, ``Modeling transmembrane domain dimers/trimers of plexin receptors: Implications for mechanisms of signal transmission across the membrane,'' *PLoS ONE*, vol. 10, no. 4, p. e0121513, 2015.

[91] B. A. Hall, J. P. Armitage, and M. S. P. Sansom, ``Mechanism of bacterial signal transduction revealed by molecular dynamics of tsr dimers and trimers of dimers in lipid vesicles,'' *PLoS Computational Biology*, vol. 8, no. 9, p. e1002685, 2012.

[92] M. Lelimousin, V. Limongelli, and M. S. P. Sansom, ``Conformational changes in the epidermal growth factor receptor: Role of the transmembrane domain investigated by coarse-grained metadynamics free energy calculations,'' *Journal of the American Chemical Society*, vol. 138, no. 33, pp. 10611--10622, 2016.

[93] F. E. Herrera and S. Pantano, ``Structure and dynamics of nano-sized raft-like domains on the plasma membrane,'' *The Journal of Chemical Physics*, vol. 136, no. 1, p. 015103, 2012.

[94] A. A. Petruk, S. Varriale, M. R. Coscia, L. Mazzarella, A. Merlino, and U. Oreste, ``The structure of the CD33$\zeta\zeta$ transmembrane dimer in POPC and raft-like lipid bilayer: A molecular dynamics study,'' *Biochimica et Biophysica Acta (BBA) - Biomembranes*, vol. 1828, no. 11, pp. 2637--2645, 2013.

[95] S. J. Singer and G. L. Nicolson, ``The fluid mosaic model of the structure of cell membranes,'' *Science*, vol. 175, no. 4023, pp. 720--731, 1972.

[96] K. Simons and G. Van Meer, ``Lipid sorting in epithelial cells,'' *Biochemistry*, vol. 27, no. 17, pp. 6197--6202, 1988.

[97] E. Rodriguez-Boulan and W. J. Nelson, ``Morphogenesis of the polarized epithelial cell phenotype,'' *Science*, vol. 245, no. 4919, pp. 718--725, 1989.

[98] R. F. Zwall, B. Roelofsen, and C. M. Colley, ``Localization of red cell membrane constituents,'' *Biochimica et Biophysica Acta (BBA) - Reviews on Biomembranes*, vol. 300, no. 2, pp. 159--182, 1973.

[99] A. Prinetti, V. Chigorno, G. Tettamanti, and S. Sonnino, ``Sphingolipid-enriched membrane domains from rat cerebellar granule cells differentiated in culture: A compositional study,'' *Journal of Biological Chemistry*, vol. 275, no. 16, pp. 11658--11665, 2000.

[100] B. Brügger, B. Glass, P. Haberkant, I. Leibrecht, F. T. Wieland, , and H.-G. Kräusslich, ``The HIV lipidome: A raft with an unusual composition,'' *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, no. 8, pp. 2641--2646, 2006.

[101] W. I. Clahoun and G. G. Shipley, ``Sphingomyelin-lecithin bilayer and their interaction with cholesterol,'' *Biochemistry*, vol. 18, no. 9, pp. 1717--1722, 1979.

[102] L. K. Bar, Y. Barenholz, and E. Thompson, Thomas, ``Effect of sphingomyelin composition on the phase structure of phosphatidylcholine-sphingomyelin bilayers," *Biochemistry*, vol. 36, no. 9, pp. 2507--2516, 1997.

[103] T. Nyholm, M. Nylund, A. Söderholm, and J. P. Slotte, ``Properties of palmitoyl phosphatidylcholine, sphingomyelin, and dihydrosphingomyelin bilayer membranes as reported by different fluorescent reporter molecules," *Biophysical Journal*, vol. 84, no. 2, pp. 987--997, 2003.

[104] A. Bunge, P. Müller, M. Stöckl, A. Herrmann, and D. Huster, ``Characterization of the ternary mixture of sphingomyelin, POPC, and cholesterol: Support for an inhomogenous lipid distribution at high temperature," *Biophysical Journal*, vol. 94, no. 7, pp. 2680--2690, 2008.

[105] R. A. Demel, J. W. C. M. Jansen, P. W. M. van Dijck, and L. L. M. van Deenen, ``The preferential interactions of cholesterol with different classes of phospholipids," *Biochimica et Biophysica Acta (BBA) - Biomembranes*, vol. 465, no. 1, pp. 1--10, 1977.

[106] S. W. Chiu, S. Vasudevan, E. Jakobsson, R. J. Mashi, and H. L. Scott, ``Structure of sphingomyeline bilayers: A simulation study," *Biophysical Journal*, vol. 85, pp. 3624--3635, 2003.

[107] G. A. Khelashvili and H. L. Scott, ``Combined Monte Carlo and molecular dynamics simulation of hydrated 18:0 sphingomyelin-cholesterol lipid bilayers," *Journal of Chemical Physics*, vol. 120, no. 20, pp. 9841--9847, 2004.

[108] S. A. Pandit, S. Vasudevan, S. W. Chiu, R. J. Mashl, E. Jakobsson, and H. L. Scott, ``Sphingomyelin-cholesterol domains in phospholipid membranes: Atomistic simulation," *Biophysical Journal*, vol. 87, pp. 1092--1100, 2004.

[109] S. A. Pandit, E. Jakobsson, and H. L. Scott, ``Simulation of the early stages of nano-domain formation in mixed bilayers of sphingomyelin, cholesterol, and dioleylphosphatidylcholine," *Biophysical Journal*, vol. 87, no. 5, pp. 3312--3322, 2004.

[110] E. Mombelli, R. Morris, W. Taylor, and F. Fraternali, ``Hydrogen-bonding propensities of sphingomyelin in solution and in a bilayer assembly: A molecular dynamics study," *Biophysical Journal*, vol. 84, no. 3, pp. 1507--1517, 2003.

[111] P. Niemelä, M. T. Hyvönen, and I. Vattulainen, ``Structure and dynamics of sphingomyelin bilayer: Insight gained through systematic comparison to phosphatidylcholine," *Biophysical Journal*, vol. 87, no. 5, pp. 2976--2989, 2004.

[112] P. S. Niemelä, S. Ollila, M. T. Hyvönen, M. Karttunen, and I. Vattulainen, ``Assessing the nature of lipid raft membranes," *PLoS Computational Biology*, vol. 3, no. 2, p. e34, 2007.

[113] D. Poger, W. F. van Gunsteren, and A. E. Mark, ``A new force field for simulating phosphatidylcholine bilayers," *Journal of Computational Chemistry*, vol. 30, no. 6, pp. 117--1125, 2010.

[114] P. R. Maulik, D. Atkinson, and G. G. Shipley, ``X-ray scattering of vesicles of N-acyl sphingomyleins: Determination of bilayer thickness," *Biophysical Journal*, vol. 50, no. 6, pp. 1071--1077, 1986.

[115] P. R. Maulik and G. G. Shipley, ``N-palmitoyl sphingomyelin bilayers: Structure and interactions with cholesterol and dipalmitoylphosphatidylcholine," *Biochemistry*, vol. 35, no. 24, pp. 8025--8034, 1996.

[116] P. J. Quinn, ``Structure of sphingomyelin bilayers and complexes with cholesterol forming membrane rafts," *Langmuir*, vol. 29, no. 30, pp. 9447--9456, 2013.

[117] T. Bartels, R. S. Lankalapalli, R. Bittman, K. Beyer, and M. F. Brown, ``Raftlike mixtures of sphingomyelin and cholesterol investigated by solid-state ^{2}H NMR spectroscopy," *Journal of the American Chemical Society*, vol. 130, no. 44, pp. 14521--14532, 2008.

[118] T. Mehnert, K. Jacob, R. Bittman, and K. Beyer, ``Structure and lipid interaction of N-palmitoylsphingomyelin in bilayer membranes as revealed by ^{2}H-NMR spectroscopy," *Biophysical Journal*, vol. 90, no. 3, pp. 939--946, 2006.

[119] M. W. Schmidt, K. K. Baldridge, J. A. Boatz, S. T. Elbert, M. S. Gordon, J. H. Jensen, S. Koseki, N. Matsunaga, K. A. Nguyen, S. Su, T. L. Windus, M. Dupuis, and J. A. Montgomery, ``General atomic and molecular electronic structure system," *Journal of Computational Chemistry*, vol. 14, no. 11, pp. 1347--1363, 1993.

[120] W. T. Carnall, P. R. Fields, and K. Rajnak, ``Electronic energy levels in the trivalent lanthanide aquo ions. i. Pr^{3+}, Nd^{3+}, Pm^{3+}, Sm^{3+}, Dy^{3+}, Ho^{3+}, Er^{3+}, and Tm^{3+}," *The Journal of Chemical Physics*, vol. 49, no. 10, pp. 4424--4442, 1968.

[121] C. Lee, W. Yang, and R. G. Parr, ``Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density," *Physical Review B*, vol. 37, no. 2, pp. 785--789, 1988.

[122] J. P. Perdew and Y. Wang, ``Accurate and simple analytic representation of the electron-gas correlation energy," *Physical Reiew. B*, vol. 45, no. 23, pp. 13244--13249, 1992.

[123] A. D. Becke, ``Density-functional thermochemistry. III. the role of exact exchange," *The Journal of Chemical Physics*, vol. 98, no. 7, pp. 5648--5652, 1993.

[124] S. Miertuš, E. Scrocco, and J. Tomasi, ``Electrostatic interaction of a solute with a continuum. a direct utilizaion of *ab initio* molecular potentials for the prevision of solvent effects," *Chemical Physics*, vol. 55, no. 1, pp. 117--129, 1981.

[125] D. Poger and A. Mark, ``On the validation of molecular dynamics simulations of saturated and *cis*-monosaturated phosphatidylcholine lipid bilayers: A comparison with experiment," *Journal of Chemical Theory and Computation*, vol. 6, pp. 325--336, 2010.

[126] W. J. Allen, J. A. Lemkul, and D. R. Bevan, ``GridMAT-MD: A grid-based membrane analysis tool for use with molecular dynamics," *Journal of Computational Chemistry*, vol. 30, no. 12, pp. 1952--1958, 2009.

[127] A. Seelig and J. Seelig, ``Effect of a single *cis* double bond on the structure of a phospholipid bilayer," *Biochemistry*, vol. 16, no. 1, pp. 45--50, 1977.

[128] T. Okazaki, A. Bielawska, R. M. Bell, and Y. A. Hannun, ``Role of ceramide as a lipid mediator of 1α,25-dihydroxyvitamin D_3-induced HL-60 cell differentiation," *Journal of Biological Chemistry*, vol. 265, no. 26, pp. 15823--15831, 1990.

[129] A. J. Sodt, R. W. Pastor, and E. Lyman, ``Hexagonal substructure and hydrogen bonding in liquid-ordered phases containing palmitoyl sphingomyelin," *Biophysical Journal*, vol. 109, no. 5, pp. 948--955, 2015.

[130] P. S. Coppock and J. T. Kindt, ``Determination of phase transition temperatures for atomistic models of lipids from temperature-dependent stripe domain growth kinetics," *The Journal of Physical Chemistry B*, vol. 114, no. 35, pp. 11468--11473, 2010.

[131] M. T. Hyvönen and P. T. Kovanen, ``Molecular dynamics simulation of sphingomyelin bilayer," *The Journal of Physical Chemistry B*, vol. 107, no. 34, pp. 9102--9108, 2003.

[132] R. Venable, A. Sodt, B. Rogaski, H. Rui, E. Hatcher, A. MacKerell Jr., R. Pastor, and J. Klauda, ``CHARMM all-atom additive force field for sphingomyelin: Elucidation of hydrogen bonding and of positive curvature," *Biophysical Journal*, vol. 107, no. 1, pp. 134--145, 2014.

[133] J. P. M. Jämbeck and A. P. Lyubartsev, ``Another piece of the membrane puzzle: Extending slipids further," *Journal of Chemical Theory and Computation*, vol. 9, no. 1, pp. 774--784, 2013.

[134] P. P. Zenatti, D. Ribeiro, W. Li, L. Zuurbier, M. C. Silva, M. Paganin, J. Tritapoe, J. A. Hixon, A. B. Silveira, B. A. Cardoso, L. M. Sarmento, N. Correia, M. L. Toribio, J. Kobarg, M. Horstmann, R. Pieters, S. R. Brandalise, A. A. Ferrando, J. P. Meijerink, S. K. Durum, J. A. Yunes, and J. T. Barata, ``Oncogenic *IL7R* gain-of-function mutations in childhood T-cell acute lymphoblastic leukemia," *Nature Genetics*, vol. 43, no. 10, pp. 932--939, 2011.

[135] K. Swaminathan, P. Flynn, R. J. Reece, and R. Marmorstein, ``Crystal structure of a PUT3-DNA complex reveals a novel mechanism for DMA recognition by a protein containing a Zn_2Cys_6 binuclear cluster," *Nature Structural and Molecular Biology*, pp. 751--759, 1997.

[136] F. K. Junius, S. I. O'Donoghue, M. Nilges, A. S. Weiss, and G. F. King, ``High resolution NMR solution structure of the leucine zipper domain of the c-Jun homodimer," *Journal of Biological Chemistry*, vol. 271, no. 23, pp. 13663--13667, 1996.

[137] M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess, and E. Lindahl, ``GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers," *SoftwareX*, vol. 1-2, pp. 19--25, 2015.

[138] W. Kabsch and C. Sander, ``Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features," *Biopolymers*, vol. 22, no. 12, pp. 2577--2637, 1983.

[139] W. G. Touw, C. Baakman, J. Black, T. A. H. te Beek, E. Krieger, R. P. Joosten, and G. Vriend, ``A series of PDB-related databanks for everyday needs," *Nucleic Acids Research*, vol. 43, pp. D364--D368, 2014.

[140] R. Nygaard, Y. Zou, R. Dror, T. Mildorf, D. Arlow, A. Manglik, A. Pan, C. Liu, J. Fung, M. Bokoch, F. Thian, T. Kobilka, D. Shaw, L. Mueller, R. Prosser, and B. Kobilka, ``The dynamic process of β2-adrenergic receptor activation," *Cell*, vol. 152, no. 3, pp. 532--542, 2013.

[141] B. Tamarit, F. Bugault, A.-H. Pillet, V. Lavergne, P. Bochet, N. Garin, U. Schwarz, J. Th□ze, and T. Rose, ``Membrane microdomains and cytoskeleton organization shape and regulate the IL-7 receptor signalosome in human CD4 T-cells," *Journal of Biological Chemistry*, vol. 288, no. 12, pp. 8691--8701, 2013.

[142] K. J. Walters, K. T. Dayie, R. J. Reece, M. Ptashne, and G. Wagner, ``Structure and mobility of the PUT3 dimer," *Nature Structural and Molecular Biololgy*, vol. 4, pp. 744--750, 1997.

[143] K. B. Koziara, M. Stroet, A. K. Malde, and A. E. Mark, ``Testing and validation of the Automated Topology Builder (ATB) version 2.0: Prediction of hydration free enthalpies," *Journal of Computer-Aided Molecular Design*, vol. 28, no. 3, pp. 221--233, 2014.

[144] A. W. Götz, M. J. Williamson, D. Xu, D. Poole, S. Le Grand, and R. C. Walker, ``Routine microsecond molecular dynamics simulations with AMBER on GPUs. 1. Generalized Born," *Journal of Chemical Theory and Computation*, vol. 8, no. 5, pp. 1542--1555, 2012.

[145] R. Salomon-Ferrer, A. W. Götz, D. Poole, S. Le Grand, and R. C. Walker, ``Routine microsecond molecular dynamics simulations with AMBER on GPUs. 2. explicit solvent Particle Mesh Ewald," *Journal of Chemical Theory and Computation*, vol. 9, no. 9, pp. 3878--3888, 2012.

Appendix A

Sphingomyelin Parameters

```
;
; Parameters for sphingomylein
; To be used in conjunction with the GROMOS 54a7 force field
;
;
;    CG1        OP1     H01—O1     H1B
;     |         ||          |       |
; CG2—N—CB—CA—OA—P—O3—C3—C2—C1—C1A═C1B—C1C—C1D−−..−−C1O
;     |          |       |      |
;    CG3        OP2(−)   N2—HN2 H1A
;                                       |
;                                      C2A═O2A
;                                       |
;                                      C2B—C2C−−...−−C2P
;
; ### HOW ITP FILE WAS CONSTRUCTED ###
;
; The charges, bond terms, bond angle and dihedrals for phosphocholine were
; taken from POPC.itp see:
; Poger D., van Gunsteren W.F. & Mark A.E. (2010) J. Comput. Chem. 31(6),
; 1117−1125 Poger D. & Mark A.E. (2010) J. Chem. Theory Comput. 6(1), 325−336
;
; The charges for the backbone carbons C3−C2−C1, the amide bond, the hydroxyl
; group and the double bond were taken from ATB results for
;ceramide−1−phosphocholine (molecule ID 9232), see also results for
; Molecule ID 9319 and 9318.
;
; The excess positive charge of 0.209 generated from taking POPC
; phosphocholine and exlcuding C3 from the phosphocholine charge group
; generated by the ATB was subtracted from C3, C2, N2, HN2, C2A, O2A
; and C2B (0.209 / 7 = 0.0299) and rounded to 3 decimal places. C3,
; C2, N2, HN2, C2A, O2A and C2B were placed into a single charge group as
; per the ATB results for molecule ID 9319.
;
; Charges over atoms were rounded to 1 decimal place.
;
; ### END ###
```

[moleculetype]

```
; Name    nrexcl
PSM       3
```

[atoms]

```
;  nr   type   resnr   resid  atom   cgnr   charge     mass      total_charge
    1   CH3p     1     PSM     CG1     1     0.400   15.0350
    2   CH3p     1     PSM     CG2     1     0.400   15.0350
    3   CH3p     1     PSM     CG3     1     0.400   15.0350
    4     NL     1     PSM       N     1    -0.500   14.0067
    5    CH2     1     PSM      CB     1     0.300   14.0270        ;   1.000
    6    CH2     1     PSM      CA     2     0.400   14.0270
    7     OA     1     PSM      OA     2    -0.800   15.9994
    8      P     1     PSM       P     2     1.700   30.9738
    9     OM     1     PSM     OP1     2    -0.800   15.9994
   10     OM     1     PSM     OP2     2    -0.800   15.9994
   11     OA     1     PSM      O3     2    -0.700   15.9994        ;  -1.000
   12    CH2     1     PSM      C3     3     0.300   14.0270
   13    CH1     1     PSM      C2     3     0.300   13.0190
   14      N     1     PSM      N2     3    -0.800   14.0067
   15      H     1     PSM     HN2     3     0.300    1.0080
   16      C     1     PSM     C2A     3     0.700   12.0110
   17      O     1     PSM     O2A     3    -0.700   15.9994
   18    CH2     1     PSM     C2B     3    -0.100   14.0270        ;   0.000
   19    CH2     1     PSM     C2C     5     0.000   14.0270        ;   0.000
   20    CH2     1     PSM     C2D     6     0.000   14.0270        ;   0.000
   21    CH2     1     PSM     C2E     7     0.000   14.0270        ;   0.000
   22    CH2     1     PSM     C2F     8     0.000   14.0270        ;   0.000
   23    CH2     1     PSM     C2G     9     0.000   14.0270        ;   0.000
   24    CH2     1     PSM     C2H    10     0.000   14.0270        ;   0.000
   25    CH2     1     PSM     C2I    11     0.000   14.0270        ;   0.000
   26    CH2     1     PSM     C2J    12     0.000   14.0270        ;   0.000
   27    CH2     1     PSM     C2K    13     0.000   14.0270        ;   0.000
   28    CH2     1     PSM     C2L    14     0.000   14.0270        ;   0.000
   29    CH2     1     PSM     C2M    15     0.000   14.0270        ;   0.000
   30    CH2     1     PSM     C2N    16     0.000   14.0270        ;   0.000
   31    CH2     1     PSM     C2O    17     0.000   14.0270        ;   0.000
   32    CH3     1     PSM     C2P    18     0.000   15.0350        ;   0.000
   33    CH1     1     PSM      C1    19     0.200   13.0190
   34     OA     1     PSM      O1    19    -0.600   15.9994
   35      H     1     PSM     HO1    19     0.400    1.0080        ;   0.000
   36    CH1     1     PSM     C1A    20    -0.100   13.0190
   37    CH1     1     PSM     C1B    20     0.100   13.0190        ;   0.000
   38    CH2     1     PSM     C1C    21     0.000   14.0270        ;   0.000
   39    CH2     1     PSM     C1D    22     0.000   14.0270        ;   0.000
   40    CH2     1     PSM     C1E    23     0.000   14.0270        ;   0.000
   41    CH2     1     PSM     C1F    24     0.000   14.0270        ;   0.000
   42    CH2     1     PSM     C1G    25     0.000   14.0270        ;   0.000
   43    CH2     1     PSM     C1H    26     0.000   14.0270        ;   0.000
   44    CH2     1     PSM     C1I    27     0.000   14.0270        ;   0.000
   45    CH2     1     PSM     C1J    28     0.000   14.0270        ;   0.000
   46    CH2     1     PSM     C1K    29     0.000   14.0270        ;   0.000
   47    CH2     1     PSM     C1L    30     0.000   14.0270        ;   0.000
   48    CH2     1     PSM     C1M    31     0.000   14.0270        ;   0.000
   49    CH2     1     PSM     C1N    32     0.000   14.0270        ;   0.000
   50    CH3     1     PSM     C1O    33     0.000   15.0350        ;   0.000
; total charge of the molecule:   0.000
```

[bonds]

```
;   ai    aj  funct   c0          c1
     1     4    2    gb_21
     2     4    2    gb_21
     3     4    2    gb_21
     4     5    2    gb_21
     5     6    2    gb_27
     6     7    2    gb_18
     7     8    2    gb_28
     8     9    2    gb_24
     8    10    2    gb_24
     8    11    2    gb_28
    11    12    2    gb_18
    13    12    2    0.1530    7.1500e+06
    13    33    2    0.1530    7.1500e+06
    14    13    2    0.1470    8.7100e+06
    14    16    2    0.1350    1.0300e+07
    15    14    2    0.1000    1.8700e+07
    17    16    2    0.1230    1.6600e+07
    18    16    2    0.1520    5.4300e+06
    18    19    2    0.1530    7.1500e+06
    19    20    2    0.1530    7.1500e+06
    20    21    2    0.1530    7.1500e+06
    21    22    2    0.1530    7.1500e+06
    22    23    2    0.1530    7.1500e+06
    23    24    2    0.1530    7.1500e+06
    24    25    2    0.1530    7.1500e+06
    25    26    2    0.1530    7.1500e+06
    26    27    2    0.1530    7.1500e+06
    27    28    2    0.1530    7.1500e+06
    28    29    2    0.1530    7.1500e+06
    29    30    2    0.1530    7.1500e+06
    30    31    2    0.1530    7.1500e+06
    31    32    2    0.1530    7.1500e+06
    33    36    2    0.1520    5.4300e+06
    34    33    2    0.1435    6.1000e+06
    35    34    2    0.1000    1.5700e+07
    36    37    2    0.1330    1.1800e+07
    38    37    2    0.1520    5.4300e+06
    38    39    2    0.1530    7.1500e+06
    39    40    2    0.1530    7.1500e+06
    40    41    2    0.1530    7.1500e+06
    41    42    2    0.1530    7.1500e+06
    42    43    2    0.1530    7.1500e+06
    43    44    2    0.1530    7.1500e+06
    44    45    2    0.1530    7.1500e+06
    45    46    2    0.1530    7.1500e+06
    46    47    2    0.1530    7.1500e+06
    47    48    2    0.1530    7.1500e+06
    48    49    2    0.1530    7.1500e+06
    49    50    2    0.1530    7.1500e+06
```

```
[ pairs ]
;   ai    aj   funct   ;   all 1−4 pairs but the ones excluded in GROMOS itp
     1     6     1
     5     8     1
     6     2     1
     6     3     1
     6     9     1
     6    10     1
     6    11     1
     7     4     1
     7    12     1
     8    13     1
     9    12     1
    10    12     1
    11    14     1
    11    33     1
    12    16     1
    12    34     1
    12    36     1
    13    17     1
    13    18     1
    13    35     1
    13    37     1
    14    34     1
    14    36     1
    15    12     1
    15    17     1
    15    18     1
    15    33     1
    16    33     1
    14    19     1
    17    19     1
    16    20     1
    18    21     1
    19    22     1
    20    23     1
    21    24     1
    22    25     1
    23    26     1
    24    27     1
    25    28     1
    26    29     1
    27    30     1
    28    31     1
    29    32     1
    33    38     1
    34    37     1
    35    36     1
    36    39     1
    37    40     1
    38    41     1
    39    42     1
    40    43     1
    41    44     1
    42    45     1
    43    46     1
    44    47     1
    45    48     1
    46    49     1
    47    50     1
```

```
[ angles ]
;   ai   aj   ak  funct   angle      fc
     1    4    2    2    ga_13
     1    4    3    2    ga_13
     1    4    5    2    ga_13
     2    4    3    2    ga_13
     2    4    5    2    ga_13
     3    4    5    2    ga_13
     4    5    6    2    ga_15
     5    6    7    2    ga_15
     6    7    8    2    ga_26
     7    8    9    2    ga_14
     7    8   10    2    ga_14
     7    8   11    2    ga_5
     8   11   12    2    ga_26
     9    8   10    2    ga_29
     9    8   11    2    ga_14
    10    8   11    2    ga_14
    11   12   13    2    110.30   524.00
    12   13   33    2    111.00   530.00
    13   14   16    2    125.00   750.00
    13   33   34    2    109.50   320.00
    13   33   36    2    111.00   530.00
    14   13   12    2    111.00   530.00
    14   13   33    2    111.00   530.00
    14   16   17    2    121.40   690.00
    14   16   18    2    120.00   560.00
    15   14   13    2    116.00   465.00
    15   14   16    2    115.00   460.00
    17   16   18    2    121.00   685.00
    16   18   19    2    111.00   530.00
    18   19   20    2    111.00   530.00
    19   20   21    2    111.00   530.00
    20   21   22    2    111.00   530.00
    21   22   23    2    111.00   530.00
    22   23   24    2    111.00   530.00
    23   24   25    2    111.00   530.00
    24   25   26    2    111.00   530.00
    25   26   27    2    111.00   530.00
    26   27   28    2    111.00   530.00
    27   28   29    2    111.00   530.00
    28   29   30    2    111.00   530.00
    29   30   31    2    111.00   530.00
    30   31   32    2    111.00   530.00
    33   36   37    2    125.00   750.00
    34   33   36    2    111.00   530.00
    35   34   33    2    108.53   443.00
    38   37   36    2    125.00   750.00
    37   38   39    2    111.00   530.00
    38   39   40    2    111.00   530.00
    39   40   41    2    111.00   530.00
    40   41   42    2    111.00   530.00
    41   42   43    2    111.00   530.00
    42   43   44    2    111.00   530.00
    43   44   45    2    111.00   530.00
    44   45   46    2    111.00   530.00
    45   46   47    2    111.00   530.00
    46   47   48    2    111.00   530.00
    47   48   49    2    111.00   530.00
    48   49   50    2    111.00   530.00
```

[dihedrals]

```
;   ai    aj    ak    al  funct     ph0      cp      mult
     1     4     5     6    1       gd_29
     4     5     6     7    1       gd_4
     4     5     6     7    1       gd_36
     5     6     7     8    1       gd_29
     6     7     8    11    1       gd_20
     6     7     8    11    1       gd_27
     7     8    11    12    1       gd_20
     7     8    11    12    1       gd_27
     8    11    12    13    1       gd_29
    12    13    33    34    1       0.00     3.77     3
    13    14    16    18    1     180.00    33.50     2
    13    33    36    37    1     180.00     1.00     6
    16    14    13    12    1       0.00     3.77     6
    14    16    18    19    1       0.00     3.77     3
    16    18    19    20    1       0.00     3.77     3
    18    19    20    21    1       0.00     3.77     3
    19    20    21    22    1       0.00     3.77     3
    20    21    22    23    1       0.00     3.77     3
    21    22    23    24    1       0.00     3.77     3
    22    23    24    25    1       0.00     3.77     3
    23    24    25    26    1       0.00     3.77     3
    24    25    26    27    1       0.00     3.77     3
    25    26    27    28    1       0.00     3.77     3
    26    27    28    29    1       0.00     3.77     3
    27    28    29    30    1       0.00     3.77     3
    28    29    30    31    1       0.00     3.77     3
    29    30    31    32    1       0.00     3.77     3
    33    13    12    11    1       0.00     3.77     3
    35    34    33    13    1       0.00     1.26     3
    38    37    36    33    1     180.00     1.53     2
    36    37    38    39    1       0.00     3.77     3
    37    38    39    40    1       0.00     3.77     3
    38    39    40    41    1       0.00     3.77     3
    39    40    41    42    1       0.00     3.77     3
    40    41    42    43    1       0.00     3.77     3
    41    42    43    44    1       0.00     3.77     3
    42    43    44    45    1       0.00     3.77     3
    43    44    45    46    1       0.00     3.77     3
    44    45    46    47    1       0.00     3.77     3
    45    46    47    48    1       0.00     3.77     3
    46    47    48    49    1       0.00     3.77     3
    47    48    49    50    1       0.00     3.77     3
```

[dihedrals]

```
; GROMOS improper  dihedrals
;   ai    aj    ak    al  funct    angle      fc
    13    12    33    14    2      35.26    334.72
    14    15    13    16    2       0.00    167.36
    16    14    17    18    2       0.00    167.36
    33    34    13    36    2      35.26    334.72
    33    36    37    38    2       gi_4
```

[exclusions]

```
;   ai    aj   funct   ;   GROMOS 1-4 exclusions
```

[virtual_sites3]

```
;   ai     aj     ak     al  funct            c0             c1
    15     14     16     13    2
```